Research Article

Maria Lialiou*, Stavroula Sotiropoulou and
Adamantios I. Gafos

Spatiotemporal coordination in word-medial stop-lateral and s-stop clusters of American English

<https://doi.org/10.1515/phon-2021-2010>

Published online October 8, 2021

Abstract: This paper is concerned with the relation between syllabic organization and intersegmental spatiotemporal coordination using Electromagnetic Articulation recordings from seven speakers of American English (henceforth, English). Whereas previous work on English has focused on word-initial clusters (preceding a vowel whose identity was not systematically varied), the present work examined word-medial clusters /pl, kl, sp, sk/ in the context of three different vowel heights (high, mid, low). Our results provide evidence for a global organization for the segments involved in these cluster-vowel combinations. This is reflected in a number of ways: compression of the prevocalic consonant and reduction of CV timing in the word-medial cluster case compared to its singleton paired word in both stop-lateral and s-stop clusters, early vowel initiation (as permitted by the clusters' phonetic properties), and presence of compensatory relations between phonetic properties of different segments or intersegmental transitions within each cluster. In other words, we find that the global organization presiding over the segments partaking in these word-medial tautosyllabic CCVs is pleiotropic, that is, simultaneously expressed in multiple phonetic exponents rather than via a privileged metric such as c-center stability or any other such given single measure employed in previous works.

***Corresponding author: Maria Lialiou**, Department of Linguistics and Cognitive Sciences, University of Potsdam, Karl-Liebknecht-Str 24/25, Haus 14, Potsdam 14476, Germany, E-mail: maria_lialiou@hotmail.gr

Stavroula Sotiropoulou, Department of Linguistics and Cognitive Sciences, University of Potsdam, Potsdam, Germany

Adamantios I. Gafos, Department of Linguistics and Cognitive Sciences, University of Potsdam, Potsdam, Germany; and Haskins Laboratories, New Haven, CT, USA

Keywords: American English; intersegmental coordination; s-stop clusters; stop-lateral clusters; syllabic structure

1 Introduction

The question of how linguistic organization is related to vocal tract action has inspired phonetic studies for many years investigating this relation in the domain of syllables. This paper is concerned with the relation between syllabic structure and intersegmental spatiotemporal coordination for word-medial stop-lateral and s-stop clusters in American English (henceforth, English) using articulatory data from seven speakers.

Several studies have investigated the spatiotemporal coordination patterns of English word-initial clusters (e.g., Browman and Goldstein 1988; Honorof and Browman 1995; Marin and Pouplier 2010 among others). Spatiotemporal coordination patterns of English word-medial clusters have not been systematically investigated. To our knowledge, only the study of Shaw and Gafos (2015) examined English clusters in the word-medial position using data from the X-ray Microbeam archive (Westbury 1994). However, the clusters examined were limited. Specifically, word-medial stop-laterals were not examined (Shaw and Gafos 2015 had only stop-rhotics) and, from s-stop clusters, Shaw and Gafos (2015) investigated only the /sp/ cluster. We therefore extend the range of clusters to the prototypical complex onsets of English, stop-lateral clusters, and the less prototypical s-stop clusters. Furthermore, our study systematically assesses the effect of vowel context in which the clusters under investigation are embedded. Each of our clusters was followed by three vowels differing with respect to height (high, mid, low). Whereas previous studies examined the effect of various characteristics of the consonants, such as their place of articulation or voicing, on the spatiotemporal patterns, the effect of vowel differences on spatiotemporal measures of syllabic organization remains understudied. With the exception of Brunner et al. (2014), who used different vowels after clusters in German, very little is known about if and how the spatiotemporal patterns thought to express syllabic organization are affected by the vowel following the clusters whose syllabic organization is at issue.

In past studies seeking phonetic correlates of syllable organization in articulation, a well-practiced approach has been the use of stability-based measures. More specifically, a prominent working hypothesis in these studies has been that different stability patterns of intervals spanning over the phonetic sequences under investigation can identify different syllabic organizations (e.g., Browman and Goldstein 1988; Byrd 1995). For example, in languages that admit complex onsets, as in English or German, it was shown in early work (Browman and Goldstein 1988;

Byrd 1995; Honorof and Browman 1995) that the most stable interval across sequences with different numbers of initial consonants (CVC, CCVC, CCCVC) is an interval which stretches from the temporal midpoint (c-center) of the whole prevocalic consonantal sequence (hence, global) up to an anchor point which represents the end of the hypothesized syllable (e.g., Browman and Goldstein 1988; Byrd 1995; Honorof and Browman 1995; Marin and Pouplier 2010; Shaw and Gafos 2015 among others). We refer to this interval as the global timing interval. Instead, in languages that do not admit complex onsets like Moroccan Arabic (Shaw et al. 2009), it has been reported that the most stable interval across CVC, CCVC, CCCVC is an interval which spans from the immediate prevocalic consonant up to the anchor point which represents the end of the hypothesized syllabic unit (e.g., Goldstein et al. 2007; Hermes et al. 2015 on Berber). We refer to this interval as the local timing interval. Thus, global timing interval stability has been assumed to be the phonetic correlate of complex onset organization, while local timing interval stability has been considered to be the phonetic correlate of simplex onset organization.

Nevertheless, subsequent empirical work has reported puzzling results of stability patterns, meaning that the expected stability pattern for both complex and simplex onsets has not been uniformly observed. A prominent example comes from German. Pouplier (2012) examined the clusters /bl, pl, gm, km, sk/ in syllable onset and coda position and reported some evidence for the expected global timing stability pattern in onsets but also strong cluster-specific effects on the stability patterns. Overall, not all clusters or cluster-vowel combinations exhibited the expected global timing stability, although there is solid phonological evidence for the syllabic onset status of all the clusters investigated (Wiese 1996). German is not the only language for which such puzzling results have been reported; for example, Marin (2013), on Romanian, reported evidence for the expected global timing stability pattern in sibilant-initial onset clusters /sk, sp, sm/ but also evidence for the unexpected local timing stability for stop-initial onset clusters /ps, ks, kt, kn/, despite native speakers classifying both cluster groups as complex onsets. Further, Hermes et al. (2017) on Polish, a language which is also said to admit complex onsets, reported the expected global timing stability pattern only for the /pl/ cluster with the clusters /pr, kr, kl/ exhibiting the unexpected local timing stability. Lastly, Shaw et al. (2009, 2011) report speaker-specific patterns in Arabic data, some of which are comparable to patterns reported previously for English, although Arabic and English are considered to be prototypical examples of the complex (English) versus simplex (Arabic) onset typological distinction. These results show that stability-based measures cannot diagnose syllabic structure consistently.

In the context of past work on evidence for syllabic organization in phonetic data which tends to consider one or two (presumed) privileged measures, our present study quantifies several different spatiotemporal measures (following Gafos et al. 2020; Sotiropoulou et al. 2020). That is, in addition to stability-based measures, we also quantified the duration of the prevocalic consonant across CV and CCV in the word-medial case, the relation between inter-gestural lag between the two consonants in the CC cluster (expressed by the inter-plateau interval between the two main oral C constrictions) and the durations of these consonants, and finally vowel initiation with respect to the preceding segment or segmental combination. We argue that joint consideration of these measures enables one to see patterns of how different measures relate to one another and provides crucial evidence for global organization of the segments that are part of the hypothesized syllabic units. This evidence, as would be argued in detail in the forthcoming, would not otherwise (i.e., when looking only at interval stabilities) be discernible.

The paper is organized as follows. Section 2 outlines the method and describes the analyses employed in our English word-medial articulatory data. Section 3 reports the results of the various spatiotemporal measures. Section 4 summarizes and discusses the obtained results, highlighting how the joint consideration of the employed spatiotemporal measures demonstrate the presence of global organization in both stop-lateral and s-stop clusters investigated herein. Section 5 concludes with implications of our results on the link between syllabic structure and spatiotemporal organization.

2 Materials and methods

2.1 Participants and data

Articulatory data from seven native speakers of American English (vp01, vp02, vp03, vp04, vp05, vp06, vp08; data from another speaker, vp07, was recorded but was excluded from the analyses due to technical issues) were analyzed. These data were registered at the University of Potsdam as part of a large articulatory corpus. Four out of the seven speakers were female and three were male (age range 20–30). All speakers were monolingual speakers of American English born and raised in the USA with no speech or hearing impairment. All speakers self-reported knowledge of a foreign language up to the level of the upper-intermediate (B2 level), provided written consent, and received reimbursement for their participation.

The data were acquired using the Electromagnetic Articulometer (EMA) 16-channel NDI Wave system which tracks the three-dimensional movement of

sensors attached inside and outside the vocal tract. Three sensors were attached to each speaker's tongue midsagittally: a tongue tip (henceforth, TT) sensor was placed 1 cm posterior to the tongue apex, a tongue mid (henceforth, TM) sensor was placed 2 cm posterior to the TT sensor, and a tongue back (henceforth, TB) sensor 2 cm posterior to the TM sensor. Additional sensors were adhered on the upper and lower lip, and the jaw (low incisors). Lastly, reference sensors (for head correction) were adhered on the upper incisors, the bridge of the nose, and behind the ears (left and right mastoid). The data were acquired at a sampling rate of 400 Hz and stored in a computer connected to the articulometer. For head movement correction, the head movement as captured by the reference sensors was subtracted from the movement of the rest of the sensors. The data obtained from the reference sensors were filtered using a cut-off frequency of 5 Hz, while the data obtained from the rest of the sensors were filtered using a cut-off frequency of 20 Hz. Audio data were also collected in parallel with the articulatory data.

2.2 Stimuli

The stimuli consist of disyllabic English words with stop-lateral and s-stop clusters or single consonants in the word-medial position.¹ Specifically, the clusters included in the analysis are word-medial /pl, kl, sp, sk/. The /pl, kl/ clusters were paired with single lateral-vowel (lV) sequences, while the /sp, sk/ clusters were paired with single stop labial-vowel (pV) and velar-vowel (kV) sequences respectively, so that in any VCV ~ VCCV pair the prevocalic C remained the same as in the pair *append* ~ *suspend*. Further, for each VCV ~ VsCV pair, three different vowel contexts (high, mid, low), and for each VCV ~ VCIV pair, two different vowel contexts (high, low) are included.² The vowel context was the same for each singleton and VCCV pair. Likewise, the postvocalic C was kept the same within word pairs to the extent possible (e.g., /n/ in *append* ~ *suspend*) and when that was not possible, the postvocalic C shared the same place or manner of articulation (e.g., *repine* ~ *despite*). The stimuli were uttered in the carrier phrase “I saw [stimulus word] on the sign.” and produced with main lexical stress on the vowel following the cluster, in case of a VCCV, or the single medial consonant, in case of a VCV, across speakers (i.e., iambic stress). There was one exception to this stress

¹ All stimuli are existing English words with the exception of the two nonce forms in the /sk/ high vowel context (*gakeep* and *gaskeep*), as well as the nonce form in the /sp/ high vowel *aspeed*, where appropriate existing forms could not be found.

² The VCIV sequence was also instantiated in the mid vowel context with the stimulus word *apoplectic*. However, our participants made several errors in producing this word which was therefore not included in any subsequent analyses.

pattern. The stimulus word *cascade* was found to be produced with main lexical stress on the vowel preceding the cluster (i.e., trochaic stress) consistently across speakers. We return to this difference in stress patterns in Sections 3.7 and 4. Table 1 presents the pairs of VCV and VCCV words across clusters and vowel contexts.

Table 1: Stimuli. Bold face indicates lexical stress in participants’ productions.

Cluster	VCV	VCCV	Vowel
/pl/	delete	replete	high /i/
	collide	applied	low /a/
/kl/	ellipse	eclipse	high /i/
	delight	decline	low /a/
	repeat	aspeed	high /i/
/sp/	append	suspend	mid /e/
	repine	despite	low /a/
	gake ep	gas ke ep	high /i/
/sk/	blockade	c ascade	mid /e/
	recover	discover	low (and back) /ʌ/

2.3 Procedure

Participants were seated in a booth (with interior dimensions in cm of 300 (*l*) × 240 (*w*) × 205 (*h*) and sound insulation with specific absorptive and diffusive wall panels, non-parallel walls and glass surfaces) in front of a computer monitor and were instructed to read the sentences presented on the monitor at a comfortable rate. The stimuli were embedded in a carrier phrase (see Section 2.2) and presented in a randomized order. There were eight blocks with 2 min break after each block, unless the participant wanted to continue without a break. For stop-lateral clusters, a total of 448 tokens were recorded (two syllable sizes × two clusters × two vowels × eight repetitions × seven speakers). For s-stop clusters, a total of 672 tokens were recorded (two syllable sizes × two clusters × three vowels × eight repetitions × seven speakers). Prior to the actual data recording, there were two short familiarization phases, one before and one after gluing the sensors. Before gluing the sensors, the participants were briefed and asked to read aloud from a written list some of the stimuli sentences. After the sensors were glued, the participants were asked to read again aloud some of the sentences so that they get familiarized to the sensors. In cases where during the recording there was a

mispronunciation of a word or any hesitation in the production of the sentence, the participants were instructed to repeat that sentence. The actual data recording lasted about 2 h, while the preparational phase including briefing, familiarization and gluing of the sensors lasted about 1 h.

2.4 Data segmentation

The articulatory segmentation of our EMA data was achieved by means of the Matlab-based mview software, developed by Mark Tiede at Haskins Laboratories. For each VCV and VCCV sequence, the consonantal gestures produced with a tongue constriction were measured using the sensor attached to the corresponding primary articulator involved in forming that constriction: the TT sensor for the coronal consonants /s, t, d, n, l/ and the TB sensor for the velar consonant /k/. Consonants produced with a labial closure such as /p, v/ were measured using the lip aperture (henceforth, LA). LA is a derived signal calculated as the Euclidean distance between the upper (henceforth, UL) and lower (henceforth, LL) lip sensors (which constitute the primary articulators for /p/ constriction). Lastly, the vowels /ʌ, i, e, ɛ, a/ were measured using the TB sensor. Using the mview software, the landmarks of both consonantal and vowel gestures are identified on the basis of the tangential velocity. Landmark identification for the vowel gestures was based on the vertical velocity signal in some exceptional cases where use of the tangential velocity did not provide a reliable parse, as for example in cases where the landmarks delineating the plateau of the vowel did not correspond to the acoustic signal (18% of the data). The process of landmark identification is semi-automatic, meaning that the gestures are first manually specified in the zoomed-in temporal range by the user. The algorithm first finds the peak velocities (to and from the constriction) and the minimum velocity within the user-specified temporal range. The achievement of target and the constriction release landmarks were then obtained by identifying the timestamp at which velocity falls below and rises above a 20% threshold of the local tangential velocity peaks. The gestural onset (onset) and gestural offset (offset) were obtained by identifying the timestamp at which velocity rises above and falls below a 20% threshold of the local tangential velocity peaks. The interval between the target and the release landmarks indicates the constriction plateau. Figure 1 illustrates an example parse of the velar gesture of the stop in the /kl/ cluster of the word *eclipse* using the TB sensor. The panels from top to bottom display the TB movement trajectory in the vertical (green line) and horizontal (yellow line) components and its corresponding tangential velocity. The vertical lines from left to right indicate the gestural landmarks: gestural onset (Onset), peak velocity towards the constriction (Peak Velocity to), the achievement

Figure 1: Identification of velar gesture in the word *eclipse* using mview. The top panel shows the tongue-back (TB) movement trajectory in the superior-inferior (green line) and anterior-posterior (yellow line) dimensions. The bottom panel shows the tangential velocity of the TB sensor. The vertical lines from left to right indicate the gestural landmarks: gestural onset (Onset), peak velocity towards the constriction (Peak Velocity to), the achievement of constriction (Target), maximum constriction (Min. Velocity), constriction release (Release), peak velocity away from the constriction (Peak Velocity fro) and gestural offset (Offset).

of constriction (Target), maximum constriction (Min. Velocity), constriction release (Release), peak velocity away from the constriction (Peak Velocity fro) and gestural offset (Offset), generated automatically by the algorithm based on the tangential velocity signal.

2.5 Data analysis

In the context of past work on evidence for syllabic organization in phonetic data which tends to consider one or two (presumed) privileged measures, we quantified several different spatiotemporal measures: specifically, duration of the prevocalic consonant across word-medial CV and CCV, relative timing of the prevocalic consonant with the vowel across word-medial CV and CCV, relation between inter-gestural timing in the CC cluster (expressed by the inter-plateau interval between the two main oral C constrictions) and consonant duration, and finally vowel initiation with respect to the preceding segment or segmental combination. Following Sotiropoulou et al. (2020), we argue that joint consideration of these measures enables one to see patterns of how different measures relate to one another and provides crucial evidence for global organization of the segments that are part of the hypothesized syllabic units. This evidence would not otherwise (i.e., when looking at isolated measures) be discernible.

2.5.1 Inter-plateau interval timing and consonant duration

Our first quantification concerns the temporal overlap between the two consonants C1 and C2 in a VCCV sequence. We quantified C1-C2 overlap in VCCV sequences by employing a frequently used measure (e.g., Gafos et al. 2010; Pouplier 2012; Sotiropoulou et al. 2020), the inter-plateau interval (hereafter, IPI). IPI is the temporal lag between the release of the first consonant (C1) and the target of the second or prevocalic consonant (C2) in milliseconds, C2 target – C1 release. Figure 2 illustrates an example of the IPI and consonant plateau duration for C1 and C2 of the /pl/ cluster in the word *replete*.

Further, to control for inter-speaker speech variability (Bombien 2011; Sotiropoulou et al. 2020), we calculated the normalized IPI by dividing the raw IPI measure by the cluster's total constriction duration $((C2 \text{ target} - C1 \text{ release}) / (C2 \text{ release} - C1 \text{ target}))$. Positive IPI values indicate that the target of the second consonant (C2) begins after the release phase of the gesture of the first consonant (C1), meaning that the two consonants show no or low overlap. Negative IPI values indicate that the target of the second consonant (C2) begins before or at the release of the first consonant (C1), meaning that the two consonants show high overlap. For any given consonant, its plateau duration was computed as the interval between the consonant's release and target landmarks (see Figure 2). Additionally, normalized plateau durations were calculated by dividing the consonant's plateau duration by the duration of the interval spanning from the target landmark of the first to the release landmark of the second consonant in the cluster.

In the present work, we also examine the relation between IPI and the prevocalic consonants C1, C2 in a C1C2 cluster and specifically how the duration of the prevocalic consonants responds to changes in duration of the IPI in a cluster. In a recent study of stop-lateral clusters in Spanish and Moroccan Arabic, Gafos et al.

Figure 2: Inter-plateau interval of the C1C2 cluster in the word *replete*. The top panel presents the lips' movement in the derived lip aperture (LA) signal. The bottom panel shows the TT movement trajectory in horizontal (solid line) and vertical (dashed line) dimensions (x, y). The intervals between the vertical lines on the LA and TT trajectories illustrate the plateau durations of C1 and C2 respectively. The interval between C1 release and C2 target is the inter-plateau interval (IPI).

(2020) find that a long lag in terms of IPI between the /k/ and the lateral is compensated in Spanish by shortening of the lateral, but that no such compensatory relation is found in Moroccan Arabic. Gafos et al. (2020) express this by saying that segments are appended linearly, one after another, in Arabic, but in Spanish adding a segment (/k/) to a sequence (/la/) or lengthening the /k/ in /kla/ has consequences for the spatiotemporal properties of the rest of the segments in the same syllable: increasing the lag between the constrictions of the two prevocalic consonants (those of the stop and the lateral) implies shortening of the second consonant (the lateral) in Spanish (see also Sotiropoulou et al. 2020). We wish to assess whether a compensatory relation of the sort described in Sotiropoulou et al. (2020) and Gafos et al. (2020) for Spanish stop-laterals holds true for word medial stop-lateral and s-stop clusters in English.

2.5.2 Stability-based analysis

Previous work on the relation between syllabic organization and articulatory data has extensively used the quantification of the stability of different intervals across CV ~ CCV pairs. Each of these intervals is delimited by left and right landmarks: the right-delimiting landmarks (hereafter, anchors) have been used as a reference point in the word and are found either on the vowel or the postvocalic consonant (e.g., the consonant's target), and the left-delimiting landmarks are found on the prevocalic consonants. As used in several studies (e.g., Browman and Goldstein 1988; Goldstein et al. 2009; Shaw et al. 2011), the two intervals left and right delimited by these landmarks were the c-center to anchor interval, spanning from the temporal midpoint of the consonant(s) to the anchor point, and the right-edge to anchor interval, spanning from the prevocalic consonant's constriction release to the anchor. In our stability analyses, we make use of two such intervals, the global and local timing interval respectively (Gafos et al. 2020). Global timing refers to the interval that stretches over the entire sequence (C)CV sequence, from the left-delimiting c-center landmark of the prevocalic consonant(s) up to a right-delimiting anchor point landmark (to be defined below). Local timing refers to the interval that stretches over the inner CV sequence to the exclusion of any preceding consonant(s), from the left-delimited release landmark of the prevocalic consonant up to an anchor point. To evaluate the robustness of our results, we used three different anchor points: two consonantal anchors such as the target of the constriction (Ctar) and the maximum constriction (Cmax) of the postvocalic C; one vocalic anchor such as the spatial extremum of the vowel (Vmax). Figure 3 schematizes the global and local timing intervals as defined above.

A descriptive characterization of our data consists in examining the relative standard deviation (RSD) of each interval calculated as the standard deviation of

Figure 3: Timing intervals in CCV (adapted from Sotiropoulou et al. 2020).

the interval divided by its mean duration and then multiplied by 100 to express the RSD value as a percentage (Shaw et al. 2009). The interval with the lowest RSD value, if a lowest exists, indicates the most stable interval for any given word pair. For instance, for a given VCV ~ VCCV word pair, if the RSD value of the global timing interval is lower compared to the RSD value of the local timing interval, the global timing interval is more stable or shows less change (in comparison to the local timing interval) as the number of consonants increases from VCV to VCCV. This would correspond to the expected stability pattern for cluster-vowel sequences parsed as one syllable.

2.5.3 Consonant-vowel timing

Consonant-vowel timing refers to the timing between the prevocalic consonant and the vowel. Relative timing is indexed by the lag between the release of the lateral and the maximum opening of the following vowel (CV lag) in the /IV/ and stop-/IV/ contexts, and between the release of the stop and the maximum opening of the vowel in the /CV/ and sibilant s-/CV/ contexts (Goldstein et al. 2009; Marin 2013). Figure 4 illustrates the CV lag measure as defined above in the word *replete*.

Figure 4: Consonant-vowel lag (CV lag) in the word *replete*. The top panel presents the tongue TB movement trajectory in horizontal (solid line) and vertical (dashed line) dimensions (x, y). The second panel shows the TT movement trajectory in horizontal (solid line) and vertical (dashed line) dimensions (x, y). The CV lag spans from the release of the consonant (C release) to the maximum opening of the vowel (Vmax).

We use this measure to examine if and how the relative timing of the prevocalic consonant and the vowel changes across different contexts, as in VCV versus VCCV.

2.5.4 Vowel initiation

Previous studies on the link between syllabic organization and the spatiotemporal dimension of articulation have not quantified the vowel gesture explicitly. Vowel-specific articulatory movements tend to be slower and longer than those of consonants and (thus) often lack clearly demarcated velocity peaks (as compared to consonants). This kinematic characteristic of vowels makes it hard for automatic algorithms to provide reliable parses of vowel gestures. Thus, previous studies have used indirect measures, such as consonant-to-consonant lag of the two non-adjacent consonants surrounding the vowel in addressing consonant-vowel timing or vowel duration in CVC and CCVC sequences (Browman and Goldstein 1988; Honorof and Browman 1995; Marin 2013; Marin and Pouplier 2010; among others).

Our present work follows Sotiropoulou et al. (2020) in seeking to quantify the timing of vowel gesture so as to gain an additional window into possible effects of syllabic organization on articulation. In particular, we study vowel initiation patterns in word-medial stop-lateral and s-stop clusters by calculating a measure we refer to as vowel onset lag. That is the interval between the c-center of the consonantal sequences (in the VC1C2V) and the gestural onset of the vowel. Figure 5 presents an example of vowel initiation with respect to the prevocalic C1C2 /pl/ cluster of the word *replete*.

Figure 5: Vowel initiation with respect to the prevocalic consonants in the word *replete*. The top panel presents the lips' movement in the derived LA signal. The second panel shows the TT movement trajectory in horizontal (solid line) and vertical (dashed line) dimensions (x , y) and the last panel the TB movement trajectory in horizontal (solid line) and vertical (dashed line) dimensions (x , y). The dark filled boxes indicate plateaus of the initial consonant (C1), the second consonant (C2) and the vowel (V). In the bottom panel, in particular, vowel initiation is indicated by the V onset landmark on the TB movement trajectory.

The vowel was parsed by using the tongue back sensor based on the tangential velocity of the signal. Given the difficulty with the gestural quantification of the vowels, we had to exclude cases (15% of the stop-lateral subset; 17% of the s-stop subset) where the algorithm either failed to identify the vowel gestural landmarks or the identified landmarks did not correspond to the acoustic signal.

2.5.5 Statistical analysis

Data were analyzed in R Studio v. 1.2.1335 (RStudio Team 2015). For a statistical evaluation of the results, linear mixed-effects models were fit using the package *lme4* v. 1.1-23 (Bates et al. 2014) with subjects and items used as random factors. The first step before fitting our data into a model was to generate histograms to examine the distribution of the data. In cases where the data were slightly skewed, we log-transformed them to approach a normal distribution. To select the best model for our data, for each analysis, we compared the ratios of the likelihoods of the models of interest (by subject and by item: varying intercepts; varying intercepts, and slopes; varying intercepts, and slopes with correlation). In case the ratio was greater than the critical chi-square value, for a given degree of freedom, the indication was that the compared models did not have the same log-likelihoods, meaning that the maximal model was better. Conversely, if there was no difference among the compared models, we chose the simpler model (Bates et al. 2018). After the model comparison, we performed a likelihood ratio test (ANOVA) on the selected model to examine the significance of predictor(s) and interactions between predictors. Post hoc pair-wise comparisons were conducted using the package *emmeans* v. 1.5.1 (Lenth et al. 2020). For multiple comparisons, the Tukey method was used for *p*-value adjustments. Separate models were fitted to the stop-lateral subset and the s-stop subset of the data. Due to technical problems with the recordings or due to failure of the algorithm to provide a reliable parse of articulatory gestures for some cases, we had to exclude for stop-laterals 18% of the data (79 out of 448) and for s-stop clusters, 22% of the data (147 out of 672). Analyses were conducted both on raw and normalized data, but we report results only on the raw data for reasons of space and because results of raw and normalized data show the same patterns. Tables of values derived from the raw EMA data and scripts for the analyses conducted on the basis of these values are available at <https://osf.io/2xahs/> on the OSF platform.

3 Spatiotemporal properties of word-medial clusters

We now come to report our results using the above multiple measures. First, we document the more basic properties of the clusters under investigation and then we report our interval stability-based results. We thus depart from the practice of limiting attention to some presumed privileged measures of assessing syllabic organization (such as interval stability) by including results on more basic properties such as inter-plateau interval durations as well as individual consonant plateau durations. This is because, as we will argue in detail in what follows, understanding more basic properties of the clusters and their dependence on the following vowel enables an understanding of the puzzling interval stability patterns we and others have observed and eventually leads to a different (more inclusive, in terms of phonetic indices) understanding of how syllabic organization is expressed in articulatory data.

3.1 Inter-plateau interval in stop-lateral clusters

In this subsection, we examine how the inter-plateau interval (IPI) changes as a function of C1 place of articulation and vowel context in stop-lateral clusters.³ Recall that the inter-plateau interval (IPI) in a C1C2V sequence is defined as the lag between the release of the initial consonant C1 and the target of the second consonant C2, that is, C2 target – C1 release, in ms. For the statistical analysis, linear mixed effects models with by subject and by item intercept variation were fitted to stop-laterals ($N = 179$) with IPI (raw in milliseconds) as dependent variable, C1 place of articulation (velar vs. labial) and vowel (high vs. low) as fixed effects. The results indicated that there is a significant effect of both C1 place of articulation and vowel context on IPI ($\chi^2(1) = 6.2, p = 0.01$). A post-hoc test showed that there is a significant difference in the IPI between /pl/ and /kl/ clusters across vowel contexts, with IPI being shorter in /pl/ clusters than in /kl/ clusters (estimate = $-7.3, p = 0.02$). Further, there is a significant difference in the IPI between low and high vowel contexts across clusters with the low vowel showing shorter IPI compared to the high vowel context (estimate = $-7.7, p = 0.01$). Mean and standard deviation for raw (in ms) and normalized IPI values across /Cl/ clusters and vowel contexts are

³ Apart from the reported IPI analysis in this section, analysis of C1 duration was also carried out. The results indicated that C1 duration changes as a function of vowel context. Specifically, stop C1 duration is found to be longer when the subsequent vowel is low compared to when it is high ($p < 0.0001$).

Table 2: Mean and standard deviation of IPI values (raw in ms and normalized) across word-medial /C1/ clusters as a function of C1 place of articulation and vowel contexts.

	Inter-plateau interval IPI (ms)		Normalized IPI	
	Mean	sd	Mean	sd
C1 velar	68.1	29.8	0.49	0.16
high	74.4	28.3	0.55	0.12
low	61.7	30.3	0.43	0.19
C1 labial	59.3	20.6	0.47	0.15
high	62.8	22.6	0.56	0.13
low	57.5	19.2	0.42	0.13

presented in Table 2. Figure 6 illustrates IPI values (in ms) as a function of C1 place of articulation and vowel context.

Overall, the IPI is larger in /k1/ clusters compared to /p1/ clusters across vowel contexts, and further, within each cluster, the IPI duration differs as a function of vowel context with the IPI being larger when the subsequent vowel is high as opposed to when it is low. These results are in line with Klatt (1975) who reports that voice onset time (hereafter, VOT) is longer for velars than labials and before high than non-high vowels.

Figure 6: IPI values in word-medial /C1/ clusters as a function of C1 place of articulation (x-axis) and vowel context.

3.2 C2 plateau duration in stop-lateral clusters

In this subsection, we present results on C2 plateau duration. First, we examine C2 plateau duration as a function of C1 place of articulation and vowel context. Plateau duration was calculated as the interval between C release and C target: C plateau = C release – C target. For the statistical analysis, linear mixed-effects models with by subject and by item intercept variation were fitted to stop-laterals ($N = 179$), with C2 plateau duration (raw in milliseconds) as dependent variable, log-transformed to reduce skewness, and as fixed effects, C1 place of articulation (velar vs. labial) and vowel (high vs. low). The results revealed that only vowel has a significant main effect ($\chi^2(1) = 38.9, p < 0.0001$). A post-hoc test showed that across /Cl/ clusters C2 duration is longer when the subsequent vowel is low compared to when it is high (estimate = 0.4, $p < 0.0001$). Table 3 presents the mean and standard deviation values of C2 plateau duration across /Cl/ clusters and vowel contexts. Figure 7 depicts C2 plateau duration (in ms) in word-medial /Cl/ clusters as a function of C1 place of articulation and vowel context.

Overall, across /Cl/ clusters, C2 duration does not differ as a function of C1 place of articulation, but it does differ as a function of vowel context; C2 duration is longer in the low compared to the high vowel context.

Table 3: Mean and standard deviation for raw (in ms) and normalized C2 duration in word-medial /Cl/ clusters as a function of C1 place of articulation and vowel context.

	Duration C2 (ms)		Normalized C2	
	Mean	sd	Mean	sd
C1 velar	31.3	19.3	0.23	0.12
high	24.0	10.4	0.19	0.08
low	38.7	23.2	0.27	0.14
C1 labial	33.6	18.9	0.25	0.09
high	24.4	12.1	0.22	0.08
low	38.6	22.0	0.27	0.09

In what follows, we examine prevocalic lateral duration in CV, CCV word-medially as opposed to just CCV above. It has been reported that consonants shorten when in (tautosyllabic) clusters than when not clustered (single consonant) (Haggard 1973). Hence, we examine whether C2 duration changes when a new segment is added to the left of the lateral-vowel sequence word-medially (from VCV to VCIV). To assess that, linear mixed-effects models with by subject and by item intercept variation were fitted with dependent variable the log-transformed C2 plateau duration ($N = 369$). As fixed effects, we used cluster size (VCV vs. VCCV), C1

Figure 7: C2 plateau duration (in ms) in word-medial /Cl/ clusters as a function of C1 place of articulation (x-axis) and vowel context.

place of articulation (velar vs. labial), and vowel (high vs. low). The results showed that there is a significant two-way interaction between cluster size and vowel context ($\chi^2(1) = 7.0, p = 0.008$). To specify the exact effects, we conducted a post-hoc test for multiple comparisons. The post-hoc test indicated that across /Cl/ cluster, C2 duration decreases significantly from VIV to VplV across both vowel contexts with the high vowel context exhibiting greater C2 compression (low V: estimate = 0.2, $p = 0.01$; high V: estimate = 0.4, $p < 0.0001$). Mean and standard deviation values for C2 plateau duration from VCV to VCIV are presented in Table 4. Mean C2 plateau duration from VCV to VCCV as a function of C1 place of articulation and vowel context is illustrated in Figure 8.

Overall, C2 duration decreases from VCV to VCIV to a greater extent when the subsequent vowel is high than when it is low across both /pl/ and /kl/ clusters.

Table 4: Mean and standard deviation for C2 duration (in ms) across VCV and VCIV contexts.

C2 duration (ms)	VCV		VCCV	
	Mean	sd	Mean	sd
C1 velar	38.4	23.1	31.3	19.3
high	36.7	24.4	24.0	10.4
low	40.3	21.7	38.7	23.2
C1 labial	43.1	25.9	33.6	18.9
high	40.6	29.3	24.4	12.1
low	45.2	22.2	38.6	22.0

Figure 8: Mean C2 plateau duration from VCV to VCIV as a function of C1 place of articulation (x-axis) and vowel context.

3.3 Compensatory effects in stop-lateral clusters

Next, we examine the relation between two phonetic parameters, IPI (inter-plateau interval) and C2 lateral duration, in word-medial stop-lateral clusters. Specifically, we investigate how the duration of the lateral, the second consonant in the word-medial C1C2V sequence, responds to an increase in duration of the lag (IPI) between the two consonants. Our aim is to assess the presence of compensatory effects in the word-medial CCV string, that is, effects where spatiotemporal modification in one local region of the string comes systematically with a change in another region of the string. The presence of such effects would indicate that the different parts of CCV are not independently planned and produced and thus such effects, if present, offer evidence for global organization.

For this analysis, we report results using the normalized IPI and lateral duration values. This is so because when looking at the relation between two phonetic parameters the normalization of their values is important in order to control for inter-speaker and inter-cluster variability. The results indicate the presence of a negative correlation between the two phonetic parameters across both /pl/ and /kl/ clusters. Specifically, within /pl/ clusters, we find a moderate negative correlation in the high vowel context ($r(26) = -0.7, p < 0.0001$), and a weak negative correlation in the low vowel context ($r(50) = -0.5, p = 0.0001$). Within /kl/ clusters, we find a strong negative correlation in the low vowel context ($r(47) = -0.9, p < 0.0001$), and a weak negative correlation in the high vowel context ($r(48) = -0.5, p = 0.0004$). These relations indicate the presence of

compensatory effects within the sequence of segments examined here; specifically, as IPI increases, C2 duration decreases (with a robust negative correlation in /kl/ clusters in the low vowel context and a weak to moderate correlation in the rest of the sequences).

Figure 9 illustrates the relation between IPI and C2 duration across /Cl/ clusters and vowel contexts. We can summarize the above result by saying that when a segment or intersegmental interval in the sequence of segments examined here is locally perturbed (for instance, by lengthening of its duration), then other segments within the sequence show a corresponding readjustment to this perturbation (e.g., by shortening). This is one indication that the entire CCV sequence is organized globally (see also Sotiropoulou et al. 2020). If instead the segments in a CCV were planned independently of the other segments, then an increase or decrease in the duration of that segment is not predicted to result in a decrease or increase in the duration of the other (as has been observed in the case of Moroccan Arabic, e.g., Gafos et al. 2020).

Figure 9: The relation between normalized IPI and normalized C2 plateau duration for velar and labial stop-lateral word-medial clusters across high and low vowel contexts.

3.4 Sibilant plateau duration in s-stop clusters

In this subsection, we examine how sibilant C1 plateau duration changes as a function of C2 place of articulation and vowel context in s-stop clusters.⁴ For the statistical

⁴ Analysis of IPI duration was also conducted. Specifically, the IPI duration was found to be larger when the subsequent vowel is high compared to when it is low ($p = 0.0001$) or mid ($p = 0.008$).

analysis of C1 plateau duration, linear mixed-effects models with by subject and by item intercept variation were fitted to s-stops ($N = 232$) with C1 plateau duration (raw in milliseconds) as dependent variable, log-transformed to reduce skewness, C2 place of articulation (velar vs. labial) and vowel (high vs. mid vs. low) as fixed effects. The results indicated that only vowel context has a significant effect on C1 duration ($\chi^2(2) = 13.6$, $p = 0.0002$). A post-hoc test showed across s-stops C1 plateau duration is longer when the subsequent vowel is low compared to when it is mid (estimate = 0.25, $p = 0.004$) or high (estimate = 0.3, $p < 0.003$). Table 5 presents the mean and standard deviation values for raw and normalized C1 plateau duration across word-medial s-stop clusters as a function of C2 place of articulation and vowel contexts. Figure 10 depicts C1 plateau duration across word-medial s-stop clusters as a function of vowel context.

Table 5: Mean and standard deviation for raw (in ms) and normalized C1 duration across s-stop clusters as a function of C2 place of articulation and vowel contexts.

	Duration C1 (ms)		Normalized C1	
	Mean	sd	Mean	sd
C2 velar	51.6	26.7	0.47	0.2
low (and back)	54.6	36.5	0.50	0.3
mid	53.9	23.4	0.49	0.2
high	46.3	19.2	0.41	0.2
C2 labial	53.4	28.1	0.46	0.2
low	67.3	31.1	0.52	0.2
mid	42.9	22.9	0.43	0.3
high	47.8	20.5	0.40	0.2

Figure 10: C1 duration in word-medial s-stop clusters as a function of vowel context.

In summary, C1 plateau duration differs in s-stop clusters across vowel contexts: /s/ duration is longer when the subsequent vowel is low compared to mid and high vowel contexts. These results are in line with the Bailey’s and Summerfield (1980) findings on English word-medial s-stop clusters.

3.5 Stop C2 plateau duration in s-stop clusters

In this subsection, we examine stop C2 plateau duration as a function of C2 place of articulation and vowel context in s-stop clusters. For the statistical analysis, linear mixed-effects models with by subject and by item intercept variation were fitted to s-stops ($N = 236$) with C2 plateau duration (raw in milliseconds) as the dependent variable, log-transformed to reduce skewness, C2 place of articulation (velar vs. labial) and vowel (high vs. mid vs. low) as fixed effects. The results revealed that only C2 place of articulation has a significant main effect ($\chi^2(1) = 9.03, p = 0.002$). A post-hoc test showed that C2 duration is longer in /sp/ compared to /sk/ clusters (estimate = $-0.2, p = 0.003$). Table 6 presents the mean and standard deviation values of C2 plateau duration across s-stop clusters and vowel contexts. Figure 11 illustrates C2 plateau duration in word-medial s-stop clusters as a function of C2 place of articulation and vowel context.

In summary, C2 plateau duration is longer in /sp/ than in /sk/ clusters. In relation to the vowel context, there is no statistical difference among the three vowel contexts, but C2 duration tends to be shorter in low than in mid or high vowel contexts.

Table 6: Mean and standard deviation for raw (in ms) and normalized C2 duration in word-medial s-stop clusters as a function of C2 place of articulation and vowel context.

	Duration C2 (ms)		Normalized C2	
	Mean	sd	Mean	sd
C2 velar	29.2	17.9	0.26	0.1
low (and back)	26.3	8.6	0.26	0.1
mid	28.2	13.6	0.26	0.1
high	32.4	25.6	0.26	0.1
C2 labial	33.8	15.3	0.27	0.1
low	32.5	10.7	0.24	0.1
mid	33.6	16.3	0.31	0.1
high	35.2	17.9	0.27	0.1

Figure 11: C2 plateau duration (in ms) in word-medial s-stop clusters as a function of C2 place of articulation (x-axis) and vowel context.

To examine C2 duration across VCV ~ VsCV, linear mixed-effects models with by subject and by item intercept variation were fitted to s-stops ($N = 438$) with C2 plateau duration (raw in milliseconds) as dependent variable, log-transformed to reduce skewness, cluster size (VCV vs. VCCV), C2 place of articulation (velar vs. labial) and vowel (high vs. mid vs. low) as fixed effects. The results revealed a significant three-way interaction among the factors ($\chi^2(7) = 24.2, p = 0.001$). To identify the exact effects, we carried out a post-hoc test for multiple comparisons. The results indicated that in /sp/ clusters, C2 duration decreases significantly from VpV to VspV across all vowel contexts (high V: estimate = 0.3, $p = 0.04$; low V: estimate = 0.3, $p = 0.006$; mid V: estimate = 0.5, $p < 0.0001$). In /sk/ clusters, C2 duration decreases significantly from VkV to VskV in high and low (and back) vowel contexts (high V: estimate = 0.3, $p = 0.009$; low V: estimate = 0.5, $p < 0.0001$), yet C2 duration does not change in the mid vowel context (estimate = 0.06, $p = 0.5$). Mean and standard deviation values for C2 plateau duration (in ms) from VCV to VsCV are shown in Table 7. Figure 12 illustrates mean C2 plateau duration from VCV to VsCV as a function of C2 place of articulation and vowel context.

In summary, C2 duration compresses from VCV to VsCV across all vowel contexts, with the low and the mid vowel context exhibiting the greatest compression compared to the high vowel context. These results confirm Bailey's and Summerfield (1980) findings with respect to stop duration in English word-medial s-stop clusters. Yet, C2 duration does not shorten from VCV to VsCV in /sk/ clusters when the subsequent vowel is mid.

Table 7: Mean and standard deviation for C2 duration (in ms) across VCV and VsCV contexts.

C2 duration (ms)	VCV		VCCV	
	Mean	sd	Mean	sd
C2 velar				
low (and back)	47.8	23.3	26.3	8.6
mid	31.3	20.6	28.2	13.6
high	42.2	29.7	32.4	25.6
C2 labial				
low	45.5	5.8	32.5	10.7
mid	51.1	15.9	33.6	16.3
high	43.9	3.6	35.1	17.9

Figure 12: Mean C2 plateau duration from VCV to VsCV as a function of C2 place of articulation (x-axis) and vowel context.

3.6 Compensatory relations in s-stop clusters

Here, we assess the presence of a compensatory relation between IPI and the duration of the consonants in word-medial s-stop clustes. IPI shows no relation to the duration of the stop (C2) in our s-stop sequences across clusters and vowel contexts ($r(293) = -0.3$, $p < 0.0001$). However, IPI and the duration of the sibilant (C1) across s-stop clusters and vowel contexts show a strong negative correlation, such that as IPI duration increases, C1 duration decreases $r(293) = -0.8$, $p < 0.0001$). Figure 13 illustrates the relation between IPI and C1 duration across s-stop clusters and vowel contexts.

Overall, we may not find a compensatory relation between IPI and C2 stop duration, yet it is the sibilant that compensates its duration as a response to the IPI lengthening, providing evidence in favor of a global organization. If each segment in the sCV were planned independently of the other segments, then an increase or decrease in the duration of that segment is not predicted to result in a decrease or increase in the duration of the other.

Figure 13: The relation between normalized IPI and normalized C1 plateau duration for s-velar and s-labial word-medial clusters across low, mid, and high vowel contexts.

3.7 Interim summary

To summarize, our results are in line with previous findings on complex onsets. Let us consider first our results concerning the duration of clustered consonants. C2 plateau duration is found to be shorter compared to its singleton pair across both stop-laterals and s-stop clusters as per previous reports on English (Haggard 1973). Further, for both stop-lateral and s-stop clusters, the vowel context affects the spatiotemporal properties of the clusters as well. Particularly, our results on s-stop clusters show that the prevocalic stop duration is shorter when the subsequent vowel is low compared to when it is mid and when it is high. These results are also in line with findings in Bailey and Summerfield (1980) on English word-medial s-stop clusters where an inverse correlation between the openness of the vowel and the duration of the prevocalic stop consonant has been reported (that is, the more open the vowel, the shorter the prevocalic stop duration).

Consider next our results concerning intersegmental measures between the consonants in the clusters. Intersegmental timing was indexed by the inter-plateau interval (IPI) measure. We found that IPI is longer when C1 is velar than labial in

stop-laterals and when the vowel context is high compared to the rest vowel contexts across stop-laterals and s-stops. This dependence of IPI on place of articulation as well as vowel height also finds clear precursors in past work on English. Thus, Klatt (1975) reports that voice onset time (hereafter, VOT) is longer for velars than labials and before high than non-high vowels.

Despite the differences in the spatiotemporal properties of the sequences under investigation, an overarching generalization in our results is the presence of compensatory relations among phonetic properties of the segments that partake in these disparate sequences. Specifically, in stop-laterals we found that the relation between the duration of the prevocalic lateral and the duration of the inter-plateau interval (IPI) of the cluster is characterized by an inverse correlation, i.e., the longer the IPI, the shorter the prevocalic lateral duration. For s-stop clusters, our results unveiled a relation between the duration of the sibilant (C1) and the IPI of the cluster. An inverse correlation characterizes this relation too; the longer the IPI, the shorter the sibilant duration. If all segments or intersegmental transitions in a C1C2V are planned as a group (globally), such compensatory relations are expected (a lengthening of one segment in the whole is compensated by a shortening of another unit in that whole). If instead each unit in C1C2V was planned independently of the others (that is, not global, but local organization), then no compensatory interactions are expected; an increase/decrease in one segment's duration is not predicted to result in a change in the duration of the others. The specific ways in which the compensatory relations are actuated are cluster-dependent: the compensatory relation concerns C2 and IPI for stop-lateral clusters, but C1 and IPI for s-stop clusters. What transpires across different cases is the presence of the compensation effect, not its specific manifestation in terms of the parochial (cluster- and language-specific) parameters that enter into its expression.⁵

One word in our stimuli, *cascade*, requires special attention. For this word, the duration of the prevocalic stop (unlike for the other words) neither gets shorter, nor changes its relative timing with the subsequent vowel from VCV to VCCV as we will report in the following section. In the context of work assessing syllabic structure with articulatory data, these results for this word show patterns in line with a heterosyllabic parse and are in line with studies in other languages where such

⁵ The reason why it is C1 in s-stop clusters but C2 in stop-lateral clusters that enter into a compensatory relation finds parallels in previous results on the manner-dependence of segmental durational modulations in English and elsewhere. Thus, in the prosodic strengthening literature, stops are reported not to be subject to prosody-driven effects to the same degree as their fricative counterparts (Berkovits 1993a, 1993b; Oller 1973; and see especially Klatt 1974, p. 62 on this differential capacity for stretchability between fricatives and stops). Discerning a compensatory relation between C1 or C2 duration and IPI requires that there is sufficiently variability in the former parameter. If stops are less variable in duration than fricatives (here, sibilants), then it is not surprising that our results reveal the presence of such a relation between /s/ and IPI in s-stop clusters but not between the stop and IPI in stop-lateral clusters.

clusters have been investigated (e.g., Shaw et al. 2009, 2011). We return to this finding in Section 4 where we argue that the different patterns that we observe for this word seem to reflect the word-specific prosodic structure which is imposed by the different lexical stress that was found to characterize its production.

3.8 Consonant-vowel timing in stop-lateral clusters

In this subsection, we ask how the relative timing of the lateral with the vowel changes from /IV/ to stop-/IV/. Recall that relative timing is indexed by the lag between the release of the lateral and the maximum opening of the following vowel in the /IV/ and stop-/IV/ contexts (Goldstein et al. 2009; Marin 2013). For the statistical analysis, we fitted linear mixed-effects models with by subject and by item intercept variation to the stop-laterals ($N = 344$) with lag duration (in milliseconds) as a dependent variable and as fixed effects, cluster size (VCV vs. VCCV), C1 place of articulation (velar vs. labial), and vowel (high vs. low). The results showed that all factors have a significant effect on the lag duration and that there is a significant two-way interaction between C1 place of articulation and vowel context ($\chi^2(4) = 13.59, p = 0.008$). A post-hoc test for multiple comparisons indicated that for both /Cl/ clusters, across vowel contexts, the CV lag decreases from VCV to VCIV (estimate = 18.6, $p < 0.0001$). For /pl/ clusters CV lag decreases more compared to /kl/ clusters when the subsequent vowel is high (estimate = 28.7, $p < 0.0001$), but there is no difference regarding the extent of the CV lag decrease between cluster groups when the subsequent vowel is low (estimate = 8.2, $p = 0.1$). Lastly, across both /Cl/ clusters the CV lag decreases more when the subsequent vowel is high (estimate = 25.3, $p < 0.0001$) compared to when it is low (estimate = 11.9, $p = 0.04$). Mean and standard deviation values for CV lag (in ms) from VCV to VCCV across /kl, pl/ clusters and vowel contexts are shown in Table 8. Figure 14 illustrates the CV lag from VCV to VCIV across /Cl/ clusters as a function of vowel context.

Overall, the lag between the prevocalic lateral and the maximum opening of the vowel when a stop C is appended in front of the lateral-vowel sequence decreases significantly across both groups and vowel contexts. Between groups, the CV lag decreases more in /pl/ clusters than in /kl/ clusters and across both groups, the CV lag decreases more when the subsequent vowel is high compared to when it is low.⁶

⁶ See Katz (2012) for a similar result on word-initial clusters in American English using acoustic data (specifically, adding a C to a CV-initial word results in various degrees of vowel compression in the CCV-initial word).

Table 8: CV lag duration (in ms) from VCV to VCCV across cluster groups and vowel contexts.

CV lag (ms)	VCV		VCCV	
	Mean	<i>sd</i>	Mean	<i>sd</i>
C1 velar				
high	90.4	39.9	60.0	36.1
low	74.6	19.7	63.0	25.4
C1 labial				
high	113.2	39.6	93.2	40.9
low	88.2	39.7	67.6	28.0

Figure 14: CV lag in VCV ~ VCCV across /Cl/ clusters as a function of vowel context (x-axis).

3.9 Vowel initiation in stop-lateral clusters

Here, we report results on vowel initiation in relation to the c-center of the stop-lateral clusters. To explore vowel initiation patterns in relation to the c-center of the clusters, we calculated the lag between the c-center and the gestural onset of the vowel of each cluster. Hereafter, we refer to this lag as vowel onset lag (in milliseconds). Considering that the results of the previous section showed that both C1 place of articulation and vowel context had a significant role on consonant-vowel timing relation, we examined vowel initiation as a function of C1 place of articulation and vowel context. Linear mixed-effects models were fitted with by subject and by item intercept variation. Vowel onset lag (raw in milliseconds) was used as a dependent variable ($N = 132$). C1 place of articulation

(velar vs. labial) and vowel context (high vs. low) served as fixed effects. The results indicated that only vowel as a factor has a significant effect on vowel onset lag ($\chi^2(1) = 4.7, p = 0.03$). Specifically, the vowel starts earlier when it is low compared to when it is high across both /kl/ and /pl/ clusters (estimate = 15.4, $p = 0.03$). Figure 15 illustrates mean vowel initiation in word-medial /Cl/ clusters across vowel contexts. The black vertical lines illustrate the intervals corresponding to the gestural plateau of the initial stop C delimited by the black dots which depict their plateau onset and offset timestamps. The blue vertical lines illustrate the intervals corresponding to the gestural plateaus of the prevocalic lateral /l/ delimited by the blue dots which indicate their plateau onset and offset timestamps. The black asterisks represent the c-center landmark of each consonantal sequence, and the red triangles denote the vowel gestural onset (plus-minus the SE of the mean) in each cluster. As can be seen in Figure 15, the vowel, when it is low (/aI/), starts 16 ms after the c-center, while when it is high (/i/), it starts 35 ms after the c-center.

Figure 15: Vowel initiation in relation to prevocalic consonants in stop-lateral clusters /kl, pl/, across vowel contexts (high, low) (x-axis). The black vertical lines illustrate the intervals corresponding to the gestural plateaus of C1 delimited by the black dots which denote their plateau onset and offset timestamps (y-axis). The blue vertical lines illustrate the intervals corresponding to the gestural plateaus of C2 delimited by the blue dots which indicate their plateau onset and offset timestamps (y-axis). The black asterisks indicate the c-center landmark of each consonantal sequence and the red triangles represent the vowel gestural onset (plus-minus the SE of the mean) in relation to the c-center of the two consonants in each consonantal cluster.

3.10 Consonant-vowel timing in s-stop clusters

In this subsection, we ask how the relative timing of the stop with the vowel changes from /CV/ to sibilant s-/CV/ in our data. For the statistical analysis of the consonant-vowel timing in s-stop clusters ($N = 494$), we fitted linear mixed-effects models with by subject and by item intercept variation, with lag duration (raw in milliseconds) as a dependent variable, and as fixed effects, cluster size (VCV vs. VCCV), C2 place of articulation (velar vs. labial), and vowel (high vs. mid vs. low). The results showed that there is a significant three-way interaction among cluster size, C2 place of articulation, and vowel context ($\chi^2(3) = 47.2$, $p < 0.0001$). To specify the exact effects, we carried out a post-hoc test for multiple comparisons. The results showed that for /sp/ clusters, the CV lag decreases from VCV to VspV across all vowel contexts (high V: estimate = 19.3, $p = 0.01$; low V: estimate = 56.8, $p < 0.0001$; mid V: estimate = 29.4, $p = 0.0001$). In similar fashion, in /sk/ clusters, from VCV to VskV, the CV lag decreases when the subsequent vowel is high or low (and back), but not when it is mid (high V: estimate = 17.1, $p = 0.05$; low V: estimate = 34.1, $p < 0.0001$; mid V: estimate = 10.2, $p = 0.2$). Mean and standard deviation values for CV lag (in ms) from VCV to VsCV across /sk, sp/ clusters and vowel contexts are shown in Table 9. Figure 16 illustrates the CV lag from VCV to VsCV across s-stop clusters as a function of vowel context.

Table 9: CV lag duration (in ms) from VCV to VsCV across cluster groups and vowel contexts.

CV lag (ms)	VCV		VCCV	
	Mean	sd	Mean	sd
C2 velar				
low (and back)	135.5	32.1	101.5	40.2
mid	104.7	33.4	94.2	27.7
high	117.1	50.3	100.8	42.2
C2 labial				
low	105.5	35.3	48.6	28.3
mid	105.0	36.5	75.9	27.3
high	109.2	38.4	90.3	33.1

Overall, the lag between the prevocalic stop and the maximum opening of the vowel when a sibilant /s/ is appended in front of the stop CV sequence decreases significantly – a pattern which is in line with global organization. An exception to

Figure 16: CV lag in VCV ~ VCCV across s-stop clusters as a function of vowel context (x-axis).

this pattern is the /sk/ cluster in the mid vowel context (*blockade ~ cascade*), where the lag between the prevocalic /k/ and the vowel does not change after the sibilant's addition to the CV sequence; a finding which is in line with local organization as we will discuss in Section 4.

3.11 Vowel initiation in s-stop clusters

In this subsection, we report results regarding the vowel initiation patterns in s-stop clusters. Recall that for the vowel initiation patterns we calculated the vowel onset lag which is the lag between the c-center and the gestural onset of the vowel of each cluster. Considering that the results of the previous subsection showed that both C2 place of articulation and vowel context had a significant role on the CV lag, we examined vowel initiation as a function of C2 place of articulation and vowel context. Linear mixed-effects models were fitted with by subject and by item intercept variation to s-stops ($N = 164$). Vowel onset lag (raw in milliseconds) was used as a dependent variable. C2 place of articulation (velar vs. labial) and vowel context (high vs. mid vs. low) served as fixed effects. The results indicated that both factors have a significant effect on vowel onset lag ($\chi^2(2) = 16.9, p = 0.0002$). Specifically, the results showed that the vowel starts earlier in /sp/ compared to /sk/ clusters across all vowel contexts. With respect to the vowel context, across s-stop clusters, the vowel is found to start earlier when this is high compared to when it is mid (estimate = 17.7, $p = 0.0002$), while there is no difference in the initiation pattern between high and low (estimate = 8.2, $p = 0.08$) or between low

and mid (estimate = 9.5, $p = 0.06$) vowel contexts. Figure 17 illustrates mean vowel initiation in word-medial s-stop clusters across vowel contexts. The black vertical lines illustrate the intervals corresponding to the gestural plateau of the initial sibilant /s/ consonant delimited by the black dots which depict their plateau onset and offset timestamps. The blue vertical lines illustrate the intervals corresponding to the gestural plateaus of the prevocalic stop /k/ or /p/ delimited by the blue dots which indicate their plateau onset and offset timestamps. The black asterisks represent the c-center landmark of each consonantal sequence, and the red triangles denote the vowel gestural onset (plus-minus the SE of the mean) in each cluster. As it can be seen in Figure 17, in /sp/ clusters, the vowel starts 9 ms after the c-center of the cluster when it is high (/i/), 18 ms after the c-center of the cluster when it is low (/a/), and 28 ms after the c-center of the cluster when it is mid (/e/). In /sk/ clusters, the vowel, when it is high (/i/) starts 40 ms after the c-center of the cluster, when it is low (and back, /a/) starts 45 ms after the c-center of the cluster, and when it is mid starts 55 ms after the c-center of the cluster.

Figure 17: Vowel initiation in relation to prevocalic consonants in s-stop sequences /sk, sp/ across vowel contexts (x-axis). The black vertical lines illustrate the intervals corresponding to the gestural plateaus of C1 delimited by the black dots which denote their plateau onset and offset timestamps (y-axis). The blue vertical lines illustrate the intervals corresponding to the gestural plateaus of C2 delimited by the blue dots which indicate their plateau onset and offset timestamps (y-axis). The black asterisks indicate the c-center landmark of each consonantal sequence, and the red triangles represent the vowel gestural onset (plus-minus the SE of the mean) in relation to the c-center of the two consonants in each cluster.

3.12 Stability analysis

In our final analysis, we present the results on the stability of timing intervals in our word-medial stop-lateral and s-stop clusters. The stability of timing intervals has been extensively used in previous work to diagnose syllabic organization (e.g., Browman and Goldstein 1988; Honorof and Browman 1995; Marin 2013; Pouplier 2012; among others).

The two intervals we used herein to assess the stability patterns are the global and local timing intervals (see Figure 3 in Section 2.5.2 for an illustrative example of the timing intervals). The global timing interval stretches over the entire sequence across word-medial CV ~ CCV, from the left-delimited c-center landmark of the prevocalic consonant(s) up to the anchor point. The local timing interval stretches over the inner CV sequence to the exclusion of any preceding consonant(s), from the left-delimited release landmark of the prevocalic consonant up to the anchor point. Further, we used three different anchor points; two consonantal: the target of the constriction (Ctar) and the maximum constriction (Cmax) of the postvocalic C, and one vocalic defined as the spatial extremum of the vowel (Vmax). Prior to the statistical analysis of our results, we conducted a descriptive characterization of the data, by computing the relative standard deviation (henceforth, RSD) for each interval across any given VCV ~ VCCV pair. The RSD is calculated by dividing the standard deviation of the interval by the mean duration and then multiply it by 100 (Shaw et al. 2009). The most stable interval is the one that exhibits the lowest RSD value. RSD values of the global and local timing intervals using two consonantal anchor landmarks (Ctar, Cmax) and one vocalic anchor landmark (Vmax) for each word pair (e.g., *append* ~ *suspend*) are provided in Appendix.

Concerning the stability-based heuristics of syllabic organization reviewed in Section 1, we expect English word-medial stop-lateral and s-stop clusters to exhibit global timing interval stability from VCV to VCCV. This is so because, in English, both cluster groups are considered tautosyllabic and thus exemplify the global organization case. According to the stability-based heuristics, global organization is assumed to be expressed in global timing interval stability. Nevertheless, by looking at the RSDs (see Appendix), we see that the results show inconsistent patterns. For example, across speakers, for stop-lateral clusters followed by a low vowel, the RSD of the local timing interval is the lowest. Yet, again across speakers, in stop-lateral clusters followed by a high vowel, the global timing interval is the one that exhibits the lowest RSD value. Hence, it is apparent that the RSD values are not always the lowest for the global timing interval as expected, but instead there is evidence for both global and local timing interval stability.

To better understand the patterns that emerged from the descriptive RSD analysis, we conducted a statistical examination, for each cluster group, of how the global and local timing intervals change as a function of cluster size, namely, as the number of the consonants increases from VCV to VCCV and if and how the timing intervals change as a function of vowel context across speakers.

3.12.1 Stop-lateral clusters

For each anchor, a series of linear mixed-effects models with by subjects and by items varying intercepts were fitted with interval duration as dependent variable ($N = 738$). The dependent variable was log-transformed to approximate a normal distribution. Interval type (global vs. local), cluster size (VCV vs. VCCV), C1 place of articulation (velar vs. labial), and vowel (high vs. low) were used as fixed effects. The results showed that all factors have a significant effect on interval duration and there is a significant interaction among them (Ctar: $\chi^2(11) = 497.4$, $p < 0.0001$; Cmax: $\chi^2(11) = 484.6$, $p < 0.0001$; Vmax: $\chi^2(11) = 468.4$, $p < 0.0001$). This means that the way the duration of each interval type changes as the number of consonants increases from VCV to VCCV depends on the C1 place of articulation (velar vs. labial) and the vowel context (high vs. low). Hence, a post-hoc test for multiple comparisons was carried out to specify the exact effects.

For /kl/ clusters, the post-hoc test showed that across Ctar and Cmax anchors, from VCV to VCCV in the low vowel context, the global timing interval increases significantly, while the local timing interval does not change significantly (Ctar: global estimate = -0.24 , $p < 0.0001$, local estimate = -0.07 , $p = 0.2$; Cmax: global estimate = -0.23 , $p < 0.0001$, local estimate = -0.07 , $p = 0.2$). Using the Vmax anchor, both timing intervals change significantly from VCV to VCCV, but the local timing interval changes less compared to the global timing interval (Vmax: global estimate = -86.6 , $p < 0.0001$; local estimate = -35.8 , $p < 0.0001$). Thus, the local timing interval is more stable compared to the global one in the low vowel context. Instead, in the high vowel context, across all anchors (Ctar, Cmax, Vmax) although the duration of both timing intervals changes from VIV to VklV, the local timing interval is the one that changes more compared to the global timing interval (Ctar: global estimate = -0.14 , $p = 0.0007$, local estimate = 0.16 , $p = 0.0001$; Cmax: global estimate = -0.11 , $p = 0.002$, local estimate = 0.14 , $p = 0.0001$; Vmax: global estimate = -16.02 , $p = 0.1$, local estimate = 29.8 , $p = 0.0004$). In other words, the global timing interval seems to remain more stable than the local timing interval when the subsequent vowel is high.

For /pl/ clusters, in the low vowel context, across all anchors (Ctar, Cmax, Vmax), from VIV to VplV, the duration of the global timing interval increases, while the duration of the local timing interval does not change (Ctar: global estimate = -0.15 , $p = 0.0001$, local estimate = 0.006 , $p = 0.9$; Cmax: global estimate = -0.15 , $p < 0.0001$, local estimate = -0.009 , $p = 0.9$; Vmax: global estimate = -37.3 , $p < 0.0001$, local estimate = 7.9 , $p = 0.7$). Thus, in the low vowel context, the local timing interval is the one that remains stable. Regarding the high vowel context, from VCV to VCCV and across all anchors (Ctar, Cmax, Vmax), the local timing interval decreases significantly, while the global timing interval does not change significantly (Ctar: global estimate = -0.05 , $p = 0.5$; local estimate = 0.1 , $p = 0.04$; Cmax: global estimate = -0.07 , $p = 0.2$; local estimate = 0.08 , $p = 0.05$; Vmax: global estimate = -11.6 , $p = 0.5$; local estimate = 24.1 , $p = 0.03$). This means that the global timing interval is the one that remains stable in the high vowel context. Figure 18 illustrates global and local timing intervals (using the Cmax as anchor point) for VCV and VCCV across /Cl/ clusters as a function of C1 place of articulation and vowel context (x-axis).

Overall, for both /kl/ and /pl/ clusters, global timing interval stability is found in the high vowel context. Instead, for the low vowel context, local timing interval stability is observed. As it is apparent, stability-based heuristics lead to unexpected or hard to interpret results as observed here, even for the same cluster in different vowel contexts.

Figure 18: Duration (in ms) of global and local timing intervals from VCV to VCCV for /kl, pl/ clusters as a function of vowel context (x-axis).

3.12.2 s-stop clusters

For each anchor, a series of linear mixed-effects models with by subjects and by items varying intercepts were fitted with interval duration as dependent variable ($N = 1214$). The dependent variable (interval duration) was log-transformed to approximate a normal distribution. Interval type (global vs. local), cluster size (VCV vs. VCCV), C2 place of articulation (velar vs. labial), and vowel (high vs. mid vs. low) were used as fixed effects. The results showed that all the factors have a significant effect on interval duration and there is a significant interaction among them (Ctar: $\chi^2(12) = 177.44$, $p < 0.0001$; Cmax: $\chi^2(12) = 189.3$, $p < 0.0001$; Vmax: $\chi^2(12) = 95.2$, $p < 0.0001$). This means that the way the duration of each interval type changes as the number of consonants increases from VCV to VCCV depends on the C2 place of articulation (velar vs. labial) and the vowel context (high vs. mid vs. low). Hence, a post-hoc test for multiple comparisons was carried out to specify the exact effects.

For /sk/ clusters, the post-hoc test showed that across all anchors (Ctar, Cmax, Vmax) from VCV to VCCV the global timing interval increases, and the local timing interval decreases in the high and mid vowel contexts, while in the low (and back) vowel context, both timing intervals decrease. Although the duration of both interval types changes from V_kV to V_skV, in low vowel context, across all anchors, the local timing interval changes significantly, while the global timing interval does not change significantly (Ctar: global estimate = 0.03, $p = 0.9$, local estimate = 0.2, $p < 0.0001$; Cmax: global estimate = 0.04, $p = 0.7$, local estimate = 0.2, $p < 0.0001$; Vmax: global estimate = -0.6, $p = 0.9$, local estimate = 30.3, $p = 0.0009$). This means that the global timing interval is the one that remains stable. On the contrary, in the high vowel context, across all anchors, although the duration of both timing intervals changes from VCV to VCCV, the global timing interval is the one that changes significantly (Ctar: global estimate = -0.1, $p = 0.03$, local estimate = 0.08, $p = 0.3$; Cmax: global estimate = -0.1, $p = 0.04$, local estimate = 0.07, $p = 0.4$; Vmax: global estimate = -29.8, $p = 0.003$, local estimate = 5.8, $p = 0.9$). In other words, the local timing interval is the one that remains stable. In similar fashion, in the mid vowel context, across all anchor points, the global timing interval changes significantly from VCV to VCCV but the local timing interval does not. Hence, the local timing interval is the one that remains stable (Ctar: global estimate = -0.1, $p = 0.05$, local estimate = 0.05, $p = 0.4$; Cmax: global estimate = -0.07, $p = 0.05$, local estimate = 0.05, $p = 0.5$; Vmax: global estimate = -28.3, $p = 0.001$, local estimate = 6.3, $p = 0.8$).

For /sp/ clusters, the post-hoc test showed that across all anchors (Ctar, Cmax, Vmax) from V_pV to V_spV the global timing interval increases, and the local timing interval decreases in the high vowel context, while in the low and mid vowel

contexts, both timing intervals decrease. Although the duration of both interval types changes from VCV to VCCV, in the low vowel context across all anchors, the local timing interval changes significantly, while the global timing interval does not change significantly (Ctar: global estimate = 0.07, $p = 0.1$, local estimate = 0.2, $p < 0.0001$; Cmax: global estimate = 0.06, $p = 0.2$, local estimate = 0.2, $p < 0.0001$; Vmax: global estimate = 15.1, $p = 0.1$, local estimate = 53.5, $p < 0.0001$). This means that the global timing interval is the one that remains stable for /sp/ clusters in the low vowel context. In similar fashion, in the mid vowel context, across all anchors, although the duration of both timing intervals changes from VCV to VCCV, the local timing interval is the one that changes significantly compared to the global timing interval (Ctar: global estimate = 0.02, $p = 0.9$, local estimate = 0.2, $p = 0.0009$; Cmax: global estimate = 0.04, $p = 0.9$, local estimate = 0.2, $p = 0.0004$; Vmax: global estimate = 7.2, $p = 0.8$, local estimate = 34.2, $p < 0.0001$). In other words, the global timing interval is the one that remains stable for /sp/ clusters in mid vowel context. Instead, in the high vowel context, using the Ctar and Cmax anchors, both timing intervals change from VpV to VspV, with the global timing interval changing more compared to the local timing interval, while using the Vmax anchor, only the global timing interval changes significantly from VpV to VspV as opposed to the local timing interval (Ctar: global estimate = -0.14, $p = 0.003$, local estimate = 0.1, $p = 0.05$; Cmax: global estimate = -0.13, $p = 0.005$, local estimate = 0.1, $p = 0.03$; Vmax: global estimate = -36.2, $p < 0.0001$, local estimate = 17.6, $p = 0.1$). Thus, the local timing interval is the one that remains stable. Figure 19 illustrates global and local timing intervals (using the Cmax as anchor point) for VCV and VCCV across s-stop clusters as a function of C2 place of articulation and vowel context (x-axis).

Figure 19: Duration (in ms) of global and local timing intervals from VCV to VCCV for /sk, sp/ clusters as a function of vowel context (x-axis).

Overall, for /sk/ clusters, global timing interval stability is found in the low (and back) vowel context. Instead, for the high and mid vowel contexts, local timing interval stability is observed. For /sp/ clusters, global timing interval stability is observed in the low and mid vowel contexts, while for the high vowel context, local timing interval stability emerges. Once again, as seen in the previous section, we find that using interval stabilities to diagnose syllabic organization results in inconclusive patterns. For some clusters and vowel contexts, stabilities seem to indicate global organization, whereas for others local organization is indicated.

4 Evidence for global organization in word-medial stop-lateral and s-stop clusters

Unlike previous research on articulatory correlates of syllabic structure which focuses on word-initial clusters, our work explores the spatiotemporal properties of English stop-lateral and s-stop word-medial clusters embedded in different vowel contexts using data from seven speakers and examining several spatiotemporal measures.

In our results, we find evidence for a global organization of the segments involved in these combinations. This is reflected in a number of ways: compression of the inner CV subsequence (in a word-medial CCV) in terms of shortening of the prevocalic consonant and reduction of CV timing (as indexed by the CV lag measure) in the word-medial cluster case compared to its singleton paired word in both stop-lateral and s-stop clusters, early vowel initiation (to the extent permitted by the cluster's phonetic properties), as indexed by the lag between vowel initiation and c-center of the cluster, and the presence of compensatory relations between phonetic properties of different segments or intersegmental transitions within each cluster.

These latter compensatory relations indicate that local perturbations (such as durational changes) on some segment or intersegmental interval in a cluster are compensated by readjustments in another part of the cluster, as would be expected if the entire cluster was planned and executed as a whole (global organization). Thus, we have seen that as the duration of the stop-to-lateral or sibilant-to-stop transition increases, the duration of the lateral in the stop-lateral cluster or the duration of the sibilant in the s-stop cluster must decrease to compensate for that increase. The specific ways in which the compensatory relations are actuated are cluster-dependent: the compensatory relation concerns C2 and IPI for stop-lateral clusters, but C1 and IPI for s-stop clusters. What transpires across different cases is

the presence of the compensation effect, not its specific manifestation in terms of the parochial (cluster- and language-specific) parameters that enter into its expression. If each segment in a C1C2V was planned and executed independently of the others, then no compensatory interactions are expected; an increase/decrease in one segment's duration is not predicted to result in a change in the rest of the sequence. If instead all segments in a C1C2V are planned and executed as a whole (globally), such compensatory relations are expected (e.g., a lengthening of one segment is compensated by a shortening of another segment in that whole).

In other words, we find that the global organization presiding over the segments partaking in these word-medial tautosyllabic CCVs is pleiotropic, that is, simultaneously expressed over a set of different phonetic parameters rather than via a privileged metric such as c-center stability or any other such given single measure (employed in prior works). This last point deserves special attention given the long-standing use of stability measures as diagnostics of syllabic organization. It is thus worth expanding on how our results demonstrate this point. Recall that stop-laterals embedded in the high vowel context exhibit global timing interval stability while the same clusters embedded in the low vowel context exhibit local timing interval stability. Similar puzzling stability results are observed for s-stops; specifically, /sk/ clusters embedded in the low (and back) vowel context and /sp/ clusters in the low and mid vowel context exhibit global timing interval stability. Instead, the same clusters embedded in the high vowel context and the /sk/ clusters in the mid vowel context show local timing stability. To understand these puzzling stability patterns, we need to consider jointly the properties of each sequence.

The starting observation is that stop-lateral clusters followed by a high vowel are less overlapped than when followed by a low vowel (as demonstrated in Section 3.1). Less overlap of the CC portion of the word-medial CCV results in more compression of the inner part (CV) of the CCV in the high vowel than in the low vowel context. Compression is in terms of C2 shortening and reduction of CV timing as indexed by the CV lag measure. Thus, for both VIV ~ VplV and VIV ~ VklV, the substantial C2 (lateral) compression and CV lag shortening result in a better alignment of the c-center landmarks compared to the C2 release landmarks. This is the reason why global timing interval stability is found in the context before high vowels. In contrast, in the context before low vowels, where the CC portion is more overlapped, compression of the inner part of the word-medial CCV does not take place to the same extent; instead, we find minimal shortening of the prevocalic lateral along with minimal reduction in the CV lag from VIV to VCIV which in turn results in a better alignment of the release landmarks compared to the c-center landmarks and thus in the stability of the local timing interval.

In similar fashion, for s-stop clusters, the starting observation is that sibilant duration is longer when the cluster is followed by a low and/or a mid vowel than when followed by a high vowel (as demonstrated in Section 3.4). Longer sibilant duration results in more compression of the inner part (CV) of the word-medial CCV in the low and mid vowel than in the high vowel context. Specifically, greater compression, in terms of C2 shortening and reduction of CV timing as indexed by the CV lag measure, is found in the low and mid vowel contexts than in the high vowel context. Thus, for both VpV ~ VspV and VkV ~ VskV in the low vowel context, the substantial C2 (stop) compression and CV lag shortening result in a better alignment of the c-center landmarks compared to the C2 release landmarks. This is the reason why global timing interval stability is found in the context before low vowels for s-stop clusters. Likewise, in /sp/ clusters followed by a mid vowel, C2 (/p/) compression and CV lag shortening is substantial (as in the case of the low vowel context described above). As a result, the c-center landmarks are better aligned compared to the C2 release landmarks, and hence the stability of the global timing interval occurs. Instead, in the context before high vowels, the compression of the inner part of the word-medial CCV does not take place to the same extent; rather, we find minimal shortening of the prevocalic stop along with minimal reduction in the CV lag from VCV to VsCV which in turn results in a better alignment of the release landmarks compared to the c-center landmarks and thus in the stability of the local timing interval. Lastly, in the /sk/ mid vowel context, we observe no compression of the inner part of the CCV, meaning that the addition of the sibilant to the left of the /k/ in the mid vowel context did not produce any shortening of the prevocalic stop or reduction in the CV lag from VkV to VskV; such a pattern, again, results in a better alignment of the release landmarks compared to the c-center landmarks and thus in the stability of the local timing interval.

It is apparent that consonant identity, IPI duration, but also vowel dependent properties (such as height) affect the patterns of timing intervals such that both global and local timing interval stabilities emerge for our English word-medial stop-lateral and s-stop clusters. Hence, it is clear that the stability heuristics fail to reliably diagnose syllabic organization, not only across different VCCV sequences but even in the same sequence in different vowel contexts. Nevertheless, these VCCV sequences share a general property which seems to reflect global organization: when a consonant is added to a CV to form a word-medial CCV, a number of readjustments ensue that effectively bring the vowel to overlap as much as possible with the cluster. Thus, for stop-lateral clusters, the prevocalic lateral duration decreases from VCV to VCCV, the relative timing of the internal /lV/ string changes, and there is a compensatory relation between IPI and C2 lateral duration such that as IPI increases, C2 lateral decreases. Similarly, for s-stop clusters, the prevocalic stop duration decreases from VCV to VCCV, the relative timing of the

internal stop CV string changes, and there is a compensatory relation between IPI and C1 sibilant duration such that as IPI increases, C1 sibilant decreases. These readjustments express effects where spatiotemporal modification in one local region of the sequence comes systematically with a change in another region of the sequence. Such compensatory effects indicate that the different parts of a CCV are not independently planned and produced (hence, globally organized).

Not only the existence but also the relative extent of the readjustments, resulting from adding a consonant in a CV to obtain a word-medial CCV, provides further evidence for global organization. Specifically, the extent of the readjustments is greater in contexts where the consonantal material before the vowel is longer or less overlapped; that is, in contexts where the parochial phonetic properties of the consonants before the vowel are such that the vowel would be distanced farther from the consonants, we find a greater degree of readjustment of the inner CV subsequence so that the consonants can overlap more with the tautosyllabic vowel. To illustrate this point, consider the observed effects in the stop-lateral low vowel context versus the stop-lateral high vowel context. Recall that in the latter context, the IPI is longer than in the former context (as shown in Section 3.1). Under these conditions, for the vowel to overlap with the cluster to the extent possible, readjustment in the prevocalic material must be of a greater extent in the stop-lateral high vowel than in the stop-lateral low vowel context. Indeed, our results show more lateral shortening and a greater change in the relative timing of the /IV/ string (as indexed by the CV lag measure) when the vowel is high compared to when it is low. This is so because stop-lateral clusters are less overlapped in the high vowel context than in the low and hence the lateral-vowel part has to compress more in the former vowel context than in the latter.

Consider also the observed effects in the s-stop low versus s-stop mid versus s-stop high vowel sequences. Recall that in the two former sequences, the sibilant duration is longer than in the latter sequence (as demonstrated in Section 3.4). Under these conditions, readjustment in the prevocalic material must be greater in the two former sequences than in the latter so that the vowel can overlap with the cluster to the extent possible. Indeed, our results show a greater compression, in terms of C2 shortening and reduction of CV timing as indexed by the CV lag measure, in the s-stop low and mid vowel than in the s-stop high vowel sequence. This is so because the sibilant duration is longer in the low and mid vowel context than in the high vowel context and thus the stop-vowel subsequence part has to compress more in the former than in the latter context. These results (the longer duration of the sibilant /s/, and the shorter duration of the prevocalic stop from low to mid to high vowel contexts) are in line with previous findings on English. Specifically, Bailey and Summerfield (1980: 545, 546, 547) observed two different relations between the openness of the vocalic context and the prevocalic

segments: a positive relation between sibilant duration and the openness of the subsequent vowel (i.e., the more open the vowel, the longer the sibilant duration) and an inverse relation between the prevocalic stop duration and the openness of the vowel (i.e., the more open the vowel, the shorter the prevocalic stop duration). Therefore, based on the phonetic properties of the segments (longer or shorter closure duration of the stop or the sibilant) and their intersegmental timing (longer or shorter IPI), we find different degrees of readjustment (in terms of prevocalic C shortening and reduction in CV timing) in attestation of the global organization in which the added consonant partakes with the rest of the CV sequence.

An exception to the pattern of the presence of readjustments was met for one of our stimuli, the word *cascade*. For this word, after the addition of the sibilant to the left of the stop-vowel sequence, we find no effects of readjustments (no C2 compression, no CV lag shortening), in attestation of an organization which is local. A closer examination of our data indicates that this word was produced by our participants with main stress on the vowel preceding the s-stop sequence, whereas in the rest of the s-stop stimuli words the main stress is found on the vowel following the s-stop sequence. Lexical stress is known to constitute an important phonological factor that affects syllabification (Derwing and Neary 1991; Redford and Randall 2005; Treiman and Danis 1988) and articulatory implementation (Gleason 1999). Particularly for English word-medial s-stop clusters, it has been found that in syllabification judgments, when the vowel of the first syllable is stressed, s-stop clusters were consistently prosodified as heterosyllabic (Redford and Randall 2005). Using the audio data collected in parallel with our articulatory data, we quantified the intensity and spectral cues of the vowels in both syllables of our stimuli with s-stop clusters. For the word *cascade*, across all repetitions, we found longer duration, higher intensity, larger amplitude, and higher pitch accent in the vocalic nucleus of the first syllable (/æ/) compared to the vocalic nucleus of the second syllable (/eɪ/). However, for the rest of the words (*discover*, *gaskeep*, *despite*, *aspeed*, *suspend*), longer duration, higher intensity, larger amplitude, and higher pitch accent were found on the vowel of the second syllable than on the one of the first syllable. Figure 20 illustrates the stress differences between the word *cascade* and the rest of the stimuli words. All parameters measured point to a strong-weak stress pattern for the word *cascade*, meaning that the main stress in this word occurs in the first syllable as opposed to the rest of our data where main stress is on the second syllable. Thus, given the heterosyllabic prosodification of the word-medial /sk/, it is not surprising that this cluster exhibits an absence of readjustment effects, in contrast to the other clusters, which in turn reflects the local metrification for this cluster in our dataset.

This exception does not contradict the rest of the findings. In fact, it supports the main claim that distinct syllabic organizations are expressed differently in

Figure 20: Duration (s), pitch (Hz), and intensity (dB) (measured midway in each of the two vowels of each stimulus word) as stress cues showing the stress differences between the word *cascade* and the rest stimuli.

terms of spatiotemporal coordination relations. Global organization is expressed through the presence of readjustment effects to local perturbations on the segmental sequence that ripple through the entire hypothesized syllabic unit, while local organization is expressed through the absence of such effects. Our findings on the influence of stress on syllabic organization call for further research; studies investigating how stress patterns affect the indices of syllabic organization using articulatory data are rather limited (e.g., Byrd and Choi 2010; Gleason 1999). Overall, our results confirm that syllabic organization modulates the spatiotemporal coordination of the segments in a multiplicity of ways and specifically that global organization is expressed via readjustment effects and compensatory relations among phonetic parameters of the segments partaking in that organization.

5 Conclusions

This paper has addressed the relation between syllable structure and intersegmental spatiotemporal coordination. Specifically, we examined the spatiotemporal properties of the word-medial stop-lateral /pl, kl/ and s-stop /sp, sk/ clusters in the context of three different vowel heights (high, mid, low) produced by seven speakers of American English.

In its most concise form, our main result can be stated by saying that adding a consonant to the left of a CV to obtain a CCV results in readjustments of the

spatiotemporal structure of the inner CV subsequence in the word-medial clusters investigated. These readjustments express effects where spatiotemporal modification in one local region of the sequence comes systematically with a change in another region of the sequence, indicating that the different parts of a CCV are not independently planned and produced (hence, globally organized).

The specific ways in which readjustments are actuated are context-dependent. For example, a greater compression, in terms of C2 shortening and reduction of CV timing (as indexed by the CV lag measure), is found for stop-laterals when the vowel is high compared to when it is low, and for s-stops, when the vowel is low or mid than when it is high. Recall that both stop-laterals and s-stops exhibit puzzling stability patterns across vowel contexts. Yet, both cluster groups across vowel contexts provide evidence for global organization in the form of a number of readjustments that effectively bring the vowel to overlap as much as possible with the entire cluster. Global organization in our stop-lateral and s-stop clusters is achieved in multiple ways: shortening of the prevocalic consonant in the word-medial cluster compared to its singleton paired word, reduction of CV timing (as indexed by the CV lag measure) in the word-medial cluster case compared to its singleton paired word, and presence of compensatory relations between phonetic properties of different segments or intersegmental transitions within each cluster.

Let us place our results in the context of related work on stop-lateral and s-stop clusters. Gafos et al. (2020) compared the spatiotemporal properties of stop-lateral clusters in Moroccan Arabic and Spanish. Moroccan Arabic and Spanish stop-lateral clusters present an opportunity to study the same segmental sequences under contrasting phonological organizations because these two languages are considered to be prototypical examples of the complex (Spanish) versus simplex (Moroccan Arabic) onset typological distinction. Moroccan Arabic does not show the set of readjustments our present work documents in detail for English and Gafos et al. (2020) document for Spanish. For instance, whereas in English when a /k/ is added to the left of /lV/ to form /klV/ we observed in the present results compression of the lateral (with the same result holding also for Spanish; Gafos et al. 2020), such an adjustment of the lateral is not observed in Moroccan. The same point can be made with respect to the compensatory relation between IPI and lateral duration, where a long(er) lag between the /k/ and the lateral must be compensated by shortening of the lateral in English. In Moroccan, this relation between the two parameters is absent (in fact, Gafos et al. 2020 report a weak positive relation when using raw duration). Hence, unlike English, the different parts of the CC subsequence are independently produced (hence, local organization) in Moroccan: spatiotemporal modification in one local region (here, IPI) of that subsequence does not result in a change in the other part (here, lateral duration). Overall, adding /k/ to a sequence of segments /lV/ produces effects that ripple through the inner lateral-vowel sequence in English but not so in Moroccan.

In parallel work, we are pursuing comparable analyses to our English analyses reported in the present work also for Spanish word-medial s-stop sequences. Spanish and English s-stop clusters have distinct phonological organizations. Whereas Spanish syllabifies obstruent-liquid clusters as complex onsets both word-initially and medially, s-stop sequences form illicit word and syllable onsets and are syllabified heterosyllabically (e.g., *mos-ca* ‘fly’, *cons-ta* ‘it consists’; Hualde 2005: 75; see also Harris 1983: 27, 30). Instead, English allows both obstruent-liquid and s-stop clusters to form complex onsets word-medially (Blevins 1995; Clements 1990; Kahn 1976), at least in the case where the cluster precedes a stressed vowel. Thus, s-stop clusters can form complex onsets in English but not in Spanish. In our results, we find that whereas English word-medial s-stop clusters show substantial compression of the stop when a sibilant /s/ joins a stop-vowel string to form a CCV, this is not the case for Spanish word-medial s-stops. A more telling example is illustrated with the compensatory relation between IPI and C1 sibilant duration, where a long(er) lag between the sibilant /s/ and the stop must be compensated by shortening of the sibilant in English. In Spanish, we find no such relation between the two parameters. Thus, in contrast to English, the different parts of the s-stop subsequence are independently produced (hence, local organization) in Spanish: an expansion or contraction in some part (here, IPI) of that subsequence does not result in a contraction or expansion of the other part (here, sibilant duration). Overall, adding a sibilant /s/ to a stop-vowel sequence produces effects that ripple through the inner sequence in English but not in Spanish.

Conceptually, perhaps the most important implication of our results is that there is no unique phonetic exponent of global organization. Uniformity in phonological organization (same organization presiding over sequences with varying segmental makeup as in /pl, kl, sp, sk/ before different vowels) does not imply uniqueness of phonetic exponents for that organization. Syllabic organization is simultaneously expressed by more than one phonetic exponent and these exponents enter into compensatory relations with one another in attestation of the global organization they instantiate.

Acknowledgements: The authors thank the editors and the two anonymous reviewers whose comments have substantially improved the paper. We would also like to thank the participants who took the time to participate in our study.

Research funding: This work has been funded by the European Research Council (AdG 249440) and by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project ID 317633480 – SFB 1287, Project C04.

Author contribution: All authors contributed equally to this work.

Statement of ethics: The experiment has been approved by the authors’ institutional ethics committee with application ID 62/2016.

Conflict of interest statement: The authors declare no conflict of interest.

Appendix

A1: Relative standard deviation (RSD) of the intervals global timing and local timing using three anchors (Ctar, Cmax, Vmax) across pairs of words for seven speakers in the s-stop subset.

Speaker	Word pair	Ctar		Cmax		Vmax	
		Global	Local	Global	Local	Global	Local
VP01	append ~ suspend	12.07	9.66	32.08	17.13	15.58	54.63
	blockade ~ cascade	6.90	7.09	18.79	9.28	9.40	20.52
	gakeep ~ gaskeep	15.99	13.21	31.50	18.74	15.44	41.54
	recover ~ discover	9.62	9.37	20.54	15.86	14.42	29.36
	repeat ~ aspeed	14.79	14.47	29.71	14.45	13.83	42.95
VP02	repine ~ despite	10.86	10.66	24.70	14.43	11.87	57.60
	append ~ suspend	7.13	5.68	29.65	12.37	9.97	30.68
	blockade ~ cascade	16.37	16.01	21.01	17.47	16.83	26.01
	gakeep ~ gaskeep	23.33	25.90	33.45	39.27	40.48	37.69
	recover ~ discover	10.00	9.71	6.42	10.89	10.17	8.55
VP03	repeat ~ aspeed	13.93	14.62	24.26	25.73	26.42	28.40
	repine ~ despite	11.75	11.05	34.96	13.72	12.91	51.99
	append ~ suspend	12.01	12.68	30.83	13.75	14.47	80.91
	blockade ~ cascade	12.23	12.23	23.46	15.97	15.79	31.05
	gakeep ~ gaskeep	6.78	6.54	18.60	13.08	12.03	29.88
VP04	recover ~ discover	10.03	9.51	20.18	27.87	24.48	45.81
	repeat ~ aspeed	16.57	13.97	26.65	13.74	13.43	35.12
	repine ~ despite	12.63	11.91	19.73	22.72	21.44	53.27
	append ~ suspend	7.61	8.45	19.53	10.06	11.47	30.28
	blockade ~ cascade	14.41	13.44	24.22	20.92	19.51	50.15
VP05	gakeep ~ gaskeep	17.75	20.03	16.79	16.62	16.79	16.06
	recover ~ discover	13.93	11.41	28.69	20.05	17.59	44.65
	repeat ~ aspeed	10.44	10.35	21.63	8.73	8.73	20.71
	repine ~ despite	11.70	11.41	22.70	16.09	15.51	37.68
	append ~ suspend	14.94	13.68	21.80	14.57	14.03	37.27
VP06	blockade ~ cascade	14.55	14.20	24.15	16.69	16.24	37.39
	gakeep ~ gaskeep	13.41	12.83	36.31	12.19	11.46	36.86
	recover ~ discover	10.70	10.23	16.15	16.58	16.64	22.22
	repeat ~ aspeed	12.87	12.90	38.34	11.24	11.37	41.22
	repine ~ despite	15.22	14.98	33.03	18.75	18.16	56.38
VP08	append ~ suspend	8.89	9.76	17.33	14.07	13.50	32.27
	blockade ~ cascade	11.23	10.92	17.89	12.12	11.72	17.12
	gakeep ~ gaskeep	10.76	8.93	25.18	17.65	14.93	41.39
	recover ~ discover	12.86	12.20	26.13	23.91	23.30	40.63
	repeat ~ aspeed	13.87	13.54	34.27	9.92	10.05	64.94
VP08	repine ~ despite	9.51	9.12	32.78	16.07	14.71	64.87
	append ~ suspend	11.17	11.76	18.14	14.93	15.02	27.17
	blockade ~ cascade	10.36	10.84	23.44	14.44	14.89	25.87
	gakeep ~ gaskeep	9.09	10.61	25.95	16.56	12.83	36.72
	recover ~ discover	14.76	14.10	15.51	17.09	16.48	19.99

A1: (continued)

Speaker	Word pair	Ctar		Cmax		Vmax	
		Global	Local	Global	Local	Global	Local
	repeat ~ aspeed	14.87	13.08	28.78	16.04	13.82	33.65
	repine ~ despite	14.50	11.68	28.96	20.71	17.42	77.32

A2: Relative standard deviation (RSD) of the intervals global timing and local timing using three anchors (Ctar, Cmax, Vmax) across pairs of words for seven speakers in the stop-lateral subset.

Speaker	Word pair	Global			Local		
		Ctar	Cmax	Vmax	Cmax	Ctar	Vmax
VP01	applied ~ collide	12.28	12.14	26.30	9.94	9.35	23.66
	delete ~ replete	16.15	13.91	15.26	22.36	18.51	23.05
	delight ~ decline	21.94	21.70	50.90	15.66	15.55	75.42
	ellipse ~ eclipse	23.09	21.24	53.89	19.84	18.73	42.35
VP02	applied ~ collide	12.79	12.69	35.26	10.54	11.38	39.88
	delete ~ replete	17.99	18.21	20.10	17.20	17.63	20.86
	delight ~ decline	16.91	17.29	46.62	12.43	12.95	54.25
	ellipse ~ eclipse	11.19	9.30	15.58	8.95	10.45	16.08
VP03	applied ~ collide	11.23	10.68	32.50	11.53	10.40	62.30
	delete ~ replete	10.73	10.31	15.29	15.32	14.80	7.35
	delight ~ decline	11.78	10.47	33.13	18.51	16.16	50.97
	ellipse ~ eclipse	8.24	11.33	34.91	22.66	19.15	119.72
VP04	applied ~ collide	9.70	9.59	28.99	5.84	5.83	29.20
	delete ~ replete	21.31	20.86	37.36	23.23	22.30	52.55
	delight ~ decline	12.30	12.86	88.05	17.12	17.88	751.36
	ellipse ~ eclipse	11.93	7.08	16.24	25.58	17.62	42.18
VP05	applied ~ collide	13.22	13.21	38.09	11.37	11.45	31.66
	delete ~ replete	22.66	18.78	17.56	32.14	24.85	30.13
	delight ~ decline	26.59	24.93	58.36	17.82	16.46	34.34
	ellipse ~ eclipse	16.01	16.78	38.71	14.92	15.77	82.98
VP06	applied ~ collide	9.87	9.53	14.52	8.10	7.74	19.91
	delete ~ replete	13.53	11.59	37.30	16.65	14.19	50.18
	delight ~ decline	16.42	15.77	50.17	15.73	14.88	42.95
	ellipse ~ eclipse	12.58	15.24	22.58	13.43	16.97	47.70

A2: (continued)

Speaker	Word pair	Global			Local		
		Ctar	Cmax	Vmax	Cmax	Ctar	Vmax
VP08	applied ~ collide	8.93	8.92	20.91	8.83	8.58	34.26
	delete ~ replete	16.88	16.02	21.90	18.35	16.83	28.23
	delight ~ decline	9.68	9.12	31.11	6.90	6.46	21.60
	ellipse ~ eclipse	20.45	13.43	27.15	27.79	19.28	42.60

References

Bailey, Peter J. & Quentin Summerfield. 1980. Information in speech: Observations on the perception of [s]-stop clusters. *Journal of Experimental Psychology: Human Perception and Performance* 6(3). 536–563.

Bates, Douglas, Reinhold Kliegl, Shravan Vasishth & Harald Baayen. 2018. Parsimonious mixed models. Available at: <http://arxiv.org/pdf/1506.04967v2>.

Bates, Douglas, Martin Mächler, Benjamin Bolker & Steven Walker. 2014. Fitting linear mixed-effects models using lme4. Available at: <http://arxiv.org/pdf/1406.5823v1>.

Berkovits, Rochele. 1993. Progressive utterance-final lengthening in syllables with final fricatives. *Language and Speech* 36(1). 89–98.

Berkovits, Rochele. 1993. Utterance-final lengthening and the duration of final-stop closures. *Journal of Phonetics* 21(4). 479–489.

Blevins, Juliette. 1995. The syllable in phonological theory. In John Goldsmith (ed.), *The handbook of phonological theory*, 206–244. London: Blackwell.

Bombien, Lasse. 2011. *Segmental and prosodic aspects in the production of consonant clusters*. Munich, Germany: University of Munich Dissertation.

Browman, Catherine P. & Louis Goldstein. 1988. Some notes on syllable structure in articulatory phonology. *Phonetica* 45. 140–155.

Brunner, Janna, Christian Geng, Stavroula Sotiropoulou & Adamantios Gafos. 2014. Timing of German onset and word boundary clusters. *Laboratory Phonology* 5(4). 403–454.

Byrd, Dani. 1995. C-centers revisited. *Phonetica* 52(4). 285–306.

Byrd, Dani & Susie Choi. 2010. At the juncture of prosody, phonology, and phonetics – The interaction of phrasal and syllable structure in shaping the timing of consonant gestures. In Cecile Fougeron, Barbara Kühnert, Mariapaola D’Imperio & Nathalie Vallée (eds.), *Laboratory phonology*, vol. 10, 31–60. Berlin, New York: De Gruyter Mouton.

Clements, George N. 1990. The role of sonority cycle in core syllabification. In John Kingston & Mary E. Beckman (eds.), *Papers in laboratory phonology*, 283–333. Cambridge: Cambridge University Press.

Derwing, Bruce L. & Terrance M. Neary. 1991. The ‘vowel-stickiness’ phenomenon: Three experimental sources of evidence. In *Proceedings of 12th international congress of phonetic science*, vol. 3, 210–213. Aix-en-Provence: Publications de l’Université de Provence.

Gafos, Adamantios, Philip Hoole, Roon Kevin & Chakir Zeroual. 2010. Variation in overlap and phonological grammar in Moroccan Arabic clusters. In Cecile Fougeron, Barbara Kühnert,

- Mariapaola D'Imperio & Nathalie Vallée (eds.), *Laboratory phonology*, vol. 10, 657–698. Berlin, New York: De Gruyter Mouton.
- Gafos, Adamantios, Jens Roeser, Stavroula Sotiropoulou, Philip Hoole & Chakir Zeroual. 2020. Structure in mind, structure in vocal tract. *Natural Language and Linguistic Theory* 38(1). 43–75.
- Gleason, Philip L. 1999. *Syllable structure and stress effects on articulatory coordination*. Boca Raton, FL: Florida Atlantic University Dissertation.
- Goldstein, Louis, Ioana Chitoran & Elisabeth Selkirk. 2007. Syllable structure as coupled oscillator modes: Evidence from Georgian versus Tashlhiyt Berber. In Jürgen Trouvain & William J. Barry (eds.), *Proceedings of the 16th international congress of phonetic sciences (ICPhS)*, 241–244. Germany: Saarbrücken.
- Goldstein, Louis, Hosung Nam, Elliot Saltzman & Ioana Chitoran. 2009. Coupled oscillator planning model of speech timing and syllable structure. In Gunnar Fant, Hiroya Fujisaki & Jiaxuan Shen (eds.), *Frontiers in phonetics and speech science*, 239–250. Beijing, China: The Commercial Press.
- Haggard, Mark. 1973. Correlations between successive segment durations: Values in clusters. *Journal of Phonetics* 1. 111–116.
- Harris, James W. 1983. *Syllable structure and stress in Spanish: A nonlinear analysis*. Cambridge, MA: MIT Press.
- Hermes, Anne, Bastian Auris & Doris Mücke. 2015. Computational modelling for syllabification patterns in Tashlhiyt Berber and Maltese. In Susanne Fuchs, Martine Grice, Anne Hermes, Leonardo Lancia & Doris Mücke (eds.), *Proceedings of the 10th international seminar on speech production (ISSP)*, 186–189. Cologne, Germany.
- Hermes, Anne, Doris Mücke & Bastian Auris. 2017. The variability of syllable patterns in Tashlhiyt Berber and Polish. *Journal of Phonetics* 64(3). 127–144.
- Honorof, Douglas & Catherine P. Browman. 1995. The centre or edge: How are consonant clusters organized with respect to the vowel? In Kjell Elenius & Peter Branderud (eds.), *Proceedings of the 13th international congress of the phonetic sciences (ICPhS) 3*, 552–555. Stockholm, Sweden.
- Hualde, José I. 2005. *The sounds of Spanish*. New York City: Cambridge University Press.
- Kahn, Daniel. 1976. *Syllable-based generalizations in English phonology*. Cambridge, MA: MIT Dissertation.
- Katz, Jonah. 2012. Compression effects in English. *Journal of Phonetics* 40(3). 390–402.
- Klatt, Dennis. 1974. The duration of (s) in English words. *Journal of Speech and Hearing Research* 17(1). 51–63.
- Klatt, Dennis. 1975. Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research* 18(4). 686–706.
- Lenth, Russel, Buerkner Paul, Maxime Herve, Jonathon Love, Hannes Riebl & Henrik Singmann. 2020. Emmeans (Version R package 1.5.1) [Computer software]. Available at: <https://CRAN.R-project.org/package=emmeans>.
- Marin, Stefania. 2013. The temporal organization of complex onsets and codas in Romanian: A gestural approach. *Journal of Phonetics* 41(3–4). 211–227.
- Marin, Stefania & Marianne Pouplier. 2010. Temporal organization of complex onsets and codas in American English: Testing the predictions of a gestural coupling model. *Motor Control* 14(3). 380–407.
- Oller, Kimbrough D. 1973. The effect of position in utterance on speech segment duration in English. *The Journal of the Acoustical Society of America* 54(5). 1235–1247.

- Poupplier, Marianne. 2012. The gestural approach to syllable structure: Universal, language and cluster-specific aspects. In Susanne Fuchs, Melanie Weirich, Daniel Pape & Pascal Perrier (eds.), *Speech planning and dynamics*, 63–96. Bern, Switzerland: Peter Lang AG.
- RStudio Team. 2015. RStudio: Integrated Development for R. RStudio, version 1.2.1335 Inc., Boston, MA. Available at: <http://www.rstudio.com/>.
- Redford, Melissa A. & Patrick Randall. 2005. The role of juncture cues and phonological knowledge in English syllabification judgments. *Journal of Phonetics* 33(1). 27–46.
- Shaw, Jason, Adamantios Gafos, Philip Hoole & Chakir Zeroual. 2009. Syllabification in Moroccan Arabic: Evidence from patterns of temporal stability in articulation. *Phonology* 26(1). 187–215.
- Shaw, Jason & Adamantios Gafos. 2015. Stochastic time models of syllable structure. *PLoS One* 10(5). e0124714.
- Shaw, Jason, Adamantios Gafos, Philip Hoole & Chakir Zeroual. 2011. Dynamic invariance in the phonetic expression of syllable structure: A case study of Moroccan Arabic consonant clusters. *Phonology* 28(3). 455–490.
- Sotiropoulou, Stavroula, Mark Gibson & Adamantios Gafos. 2020. Global organization in Spanish onsets. *Journal of Phonetics* 82. 100995.
- Treiman, Rebecca & Catalina Danis. 1988. Syllabification of intervocalic consonants. *Journal of Memory and Language* 27(1). 87–104.
- Westbury, John R. 1994. *X-ray microbeam speech production database user's handbook 1994*. Madison, WI: University of Wisconsin.
- Wiese, Richard. 1996. *The phonology of German*. Oxford, UK: Oxford University Press.