Probing syllabic affiliation of word-initial and word-medial consonant sequences in north-central Peninsular Spanish

Karthik Durvasula
karthikd@msu.edu
Department of Linguistics, Languages, and Cultures
Michigan State University
East Lasing, MI
United States

Silvina Bongiovanni
sbongio@msu.edu
Department of Romance and Classical Studies
Michigan State University
East Lasing, MI
United States

Abstract

There is now a large literature probing syllable affiliation of consonant sequences through phonetic measurements. These studies often use one of two diagnostic measures: (1) temporally stable intervals using relative standard deviation, and (2) compensatory shortening effects. In this study, we argue that both measures are difficult to infer from without precise theoretically predicted expectations and additional controls. We studied eleven native speakers of North-Central Peninsular Spanish who pronounced disyllabic real/nonce Spanish words with varying consonant sequences. On the face of it, our temporal stability and compensatory shortening results challenge the standard analysis of syllabic affiliation in Spanish phonology, potentially supporting a complex onset analysis for /sl/ and /sm/. However, in post hoc analyses we observed shortening effects outside the target syllable due to consonant sequences, indicating evidence for poly-constituent shortening. Therefore, compensatory shortening effects within a syllable cannot automatically be assumed to be due to syllable structure. Our results and simulations suggest that, despite superficial evidence of a c-centre alignment, the clusters are more consistent with a right-edge alignment once poly-constituent shortening and domain-initial lengthening are taken into consideration.

1 Introduction

The last 40 years have seen many advances in connecting phonetic signatures to syllabic affiliation of segments. Using articulatory methods, and more recently replicated with acoustics, measures of temporal stability, like C-CENTRE-TO-ANCHOR and RIGHT-EDGE-TO-ANCHOR patterns, have been probed to see if they reflect syllabic affiliation of consonant sequences (Browman and Goldstein 1988; Durvasula et al. 2021; Goldstein et al. 2007; Hermes et al. 2017, 2013; Marin and Pouplier 2014; Shaw et al. 2009; Sotiropoulou et al. 2020). Generally, two patterns of stability have been observed in the literature, each theorized to be characteristic of a particular syllabic affiliation. In languages allowing complex onsets, such as English, studies reveal that typically word-initial consonant sequences temporally reorganize as a unit. They synchronize with the nuclear vowel, meaning that the midpoint of consonantal gestures aligns temporally with the end of the nuclear vowel, maintaining a constant duration between these two points, regardless of the number of consonants in the onset. In contrast, languages that do not admit complex onsets typically exhibit a different articulatory behavior (Durvasula et al. 2021; Goldstein et al. 2007; Hermes et al. 2017; Shaw et al. 2009, 2011). In such languages, the articulatory gesture for the second consonant in word-initial sequences maintains a stable relationship with the following vowel. This results in temporal stability between the second consonant (rather than the entire sequence) and the end of the nuclear vowel. Although these patterns have been identified in various languages, not all languages adhere strictly to these distinctions. Sometimes they can even vary within the same language (Hermes et al. 2013). German (Wiese 1996), French (Dell 1995), Hebrew (Bolozky 1997), and Italian (Davis 1987) have all been shown to exhibit different types of temporal stability patterns with different types of consonant sequences, arguably undermining the relationship between stability patterns and syllabic affiliation (an issue we expand on in Section 2).

In this article, we build upon existing research by examining temporal stability patterns within consonant sequences in North-Central Peninsular Spanish. In addition to using acoustic techniques, we also incorporate word-medial environments into our study. More specifically, we look at the temporal stability patterns observed in consonant sequences and the phonetic shortening effect observed in the C_2 position of particular word-initial /fl/ sequences and word-medial /fl sl sm/ sequences in Spanish. In contrast to the standard phonological analysis, our findings show patterns which could be interpreted as being consistent with being complex onsets, for all consonant sequences under analysis, and in all analyzed word positions. However, post-hoc analyses show shortening effects outside the target syllable as well, indicating that the observed C-CENTRE effects are confounded by poly-constituent shortening and perhaps domain-initial lengthening. Although the examination of poly-constituent shortening was not originally a part of our research design, its consideration in the post-hoc analysis opened an alternative explanation to the patterns uncovered, namely that the observed durational changes are not just local to the syllable and therefore are not evidence of syllabic affiliations. We followed up with simulation work along the lines of Shaw and Gafos (2015) and Shaw et al. (2009) and confirmed that the general direction of the observed temporal stability patterns was possible under both complex and simplex onset organisation, and that this was true even when we modelled in poly-constituent shortening based on our own results. However, as we note below, there is some evidence that the observed temporal stability patterns were more consistent with RIGHT-EDGE-TO-ANCHOR stability.

In the next two sub-sections, we review how temporal stability metrics have generally been viewed to relate to onset affiliation, followed by a review of relevant phonological and phonetic facts of Spanish syllable structure, specifically as it relates to consonant sequences. We then present the methods and results of two experiments in Sections 5 and 6, respectively. In Section 7 we present the post-hoc analyses, followed by the relevant simulations in Section 8. We conclude the article with a discussion of some implications of this research in Section 9.

2 Temporal stability metrics and onset affiliation

In an important first assay on the topic, Browman and Goldstein (1988) found that onset consonants in American English showed a temporal stability pattern around the centre of the mid-points of their oral gestures. They termed this abstract centre point the C-CENTRE. Specifically, they analysed articulatory data from the Tokyo x-ray microbeam database, which consisted of sets of nonsense words/phrases (*e.g.*, pi lats vs. pi plats vs. pi splats), and found that irrespective of the number of consonants (1 vs. 2 vs. 3) in the onsets of the second word in such phrases, the C-CENTRE point was in a stable relationship with the following vowel (the ANCHOR). So, the addition of more consonants to the onset did not seem to affect the duration between the C-CENTRE and the ANCHOR. This pattern of stability is schematised in Figure 1 (left), and has since been termed C-CENTRE-TO-ANCHOR interval stability. The pattern of stability of the C-CENTRE-TO-ANCHOR interval for onset consonants has been replicated for at least some consonant sequences, in American English (Marin and Pouplier 2010), Romanian (Marin and Pouplier 2014), Georgian (Goldstein et al. 2007), Italian (Hermes et al. 2013), Polish (Hermes et al. 2017), and Spanish (Sotiropoulou et al. 2020).

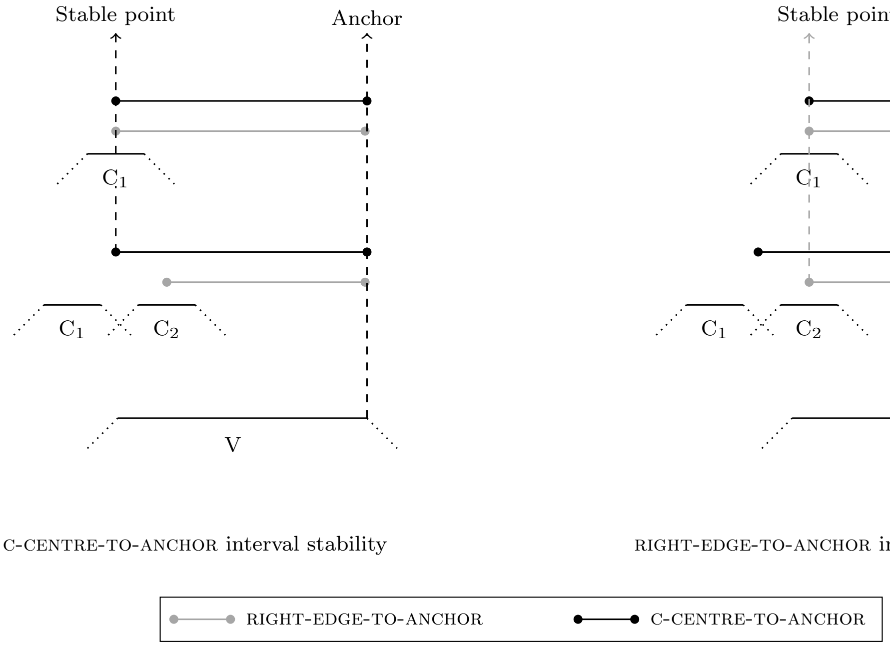


Figure 1: Schematic representations of C-CENTRE-TO-ANCHOR interval stability patterns (left) and RIGHT-EDGE-TO-ANCHOR interval stability patterns (right) (figure adapted from Durvasula et al. (2021) and Shaw et al. (2009). The x-axis in the figure represents time. The ANCHOR marks the end of the following vowel, and C_1-C_2 represent word-initial consonants.

In contrast to the above stability pattern observed in languages that allow complex onsets, languages that do not allow complex onsets typically show a different, RIGHT-EDGE-TO-ANCHOR, stability. For example, Tashlhiyt Berber, despite having word-initial consonant sequences, disallows complex onsets. Goldstein et al. (2007) and Hermes et al. (2017) observed that the right-most consonant of a word-initial consonant sequence (the RIGHT-EDGE) is in a stable relationship with the following vowel, as schematised in Figure 1 (right). This pattern of RIGHT-EDGE-TO-ANCHOR interval stability has also been observed in Moroccan Arabic (Shaw et al. 2009, 2011) and Jazani Arabic (Durvasula et al. 2021).

In an important result in this line of research, Hermes et al. (2013) showed that the stability patterns can vary within the same language. In Italian, phonologists have argued for different types of onset complexity for different consonant sequences (Davis 1987). For example, there is a clear morpho-phonological pattern related to the definite article that consonant sequences exhibit in Italian — nouns with some types of word-initial sibilant-initial sequence (*e.g.*, /sp/) appear with the allomorph /lo/, while other consonant sequences (*e.g.*, /pr/) and singleton consonants (*e.g.*, /p/) appear with the allomorph /il/. Based on such patterns, it has been argued that the different consonant sequences have different types of onset complexity — wherein some types of word-initial sibilant-initial consonant sequences do not form a complex onset, while the other word-initial consonant sequences do (Davis 1987). In line with this distinction, Hermes et al. (2013) argued that the relevant sibilant-initial consonant sequences show RIGHT-EDGE coordination, but non-sibilant-initial consonant sequences do not; they seem to show a C-CENTRE co-ordination. This particular set of facts from Italian argues against the possibility that the previously observed stability patterns in other languages were simply a language specific setting of gestural co-ordination between pre-vocalic consonants, and allows us to clearly establish the link between the stability patterns and onset complexity.

In summary, the results discussed above suggest a potential linking hypothesis between onset complexity and temporal stability patterns associated with the following vowel, namely, that consonant sequences that form complex onsets have a C-CENTRE-TO-ANCHOR interval stability, while those that form simplex onsets have a RIGHT-EDGE-TO-ANCHOR interval stability. Consequently, researchers can potentially use the C-CENTRE-TO-ANCHOR vs. RIGHT-EDGE-TO-ANCHOR interval stability pattern to probe onset complexity. If a consonant sequence belongs to the same onset (or syllable), then there should be a C-CENTRE effect. If a consonant sequence has consonants that are not part of the same onset (or syllable), there should not be a C-CENTRE effect.

However, there are some results that contradict the above linking hypothesis. Some arguments in the phonological literature suggest that word-initial consonant sequences in Hebrew (Bolozky 1997), French (Dell 1995) and German (Wiese 1996) form complex onsets. However, the three languages have been observed to show a RIGHT-EDGE alignment, at least for some consonant sequences (Brunner et al. 2014; Pouplier 2012; Tilsen et al. 2012). While the observations might at first blush seem problematic for the linking hypothesis discussed above, there are at least three different ways one could account for them. First, Mücke et al. (2020) argued that the patterns observed in the languages are consistent with a complex onset organisation, and previous research likely misinterpreted the relevant articulatory data. More specifically, they suggest that explicitly modelling the speaker-specific coupling strength between gestures and speaker-specific biomechanical interactions between articulators allows us to still understand the patterns in such languages as stemming from a C-CENTRE organisation. They argue that this is likely the case for at least German. Second, Sotiropoulou et al. (2020) suggest that relevant cues of different syllabic affiliations are in fact distributed over a variety of gestural adjustments within a syllable, and may not show up in each such aspect. They thereby suggest a "global" organisation over syllables. Finally, Durvasula et al. (2021) suggest that, when the intervals are extracted from acoustic measurements, the relevant articulatory data is in fact indirect evidence of the the stability patterns present in the acoustics, and that there might be more stability for the C-CENTRE-TO-ANCHOR than for the RIGHT-EDGE-TO-ANCHOR for the three languages. This last possibility receives further support from recent work that did find a C-CENTRE-TO-ANCHOR stability pattern using the acoustic method for the same sequences that didn't show the pattern in the articulations

(Franke et al. 2023). All three of the above suggestions raise the possibility that the articulatory stability patterns observed in the three languages may not be counterexamples to the linking hypothesis after all.

While almost all previous related work has studied the phenomenon by observing gestural coordination using articulatory data, in work that is most relevant for the present article, there is clear evidence that acoustic recordings can be used to observe a C-CENTRE-TO-ANCHOR interval stability pattern for word-initial consonant sequences in a complex onset language like English (Durvasula et al. 2021; Selkirk and Durvasula 2013; Shaw and Gafos 2015), and a RIGHT-EDGE-TO-ANCHOR interval stability pattern for word-initial consonant sequences in a language that does not allow complex onsets, like Jazani Arabic Durvasula et al. (2021). Durvasula et al. (2021, p. 198) point out that their particular result opens up the possibility of probing such effects both in the lab and in field work, and express hope that the technique "will be employed in a variety of languages and contexts – not only to test its viability, but also to examine its correlation with more traditional analytical techniques for inferring syllable structure." We follow up on this hope in this paper by probing the syllable structure of word-initial and word-medial consonant sequences in Spanish through acoustic techniques.

3 Syllabic affiliation of consonant sequences in Spanish

The standard analysis of the syllabic affiliation of onset consonant clusters is not generally regarded as a controversial topic in Spanish phonology (Colina 2009, 2012; Harris 1983; Hualde 1991, 2005; Morales-Front 2018; Real Academia Española 2011; Saporta and Contreras 1962). Under this standard view, Spanish onset clusters (either word-initially or word-medially) may have at most two consonants and their structure is very constrained. Licit sequences consist of an obstruent (including a labiodental fricative),[1] /p t k b d g f/[2], as the first member of the cluster, and a liquid, /l/ or /ɾ/, as the second (see examples in Table 1). All other consonant sequences are standardly described as heterosyllabic. Onset clusters with coronals /d t/ followed by /l/ are typically seen as exceptions to the generalisation, as they are not observed word-initially in most dialects.[3] The standard analysis of the syllabic affiliation of onset consonant clusters is not generally regarded as a controversial topic in Spanish phonology (Colina 2009, 2012; Harris 1983; Hualde 1991, 2005; Morales-Front 2018; Real Academia Española 2011; Saporta and Contreras 1962). Under this standard view, Spanish onset clusters (either word-initially or word-medially) may have at most two consonants and their

---

[1] While most analyses agree that the well-formedness of the onset cluster is driven by sonority, there is less agreement on why /f/ is the only fricative to pattern with plosives. Both Martínez-Gil (2001) and Colina (2016) argue that /f/ is unspecified for the feature [continuant] in the phonological representation. In their analyses, [+continuant] contributes to sonority. Segments that are specified as [+continuant] are more sonorous than those with the opposite specification (like stops), or with underspecification (like /f/). Thus, /f/ belongs in the same class as stops, which are uncontroversially [-continuant].

[2] Note, /sC/ sequences do not form onset clusters as per the standard analysis.

[3] The word-initial sequence */dl/ is not observed in any dialect, but /tl/ is observed in many varieties of Latin America (Hualde 2005), likely due to contact with languages where it is a licit sequence, such as Nahuatl.

structure is very constrained. Licit sequences consist of an obstruent (including a labiodental fricative),[4] /p t k b d g f/, as the first member of the cluster, and a liquid, /l/ or /ɾ/, as the second (see examples in Table 1). All other consonant sequences are standardly described as heterosyllabic. Onset clusters with coronals /d t/ followed by /l/ are typically seen as exceptions to the generalisation, as they are not observed word-initially in most dialects.[5]

Word-initial		**Word-medial**	
<brazo>	arm.MASC.SG	<abrazo>	hug.MASC.SG
<drena>	drains.3SG	<adrenalina>	adrenaline.FEM.SG
<grado>	grade.MASC.SG	<sagrado>	sacred.MASC.SG
<premio>	prize.MASC.SG	<apremio>	urgency.MASC.SG
<traer>	bring.INF	<distraer>	distract.INF
<creible>	believable.SG	<increible>	incredible.SG
<fresco>	chilly.MASC.SG	<refresco>	soft drink.MASC.SG
<bloquear>	block.INF	<desbloquear>	unblock.INF
<gluten>	gluten.MASC.SG	<degluten>	swallow.3PL
<plaza>	park.FEM.SG	<aplaza>	postpones.3SG
<clon>	clone.MASC.SG	<ciclon>	cyclone.MASC.SG
<flan>	Spanish cream caramel.MASC.SG	<inflan>	inflate.3PL

Table 1: Word-initial and word-medial consonant sequences with /r/ (top) and /l/ (bottom)

Early experimental work on complex onsets looked primarily at the distribution and properties of epenthetic vowels[6] in word-initial consonant sequences. It is only more recently that instrumental studies have directly examined the phonetic consequences of syllabic affiliation.

Three sets of studies are relevant to the research we report on in this article. One strand of laboratory work has examined vowel compression in a number of syllable structures (Aldrich and Simonet 2019; Marchini and Ramsammy 2022a,b). For example, Aldrich and Simonet (2019) examined vowel duration of a mixture of real and nonce words with the templates *p*V*pa* (*e.g.*, *papa*), *p*VC*pa* (*e.g.*, *palpa*), *p*VCC*pa* (*e.g.*, *panspa*), and *p*CV*pa* (*e.g.*, *plapa* and *prapa*), where V in all cases was the target vowel. With regard to the forms that

---

[4] While most analyses agree that the well-formedness of the onset cluster is driven by sonority, there is less agreement on why /f/ is the only fricative to pattern with plosives. Both Martínez-Gil (2001) and Colina (2016) argue that /f/ is unspecified for the feature [continuant] in the phonological representation. In their analyses, [+continuant] contributes to sonority. Segments that are specified as [+continuant] are more sonorous than those with the opposite specification (like stops), or with underspecification (like /f/). Thus, /f/ belongs in the same class as stops, which are uncontroversially [-continuant].

[5] The word-initial sequence */dl/ is not observed in any dialect, but /tl/ is observed in many varieties of Latin America (Hualde 2005), likely due to contact with languages where it is a licit sequence, such as Nahuatl.

[6] These vowels might very well be intrusive or excrescent vowels (Bradley 2006; Colantoni and Steele 2005; Gili y Gaya 1921) — we take no position on this distinction here, as it is not relevant to our current purposes.

are of interest for the present study, they found that nuclear vowels in *p*CV*pa* were systematically shorter when preceded by a word-initial consonant sequence. That is, the presence of two consonants –a complex onset– caused an acoustic shortening of the nuclear vowel, thus offering evidence for acoustic cues signaling syllabic affiliation. Note, however, the vowel of interest in their study was always in the word-initial syllable. As a result, complex onsets were only examined word-initially. We build on Aldrich and Simonet (2019) by precisely including these word-medial [fC sC] sequences to further probe the acoustic correlates of syllable affiliation of consonant sequences.

Another strand of work examines consonant durations in different positions of the word. Prieto (2002) showed that lateral consonants in complex onsets (*e.g.*, /bl/ in <obligada> "obligated.FEM.SG") were shorter than in words with intervocalic simple onsets (*e.g.*, /l/ in <holograma> "hologram.MASC.SG"). Our research design (see Section 5.1) follows up on this issue as well, particularly with respect to word-medial /fC sC/ sequences.

Finally, articulatory work examining the temporal coordination of consonantal gestures has also probed syllabic affiliation of Spanish consonant sequences. Sotiropoulou et al. (2020) observed a number of gestural adjustments. There are two key insights of this research. First, regarding consonant-lateral sequences, they found that prevocalic laterals in word-initial consonant sequences (*e.g.*, <plomo> "lead.MASC.SG") were shorter compared to their singleton counterparts (*e.g.*, <lomo> "back.MASC.SG"), similarly to Prieto (2002). Additionally, they found that the relative timing between the release of [l] and the maximum opening of the nuclear vowel decreased with the tautosyllabic sequences. In their view, these two adjustments result in an increase in the overlap between the vowel and the onset consonant cluster. Second, they argue that their findings also suggest that information regarding syllable structure, or even onset organisation, is observable in an array of "global" gestural adjustments, and not an individual gesture necessarily.

4 Linking hypotheses used in this paper

Based on the discussion in the preceding sections, two mutually compatible linking hypotheses seem to be available to probe consonant sequences, which we present in (1). The first linking hypothesis comes directly from the discussion on temporal stability metrics. The second linking hypothesis comes from prior experimental work on consonant sequences in Spanish as discussed in Section 3, though it has also been found to be useful in distinguishing alignment patterns in American English and Jazani Arabic (Durvasula 2023). Note, the second linking hypothesis is also consistent with the first — Sotiropoulou et al. (2020) suggest the relevant temporal stability patterns are only a part of the set of correlates that distinguish *global* timing stability in complex onset languages from *local* timing stability in simple onset languages, and really the distinction between the two is simultaneously expressed over a set of different phonetic parameters rather than just through a single measure such as C-CENTRE-TO-ANCHOR or RIGHT-EDGE-TO-ANCHOR interval stability. It is important to re-iterate, based on the review of the literature in the previous sections, that both these potential linking hypotheses are meant to be true of languages in general, and not specific to just Spanish.

(1) Linking hypotheses

(a) **Linking hypothesis 1:** A consonant sequence that is a complex onset will show C-CENTRE-TO-ANCHOR interval stability, while one that is a simplex onset will show RIGHT-EDGE-TO-ANCHOR interval stability.

(b) **Linking hypothesis 2:** The second member (C_2) will be shorter in a word containing a C_1C_2 sequence than in a minimal-pair word without C_1, if the sequence is a complex onset in the original word.

Based on the above linking hypotheses, word-initial and word-medial /fl/ sequences should show a C-CENTRE-TO-ANCHOR interval stability and the [l] should shorten when compared to stimuli without the preceding [f]. Similarly, the word-medial sC sequence should show RIGHT-EDGE-TO-ANCHOR interval stability and the C following the /s/ should not shorten when compared to stimuli without the preceding /s/.

These are the hypotheses the study was designed to test. However, as we will point out in later sections, they are actually difficult to maintain without further elaboration of the effect of poly-constituent shortening. There is evidence from many languages that as words become longer, individual segments exhibit shortening (Cuenca 1997; Farnetani and Kori 1986; Fowler 1981; Katz 2012; Marin and Pouplier 2010; Munhall et al. 1992). Therefore, shortening of C_2 (linking hypothesis 2) maybe an artefact of a longer word (by one segment here). As a result, the stability for the RIGHT-EDGE-TO-ANCHOR interval would decrease as well, unrelated to the underlying syllable structure. Thus, an observed C-CENTRE-TO-ANCHOR interval stability (the putative phonetic signature of complex onsets), could actually be masking a true RIGHT-EDGE alignment. Sections 7 and 8 expand on this issue.

The predictions stemming from the above hypotheses are also affected by domain-initial lengthening, whereby segments at the beginning of higher prosodic domains are longer (Cho et al. 2003; Fougeron and Keating 1997). Specifically, word-initial segments may be lengthened due to domain-initial lengthening, an effect that is reduced or absent for consonants in non-initial positions. Consequently, a word-initial consonant may appear to shorten when in a word-initial complex onset (singleton /l/ vs. /l/ in a Cl sequence, for example). Additionally, the shortening of the consonant increases the stability measure for

RIGHT-EDGE-TO-ANCHOR. Such an interaction could account for the observation of C-CENTRE-TO-ANCHOR stability word-initially, *i.e.*, this interaction could account for the apparent C-CENTRE alignment at word-initial positions. Although domain-initial lengthening is a potential confound, our experimental stimulus design does not allow us to observe it independently from the effect under study. We address this issue further in Section 9, where where we discuss its implications for our overall findings.

We first present two experiments that use the above linking hypotheses to study four different consonant sequences in Spanish: word-initial /fl/ sequences, and word-medial /fl/ and /sl/ in Experiment 1, and word-medial /sm/ sequences in Experiment 2.

5 Experiment 1

5.1 Methods

5.1.1 Participants

Prior to data collection, a pre-selection study was set up in Prolific (www.prolific.co) for native speakers of north-central Peninsular Spain (self-reported). North-central Peninsular Spanish was targeted because it is a [s]-preserving dialect (Hualde 2005), which was crucial for ease of consistent demarcation (see below), and the availability of participants in the Prolific platform. Note, in Prolific, we were able to limit the participant pool to those who resided in Spain, held Spanish nationality and reported Spanish as their L1. Additionally, they also reported having lived in Spain their whole lives. The participants were invited to submit recordings of the words in Table 2.

We took inspiration from Durvasula et al. (2021) in deploying the study over the internet. However, we went beyond the procedure established by them by adding a pre-selection task, for two reasons: (a) we wanted to ensure that we got sufficiently high-quality recordings for us to be able to annotate our recordings for segment boundaries with little confusion; (b) we wanted to ensure that the speakers were indeed from our target population. The stimuli for the pre-selection in Table 2 were chosen specifically to ensure that the speakers were from an /s/ preserving dialect and exhibited /θ/ in their pronunciations, as would be expected for speakers of the target dialect. Of the 31 submissions, 11 were selected based on the quality of the recordings. The spectrograms in Figures 2-5 are representative of the overall corpus of analysis.

Item	IPA transcripction	Gloss
<felicidad>	/feliθidad/	happiness.FEM.SG
<ensaladas>	/ensaladas/	salads.FEM.PL
<familia>	/familia/	family.FEM.SG
<tristeza>	/tristeθa/	sadness.FEMSG
<gente>	/xente/	people.FEM.SG
<formación>	/formaθion/	formation.FEM.SG
<pasta>	/pasta/	pasta.FEM.SG
<jirafas>	/xiɾafas/	giraffes.FEM.PL
<español>	/espaɲol/	Spanish.MASC.SG
<incompletas>	/inkompletas/	incomplete.FEM.PL
<jabalí>	/xabali/	wild boar.MASC.PL

Table 2: Pre-selection stimuli

5.2 Materials

Stimuli for the actual experiment consisted of disyllabic real and nonce Spanish words, with a word-initial or word-medial sequence of the form $C_1C_2V(C_3)$, and penultimate stress. In all cases, we were careful in choosing words where there was no morpheme boundary (or the appearance of one for nonce words) between the consonants in the target sequence. For word-initial sequences, the C_1 was /f/, C_2 was /l/, V was either /a/ or /o/ and C_3 was /s/; and for word-medial cases, C_1 could also be /s/. Thus, the consonant sequence of interest was in word-initial or word-medial positions (see Table 3). Their paired single-consonant words included a first or final syllable in the form of $C_1V(C_2)$, where C_1 =/l/, V = /a o/ and C_2 = /s/. Thus, a word like <flaca> "skinny.FEM.SG" was paired with <laca> "lacquer.FEM.SG" and <naflas> (nonce word) was paired with <nalas> (nonce word).

We used /f/ and /s/ as C_1 because: (a) their acoustic boundaries are easier to demarcate, in comparison to segments such as stops, and (b) because we assumed that they allowed us to probe both cases of tautosyllabic and hetorosyllabic consonant sequences. The lateral was chosen as C_2 in order to avoid the epenthetic vowels often found with production of Spanish /Cɾ/ sequences (Bradley 2006; Colantoni and Steele 2005). Additionally, using the rhotic as C_2 would not allow the comparison of word-internal and word-initial contexts, since only the trill is licit in word-initial position whereas only the tap is in complex onsets. When in word-final position, the addition of C_3 (C_2 in the single-consonant member of the pair; always /s/) ensured that the offset of the vowel following the crucial consonant sequence could be easily identified. Nonce words were included in the study because perfect Spanish minimal pairs were not always possible. There were no observations of productions of these nonce words with variable stress. A set of fillers items, both real and nonce, were also included to be used in a different study.

	Real word pair	Gloss	Nonce word pair
	<flaca, laca>	skinny.FEM.SG, lacquer.FEM.SG	<flato, lato>
/fl/ word-initial	/ˈfla.ka/, /ˈla.ka/		/ˈfla.to/, /ˈla.to/
	<flan, lan>	flan.MASC, LAN.FEM.SG	<flape, lape>
	/ˈflan/, /ˈlan/		/ˈfla.pe/, /ˈla.pe/
	<flote, lote>	float.3SG.SJV, plot-of-land.MASC.SG	<floque, loque>
	/ˈflo.te/, /ˈlo.te/		/ˈflo.ke/, /ˈlo.ke/
			<naflas, nalas>
/fl/ word-medial			/ˈna.flas/, /ˈna.las/
			<baflos, balos>
			/ˈba.flos/, /ˈba.los/
			<goflas, golas>
			/ˈgo.flas/, /ˈgo.las/
	<muslos, mulos>	thigh.MASC.PL, mule.MASC.PL	<queslas, quelas>
/sl/ word-medial	/ˈmus.los/, /ˈmu.los/		/ˈkes.las/, /ˈke.las/
	<islas, hilas>	island.MASC.PL,thread.2SG	<toslos, tolos>
	/ˈis.las/, /ˈi.las/		/ˈtos.los/, /ˈto.los/
	<teslas, telas>	Tesla.PL, fabric.MASC.PL	<poslas, polas>
	/ˈtes.las/, /ˈte.las/		/ˈpos.las/, /ˈpo.las/

Table 3: Experiment 1 Stimuli

5.3 Procedure

After recruitment on Prolific, participants were directed to a survey on JotForm.[7] We chose this platform because it includes a participant-friendly recording widget. Participants read five repetitions of test items in Table 3 (as well as fillers).[8] Each repetition was its own pseudo-randomised list, where the words were

---

[7] The experiment interface is viewable here: [link redacted for anonymity]

[8] Previous (articulatory) studies have many more repetitions, but our concern was that such repetitions are almost always assumed to be "independent" repetitions. However, given that they are produced by the same participant in the same experimental session, that assumption is likely wrong. When the assumption of independence is violated, mean and standard deviation estimates may not be trustworthy. At the same time, we do need some repetitions to get an estimate of the standard deviation and the mean needed for RSD calculations, so using 1 repetition was untenable (which would actually be statistically ideal to avoid independence violations). One would be right to worry about a power issue; however, we don't believe this is the case in the present study. To foreshadow our results, our findings are quite consistent across all measures and across all the tested cases. This would not usually be the case if power was an issue. If there were a power issue, the chance of null results would be much higher, and in cases where there is a statistically clear effect, there should be large inconsistencies in magnitude (Type M errors) and even sign (Type S errors) of the effect, across cases (Gelman and Carlin 2014). This was not the issue in our results, as we will show. Of course, the interpretation of the results is a separate issue, and there we provide extended discussion of why the results shouldn't be taken at face value.

presented in blocks of about 10. The total of tokens produced by speaker was 270 (54 stimuli x 5 repetitions), of which 150 were test items (30 stimuli x 5 repetitions).

5.4 Measurements

The recordings were first automatically forced-aligned using the Montreal Forced Aligner (McAuliffe et al. 2017), and then the annotations were manually corrected by both authors in Praat (Boersma and Weenink 2023). One author corrected repetitions one, two and three, while the other author corrected repetitions three, four and five. The third repetition, annotated by both authors, allowed us to test for annotation reliability (see below). The annotation contained three tiers: a word tier, a phone tier and a quality tier. If participants misread the item, or there were any other types of disfluencies (*e.g.*, a pause), the token was marked as 'bad'; it was otherwise kept empty. Approximately 11.5% of the data was removed due to poor quality (defined here as 'bad/weird' in our annotation). During the first phase of annotation, the label 'unclear' was used for cases in which boundaries between segments were not straightforward, and the label 'weird' was used for cases where the production was unexpected based on the stimulus prompt. These cases were reviewed by both authors. There were no cases of disagreement. After these exclusions, a total of 1328 observations were submitted for analysis.

For each token, the focus of annotation was the target C_1, C_2, V, and C_3. The interval for the fricative was identified by the presence of a noisy spectrum; the offset of the interval was identified as the point in which the onset of formant structure for the following segment was observable. The interval for [l], which could be word-initial or word-internal, was identified at the start clear onset of voiced waveform and formants. Finally, the vowel interval was identified based on the strong presence of formant structure, and in the waveform, a noticeably higher intensity when compared to its surrounding segments. When followed by /s/ (*e.g.*, in the word-final syllable, like in <muslos> "thigh.MASC.PL", <mulos> "mule.MASC.PL"), the vowel often included a voiceless period prior to the onset of frication. This period was included as part of the vocalic interval. Figures 2 through 5 present sample of the annotation schemes.

As noted above, in order to determine inconsistencies across annotators, the third repetition was corrected by both authors and their percentage of agreement was calculated at various levels. Segmental boundaries were within 5ms of each other in 54% of the cases, within 10ms in 79%, and within 15ms in 89%. Given these numbers, we assume that there was reasonable consistency across authors in annotation.

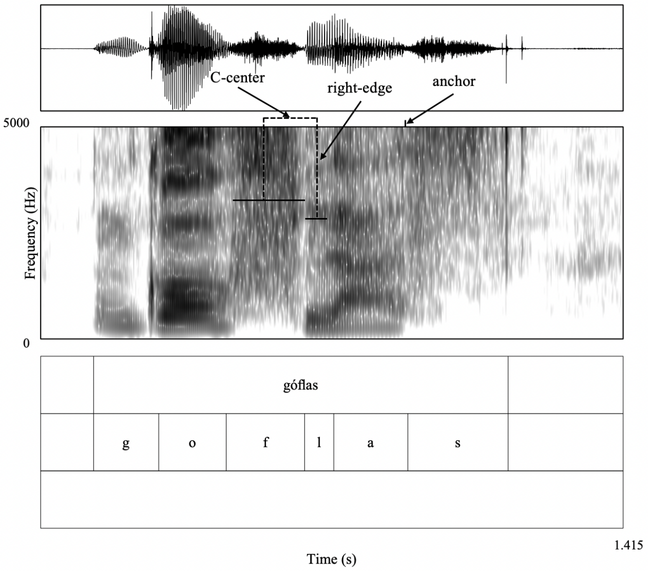


Figure 2: Sample annotation of the nonce word <góflas>. c-centre-to-anchor, right-edge-to-anchor and anchor landmarks are indicated in the figure

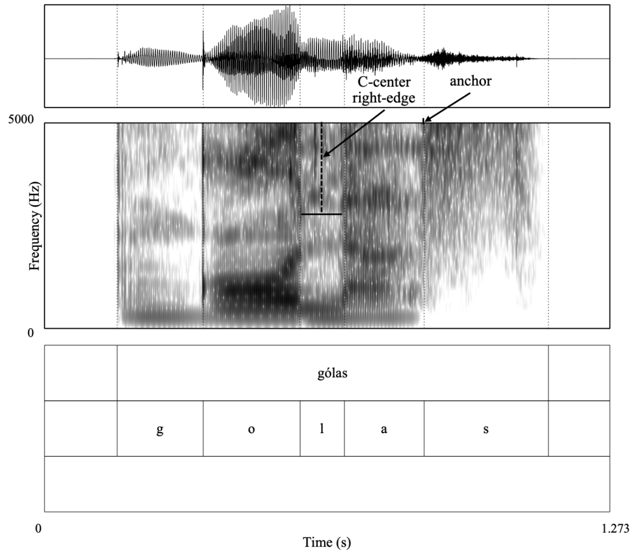


Figure 3: Sample annotation of the nonce word <golas>. c-centre-to-anchor, right-edge-to-anchor and anchor landmarks are indicated in the figure

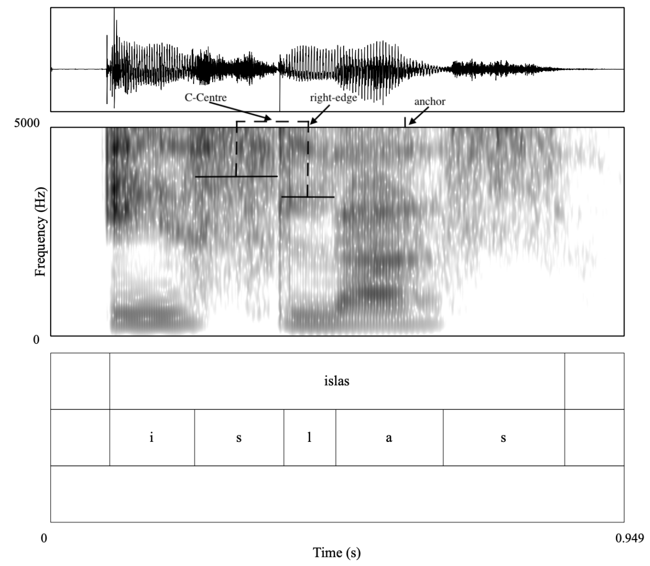


Figure 4: Sample annotation of the word <islas> ‘islands’. C-CENTRE-TO-ANCHOR, RIGHT-EDGE-TO-ANCHOR and ANCHOR landmarks are indicated in the figure

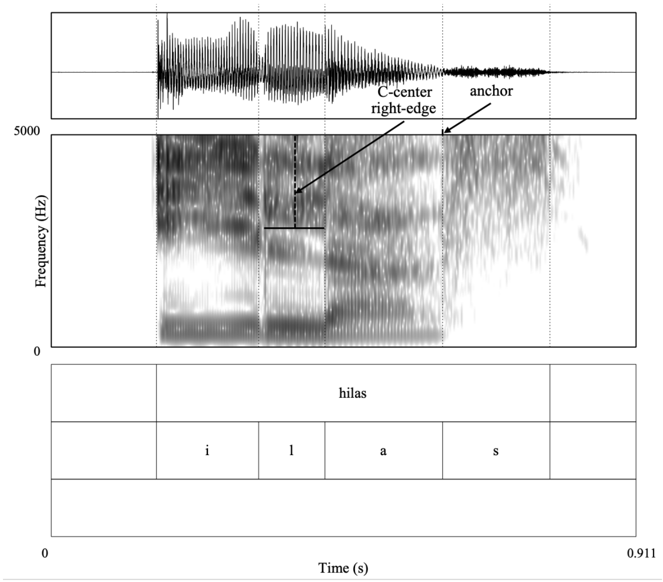


Figure 5: Sample annotation of the word <hilas> ‘thread.2SG’. C-CENTRE-TO-ANCHOR, RIGHT-EDGE-TO-ANCHOR and ANCHOR landmarks are indicated in the figure

We then extracted: (a) the duration from the mid-point of the right-most prevocalic consonant to the end of the following vowel (i.e, RIGHT-EDGE-TO-ANCHOR), (b) the duration from the mean of the mid-points of the word-initial consonants to the end of the following vowel (i.e., C-CENTRE-TO-ANCHOR).

Following Shaw et al. (2009, 2011), we calculated the Relative Standard Deviation (RSD) of the durations for each pair of words using the formula below (Equation 1) to estimate the spread, and therefore the stability, in the durations. To measure the RSD for each pair of words for each participant, we used all the repetitions of the pair produced by the participant. Note, we used RSDs as our measure of stability since they have been argued to control for the larger variance that is typically associated with longer durations; in contrast, an uncorrected measure such as standard deviation or variance would have an inherent bias against measures involving longer durations (in this case, C-CENTRE-TO-ANCHOR) over those involving shorter durations (RIGHT-EDGE-TO-ANCHOR).

$$\text{Relative Standard Deviation (RSD)} = \frac{\text{Standard Deviation} * 100}{\text{Mean}} \tag{1}$$

In our data, the RSD value was calculated for the C-CENTRE-TO-ANCHOR interval and RIGHT-EDGE-TO-ANCHOR interval separately. For each interval, within each word-pair and subject, the standard deviation of all the interval durations was calculated, and then divided by the mean of all the interval durations. The RSD value is expected to be higher when there is a greater increase in interval duration between words with consonantal sequences and those without, and is expected to be 0 when there is absolutely no variation in interval durations.

In cases where the underlying syllable co-ordination is one of C-CENTRE (complex) alignment, the RSD of the C-CENTRE-TO-ANCHOR interval is expected to be generally lower than that of the RIGHT-EDGE-TO-ANCHOR interval. In cases where the underlying syllable co-ordination is one of RIGHT-EDGE (simplex) alignment, the RSD of the RIGHT-EDGE-TO-ANCHOR interval is expected to be generally lower than that of the C-CENTRE-TO-ANCHOR interval. However, as Shaw et al. (2009) point out through simulations, the latter expectation needs more nuance. The RSDs for C-CENTRE-TO-ANCHOR interval stability can be lower than the RSDs for RIGHT-EDGE-TO-ANCHOR interval stability *if there is sufficient variance in the durational mesarurements.* Therefore, with sufficient variance in durations, a lower RSD value for C-CENTRE-TO-ANCHOR interval stability becomes difficult to interpret. In contrast, if the RSD for the RIGHT-EDGE-TO-ANCHOR interval is generally lower in the results, then that can be interpreted as evidence in favour of RIGHT-EDGE (simplex) alignment.

It is important to note here that Shaw and Gafos's (2015) and Shaw et al.'s (2009) result was a simulation with a very specific set of parametric estimates stemming from the observations in their dataset. As noted in their work and in Gafos et al. (2014), one needs parametric estimates from the relevant production study to identify the threshold/tipping point variance value. Therefore, we are not able to use an a priori threshold to see if the issue they raise is a problem for our results. The issue is further compounded by the phenomenon of poly-constituent shortening that we discussed above and return to in Section 7, which was not part of their simulations. To resolve this issue, we present our own simulations using parametric estimates obtained from the data presented in this manuscript in Section 8, and request the reader's indulgence for now.

5.5 Results

All plotting and statistical modelling in this article were done using the programming language R (R Core Team 2021) within the Rstudio IDE (RStudio Team 2020). The plotting and data munging were done using the package `tidyverse` (Wickham 2017). For each experiment, we first visually inspected the results and then followed the visual inspection with linear mixed effects modelling with the packages `lme4` (Bates et al. 2015) and `lmerTest` (Kuznetsova et al. 2017). Finally, the statistical models were converted to LaTeX code using the package `stargazer` (Hlavac 2018).

As can be seen in Figure 6, the overall RSDs for each pair of consonants are lower for the c-centre-to-anchor interval.

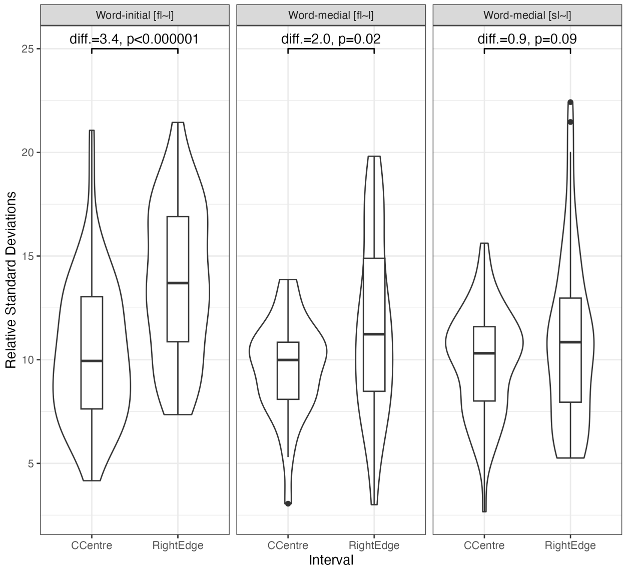


Figure 6: Overall Relative Standard Deviations (RSDs) for Experiment 1. Each boxplot/violin plot represents the RSDs calculated for each CC$\sim$ C pair (left of each facet = RSDs of c-centre-to-anchor durations, right of each facet = RSDs of right-edge-to-anchor durations). Included statistics are based on mixed-effects models discussed in the prose.

We followed up on the visual inspection of the overall RSDs with linear mixed-effects modelling. As mentioned above, the crucial dependent variable when looking at interval stability is the Relative Standard Deviations (RSD), and the independent variable considered was interval (c-centre-to-anchor, right-edge-to-anchor; baseline = c-centre-to-anchor). The random-effects structure included a random intercept of participant, word-pair, and nonce status.[9] We modelled each consonant pair case separately in order to see which of the two intervals was more stable. The modelling results are presented in Table 4.

As can be seen in Table 4, the right-edge-to-anchor interval had a higher RSD value for each consonant pair. In the case of /fl∼l/, there is a statistically clear difference between the RSD values of the c-centre-to-anchor interval and the right-edge-to-anchor interval.[10] In the case of the word-medial /sl∼l/, although the difference is in the same direction, the difference is statistically not clear.[11]

Table 4: Linear mixed-effects models for each consonant pair in Experiment 1 (reference: c-centre)

Consonant Pair	Position		Estimate	S.E.	df	t-value	Pr(<|t|)
fl∼l	Word-initial	(Intercept)	10.4	0.5	25.0	20.8	< 0.00001
		right-edge	3.4	0.6	109.0	5.3	< 0.00001
fl∼l	Word-medial	(Intercept)	9.6	0.7	7.5	12.9	< 0.00001
		right-edge	2.0	0.8	47.0	2.4	0.02
sl∼l	Word-medial	(Intercept)	9.9	0.7	13.0	14.9	< 0.00001
		right-edge	0.9	0.5	109.0	1.7	0.09

We then looked at the acoustic durations of the lateral consonant in both members of each pair. As a reminder, Prieto (2002) and Sotiropoulou et al. (2020) observe that the pre-vocalic consonant is shorter in the case of a complex onset alignment. The durations in Figure 7 suggest that there is a general shortening of the pre-vocalic consonant in words with a sequence of two consonants in the relevant position for all three cases.

---

[9] We maintain the same random-effects structure for each consonant pair in both experiments. Only random intercepts were considered because with random slopes the model did not converge across all the models.

[10] Throughout, we use the phrase "statistically clear difference" (and variants) instead of statistically significant, and the phrase "statistically unclear difference" (and variants) instead of statistically non-significant on the recommendation of Dushoff et al. (2019). They make this recommendation in order to avoid the common confusion between statistical significance and practical significance.

[11] On the recommendation of a reviewer, we fitted a post-hoc model to all the fl∼l nonce word pairs, with interval and position as fixed effects, and the same random effects structure as the other models. There was no clear interaction of interval and position ($\hat{\beta}$=-0.4215, p=0.7). This suggests that there is no clear evidence that the word-medial and word-initial cases have different estimates for the difference between the RSDs for c-centre-to-anchor interval and right-edge-to-anchor interval.

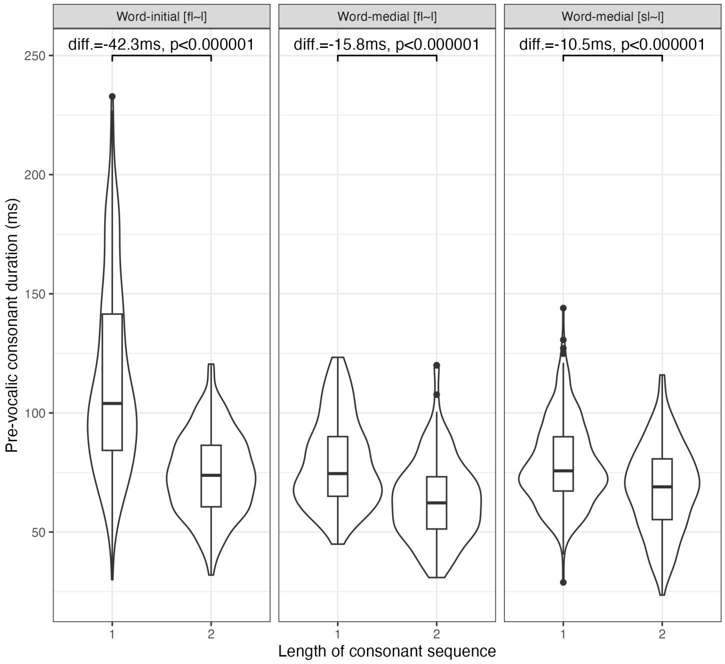


Figure 7: Durations of the prevocalic consonants for the consonant pairs in Experiment 1. Included statistics are based on mixed-effects models discussed in the prose.

Again, we followed up the visual inspection with linear mixed-effects modelling. The dependent variable was the duration of the pre-vocalic consonant, and the independent variable considered was LENGTH OF CONSONANT SEQUENCE (Length-1, Length-2; baseline = Length-1). The random-effects structure included a random intercept of participant, and word-pair.[12] The modelling results are presented in Table 5. The results showed a statistically clear shortening of the pre-vocalic consonant in the consonant sequences for all three cases. Note further, that the unstandardised effect sizes in each of the cases observable in the Table (namely, -42.3 ms, -15.8 ms, -10.5 ms) are substantial, and are therefore likely to be practically significant.

---

[12] Again, we maintain the same random-effects structure for each consonant pair in both experiments.

Table 5: Linear mixed-effects models for the duration of the pre-vocalic consonant for each consonant pair in Experiment 1 (reference: c-centre)

Consonant Pair	Position		Estimate (ms)	S.E.	df	t-value	Pr(<|t|)
fl∼l	Word-initial	(Intercept)	156.1	8.0	11.7	19.5	< 0.00001
		Length-2	-42.3	2.0	498.2	-21.6	< 0.00001
fl∼l	Word-medial	(Intercept)	94.8	4.7	16.8	20.0	< 0.00001
		Length-2	-15.8	1.6	258.0	-9.8	< 0.00001
sl∼l	Word-medial	(Intercept)	89.3	4.3	13.7	20.7	< 0.00001
		Length-2	-10.5	1.2	532.5	-8.8	< 0.00001

On the recommendation of a reviewer, we fitted a post-hoc model to all the fl∼l nonce word pairs, with length of consonant sequence and position and the interaction between them as fixed effects, and the same random effects structure as the above models (Table 6). There is a statistically clear interaction between length of consonant sequence and position, suggesting that the degree of C_2 shortening observed in the word-initial case is larger than in the word-medial case.

Table 6: Linear mixed-effects models for the duration of the pre-vocalic consonant for fl∼l nonce words in Experiment 1 (reference: c-centre, Word-medial)

	Estimate (ms)	S.E.	df	t-value	Pr(<|t|)
(Intercept)	95.0	6.4	20.1	14.8	< 0.00001
Length-2	-16.2	2.4	511.0	-6.7	< 0.00001
Position (Word-initial)	59.8	5.4	511.1	11.1	< 0.00001
Length-2 : Position (Word-initial)	-25.1	3.5	511.2	-7.1	< 0.00001

5.6 Discussion

The results of the current experiment suggest that the c-centre-to-anchor interval is more stable than the right-edge-to-anchor interval for both word-initial and word-medial /fl/ sequences. Furthermore, there is some statistically unclear evidence that the same is true for word-medial /sl/ sequences. Based on the linking hypotheses, the pattern of interval stability in turn would be interpreted as showing that both word-initial and word-medial /fl/ sequences are consistent with being complex onset clusters, while there is no statistically clear difference in the stability results for the word-medial /sl/ case.

The results from the analysis of the duration of the pre-vocalic consonant also would have a similar inference. In all three cases, the presence of a preceding consonant shortens the pre-vocalic consonant. The pattern of pre-vocalic consonant duration shortening is consistent with a pattern of complex onset clusters for all three cases, despite standard analyses of tautosyllabic vs. heterosyllabic constituency.

One potential reason for the lack of a clear statistical difference with /sl/ when looking at the stability metrics is the general issue of annotating the boundary between /sl/. More specifically, there were cases of short acoustic silences between the /s/ and the /l/, which we interpreted as excrescent stops. While we consistently included them as part of the preceding fricative, one could argue that this is inappropriate. Our worry was that there is no "right" annotation scheme in such cases, and so we simply proceeded in a consistent fashion. This issue could however have led to noisy measurements of the relevant intervals for the /sl/ case, thereby affecting the effect size and the p-value calculations.

It is for this reason that we turn to word-medial /sm/ sequences in Experiment 2.

6 Experiment 2

In Experiment 1, we generally observed that the consonantal sequences considered generally appeared to show C-CENTRE alignment, both word-medially and word-initially. However, the interval stability pattern had no clear statistically clear difference in the case of /sl/ sequences. As stated above, we conjectured that a part of the problem with the sequences was with annotating the boundaries between /sl/. For this reason, in this experiment, we turned to /sm∼m/ pairs. Additionally, given that word-medial effects in Experiment 1 were notably smaller than word-initially, in Experiment 2 we only examine word-medial cases.

6.1 Methods

6.1.1 Participants

Participants in Experiment 1, and only those participants, were invited to take part in Experiment 2. The participants were re-invited for this experiment through Prolific and directed to a survey on JotForm[13]. Of the original eleven subjects in Experiment 1, eight returned; therefore, Experiment 2 had eight participants. Participant label was retained; that is, Subject 1 in the first experiment is Subject 1 in this second one.

6.2 Materials

Stimuli consisted of real words, with word-medial /sm/ sequences and the singleton counterpart /m/, and penultimate stress. We used this sequence, instead of /sn/ due to availability of real word pairs. Table 7

---

[13] The experiment interface is viewable here: [link redacted for anonymity]

presents the list of target words. A set of fillers of real words were also included to be used in a different study (see Table 13 in the appendix.)

Real word pair	IPA transcription	Gloss
<cosmos, cómos>	/ˈkos.mos/, /ˈko.mos/	cosmos.MASC.SG, how.MASC.PL
<cismas, cimas>	/ˈθis.mas/, /ˈθi.mas/	schism.MASC.PL, summit.MASC.PL
<asmas, amas>	/ˈas.mas/, /ˈa.mas/	asthma.MASC.SG, love.2SG.IND
<husmeas, humeas>	/us.ˈme.as/, /u.ˈme.as/	sniff.2SG.IND, smoke.2SG.IND
<mismas, mimas>	/ˈmis.mas/, /ˈmi.mas/	same.FEM.PL, pamper.2SG.IND
<mismos, mimos>	/ˈmis.mos/, /ˈmi.mos/	same.MASC.PL, cuddle.MASC.PL

Table 7: Experiment 2 Stimuli

6.2.1 Procedure

Participants read five repetitions of the stimuli in isolation, each its own pseudo-randomised list, presented one at a time. Each participant produced a total of 135 test items (27 stimuli x 5 repetitions), of which 60 were the test words analysed for this experiment (12 stimuli x 5 repetitions).

6.2.2 Measurements

Following Experiment 1, recordings were automatically forced-aligned with the Montreal Forced Aligner (McAuliffe et al. 2017) and manually corrected by both authors. For revising the automatic annotation, the same division of labour among authors was followed in this experiment as well (repetitions one, two and three corrected by one author, repetitions three, four and five by the other one, with repetition three serving as inter-annotator reliability). The annotation scheme of three tiers (word, segment and quality) used in Experiment 1 was adopted here as well. The same measurements as in Experiment 1 were taken for Experiment 2. Approximately 9.5% of the data was removed due to poor quality. A total of 431 observations were submitted for analysis.

Much like in Experiment 1, the third repetition allowed to assess inter-annotator reliability. Within 5ms of each other, annotators were in 51% agreement; within 10ms, 78%; and within 15ms, 90%.

6.3 Results

As with all the consonant pairs in Experiment 1, it can be seen in Figure 8 that the RSDs for the word-medial /sm∼m/ pair are generally lower for the C-CENTRE-TO-ANCHOR interval.

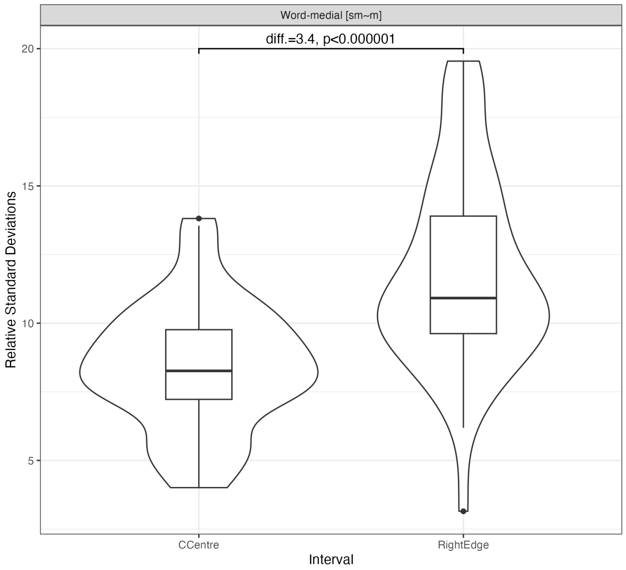


Figure 8: Overall Relative Standard Deviations for Experiment 2. Included statistics are based on mixed-effects models discussed in the prose.

Again, as with the data in Experiment 1, we followed up on the inspection of the overall RSDs with linear mixed-effects modelling, where the crucial dependent variable is the Relative Standard Deviations (RSD), and the independent variable considered was interval (c-centre-to-anchor, right-edge-to-anchor; baseline = c-centre-to-anchor). The random-effects structure included a random intercept of participant and word-pair.

The right-edge-to-anchor interval had a higher RSD than the c-centre-to-anchor interval for the word-medial /sm∼m/ pair, and the difference was statistically clear (see Table 8).

Table 8: Linear mixed-effects models for the word-medial /sm∼m/ pair in Experiment 2

	Estimate	S.E.	df	t-value	Pr(<|t|)
(Intercept)	8.4	0.7	11.7	12.7	< 0.00001
right-edge	3.4	0.5	82	6.3	< 0.00001

As with Experiment 1, we then looked at the acoustic duration of the consonant immediately before the crucial vowel. The durations in Figure 9 again suggest that there is a general shortening of the pre-vocalic consonant, *i.e.*, /m/, in word-medial /sm/ cases.

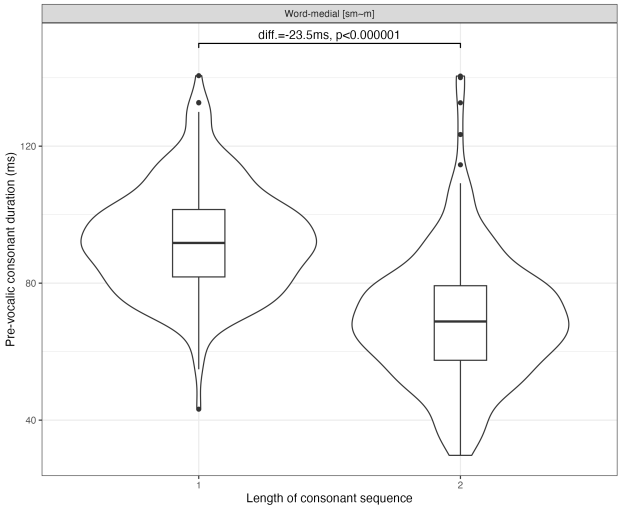


Figure 9: Durations of the prevocalic consonants for the word-medial [sm∼m] pair in Experiment 2. Included statistics are based on mixed-effects models discussed in the prose.

We followed up the visual inspection with linear mixed-effects modelling of the pre-vocalic consonant as the dependent variable, and the Length of consonant sequence (Length-1, Length-2; baseline = Length-1) as the independent variable. The random-effects structure included a random intercept of participant, and word-pair. The results showed a statistically clear shortening of the pre-vocalic vowel in the consonant sequences for both cases (see Table 9). Note further that, as with Experiment 1, the unstandardised effect size (namely, -23.5 ms) is substantial, and is therefore likely to be practically significant.

Table 9: Linear mixed-effects models for the duration of the pre-vocalic consonant for the word-medial [sm∼m] pair in Experiment 2 (reference: Length-1)

	Estimate (ms)	S.E.	df	t-value	Pr(<|t|)
(Intercept)	116.3	3.3	33.6	35.6	< 0.00001
Length-2	-23.5	1.6	419.3	-14.9	< 0.00001

6.4 Discussion

The results of the current experiment not only replicate the interval stability patterns observed in Experiment 1, in that c-centre-to-anchor interval is more stable than right-edge-to-anchor interval for both the word-initial and word-medial sequences, they also show the same pattern of shortening of the pre-vocalic consonant /m/ in /sm/ sequences. At first blush, both observed patterns appear to be consistent with the temporal stability patterns of word-medial complex onsets in other languages.

If correct, our results are quite intriguing, as they stand in contrast to standard analyses of syllable structure in Spanish. However, as we point out, such an inference would be premature as the analyses haven't controlled for potential confounds.

7 Post-hoc analyses reveal that the C-centre effect is confounded by poly-constituent shortening

In Experiments 1 and 2, /sm/ and /sl/ were observed to have a C-CENTRE-TO-ANCHOR stability and a shortening of the C_2 similar to complex onsets. This is superficially consistent with a C-CENTRE-TO-ANCHOR organisation. As per standard accounts /sm/ and /sl/ are hetero-syllabic, and consequently, our findings are quite surprising given the traditional analysis of syllabic affiliation in Spanish. One possible way to interpret our results is that the standard analysis of Spanish syllables, particularly that for word-medial /sm/ and /sl/ cases, is wrong, and that indeed such word-medial sequences are tauto-syllabic and form complex onsets. However, such an inference would be rather hasty, in our opinion, if the experimental probe has not been sufficiently vetted against possible confounds. For this reason, we wanted to explore if there were potential confounds that could explain our findings.

One potential confound discussed earlier is poly-constituent shortening, which is observed in many languages. As words become longer, individual segments tend to shorten (Cuenca 1997; Farnetani and Kori 1986; Fowler 1981; Katz 2012; Marin and Pouplier 2010; Munhall et al. 1992). Thus, the shortening of C_2 may simply be an artefact of a longer word (by one segment, in our case). If C_2 in CC sequences shortens due to poly-constituent shortening (or other factors), that could make the RIGHT-EDGE-TO-ANCHOR interval appear less stable across words with and without consonant sequences, accounting for reduced interval stability. In order to test this possibility, we examined the duration of non-target segments in each experiment. If these segments show no changes in acoustic durations, this confound can be ruled out.

A second potential confound brought up earlier is that of domain-initial lengthening (Cho et al. 2003; Fougeron and Keating 1997). As mentioned earlier, our experimental stimulus design doesn't allow us to observe domain-initial lengthening independent of the effect under study. However, we return to the issue in the conclusion (Section 9), where we suggest that the effect has a bearing on the full account of the facts.

7.1 Post-vocalic consonant durations

In this sub-section, we specifically look at the duration of the consonant following the crucial vowel (*e.g.*, /t/ in <flato∼lato>). Our hypothesis was that consonants in such positions should not be affected by standard C-CENTRE-based stability accounts of the C-CENTRE-TO-ANCHOR pattern, as the effect is local to the consonant sequence and the following vowel.

In the case of word-initial /fl∼l/ sequences, the post-vocalic consonant is standardly analyzed as part of the following syllable (*e.g.*, /ˈfla.**t**o/-/ˈla.**t**o/). In contrast, for the word-medial /fl∼l/ and /sl∼l/ sequences in Experiment 1 and the word-medial /sm∼m/ sequences in Experiment 2, the post-vocalic consonant is standardly analysed as part of the same syllable as the pre-vocalic consonant (*e.g.*, /ˈna.fla**s**/-/na.la**s**/, /ˈmus.lo**s**/-/ˈmu.lo**s**, and /ˈsis.ma**s**/-/ˈsi.ma**s**/, respectively).

A visual inspection of all the relevant cases suggested a consistent shortening of the post-vocalic consonant in all the cases we looked at (Figure 10.)

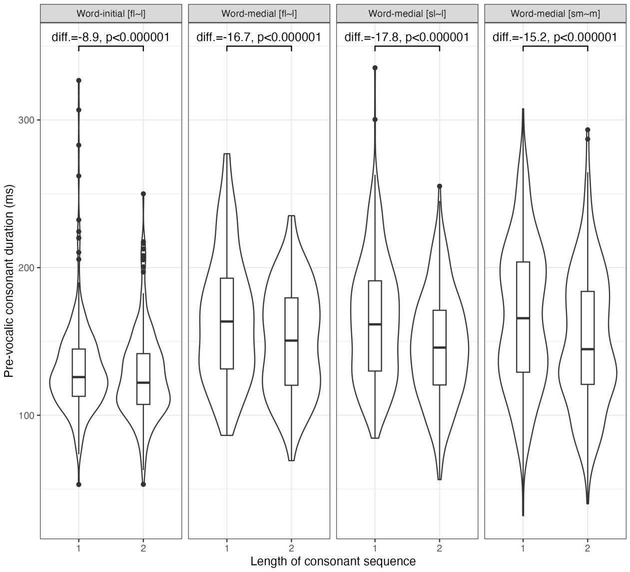


Figure 10: Durations of the post-vocalic consonants for the relevant pairs in Experiments 1 and 2. Included p-values are based on mixed-effects models discussed in the prose.

A linear mixed-effects model was fitted for each of the cases with the number of consonants in the target consonant sequence (LENGTH) as the independent variable, and random intercepts of participants and word-pair. These models are shown in Table 10. Importantly, irrespective of syllabic affiliation of the target consonant, each case shows a shortening effect correlated with the presence of an additional consonant in the target sequence. This suggests that the pre-vocalic consonant shortening observed in the main experiments is confounded by poly-constituent shortening across the whole word, *i.e.*, the pre-vocalic consonant shortening could have been from general poly-constituent shortening that applies beyond the syllable and not due to syllable-structure *per se*. Consequently, the C-CENTRE-TO-ANCHOR stability pattern observed may also be confounded by the same finding.

Table 10: Linear mixed-effects models for the duration of the word-initial consonant for the word-medial /fl∼l/ and /sl∼l/ pairs in Experiment 1, and the /sm∼m/ pair in Experiment 2 (reference: Length-1)

Experiment		Estimate (ms)	S.E.	df	t-value	Pr(<|t|)
Exp. 1 /fl∼l/ Word-initial	(Intercept)	140.7	8.1	14.2	17.3	< 0.0001
	Length	-8.9	1.9	493.4	-4.6	< 0.0001
Exp. 1 /fl∼l/ Word-medial	(Intercept)	181.0	11.6	12.3	15.6	< 0.0001
	Length	-16.7	3.0	260.0	-5.6	< 0.0001
Exp. 1 /sl∼l/	(Intercept)	181.6	10.7	11.1	17.0	< 0.0001
	Length	-17.8	2.1	532.4	-8.4	< 0.0001
Exp. 2 /sm∼m/	(Intercept)	182.4	15.4	8.0	11.9	< 0.0001
	Length	-15.2	2.5	419.2	-6.0	< 0.0001

7.2 Word-initial consonant durations when looking at medial C-CENTRE effects

In this post-hoc analysis, we looked at the word-initial consonant durations in the cases where the C-CENTRE effect being probed was word-medial. For Experiment 1, we chose the word-pairs /naflas∼nalas/, /muslo∼mulo/; and, for Experiment 2, we examined the word-pairs /mismas∼mimas/, /mismos∼mimos/. We specifically chose these words as they begin with a nasal consonant, for which the annotation of the acoustic onset/offset is cleaner than the other word-initial consonants (*e.g.*, *mulos* vs. *telas*) in our stimuli.

A visual inspection of all the relevant cases suggested a consistent shortening of the initial consonant in all the cases we looked at (Figure 11.)

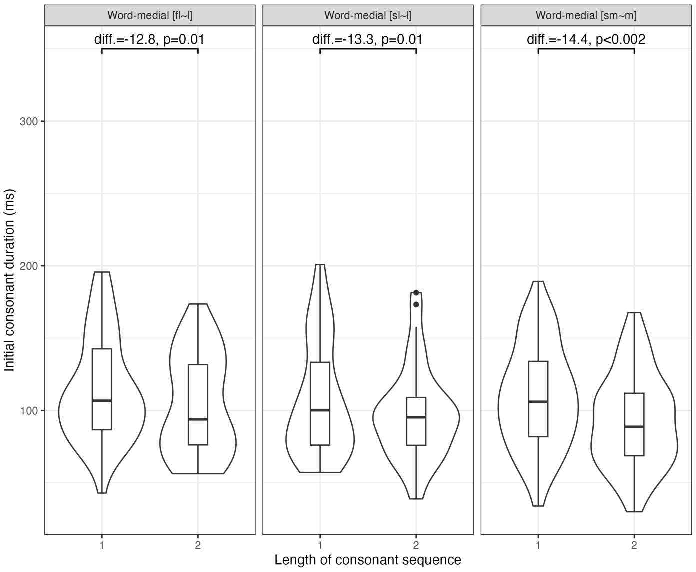


Figure 11: Durations of the initial consonants for the relevant pairs in Experiments 1 and 2. Included statistics are based on mixed-effects models discussed in the prose.

Table 11 shows the mixed effects model results for each of the pairs. There is a consistent shortening of the word-initial consonant in cases where there was an additional consonant word-medially. Crucially, this result cannot be attributed to syllabic structure, per se, as it is both outside the syllable under study and not even immediately adjacent to the C(C)V sequence under study. As with the previous post-hoc analysis, the results here suggest that the findings observed in the main experiments should be interpreted with caution, as they are consistent with a poly-constituent shortening effect from the presence of an additional segment in the word.

Table 11: Linear mixed-effects models for the duration of the word-initial consonant for the word-medial /fl∼l/ and /sl∼l/ pairs in Experiment 1, and the /sm∼m/ pair in Experiment 2 (reference: Length-1)

Experiment		Estimate (ms)	S.E.	df	t-value	Pr(<|t|)
Exp. 1 /fl∼l/	(Intercept)	127.5	11.5	25.8	11.1	0
	Length	-12.8	5.1	83.1	-2.5	0.01
Exp. 1 /sl∼l/	(Intercept)	120.8	11.8	22.9	10.2	0
	Length	-13.3	5.0	77.2	-2.7	0.01
Exp. 2 /sm∼m/	(Intercept)	122.8	10.5	18.9	11.7	0
	Length	-14.4	4.6	133.3	-3.1	0.002

8 Model simulations of onset alignment with poly-constituent shortening

As pointed out by Shaw and Gafos (2015) and Shaw et al. (2009), RSD measurements need to be understood in a nuanced way. They show through simulations that: (a) with sufficient variation, the RSD for C-CENTRE-TO-ANCHOR interval will be less than the RSD for RIGHT-EDGE-TO-ANCHOR interval, *even if the underlying structure actually has* RIGHT-EDGE *alignment*; (b) the RSD for C-CENTRE-TO-ANCHOR interval will always be less than that of RSD for RIGHT-EDGE-TO-ANCHOR interval, if the underlying structure has C-CENTRE alignment. That is, if there is sufficient variance in the duration measurements, a lower RSD value for C-CENTRE-TO-ANCHOR interval is difficult to interpret as evidence that aligns with a specific underlying stability pattern (C-CENTRE vs. RIGHT-EDGE alignment), but lower RSD value for RIGHT-EDGE-TO-ANCHOR interval is always interpretable as evidence that aligns with RIGHT-EDGE alignment. In short, with sufficient variation in durations, the RSD measure becomes an asymmetric evidentiary source.

However, Shaw and Gafos's (2015) and Shaw et al.'s (2009)'s inferences were based on simulations with a very specific set of parametric values stemming from the observations in their dataset; consequently, it is not clear if the issue they observe in (a) above is also true for our data. Furthermore, given our results in the previous sections, the inclusion of poly-constituency is important to assess whether their inferences hold in this study as well.

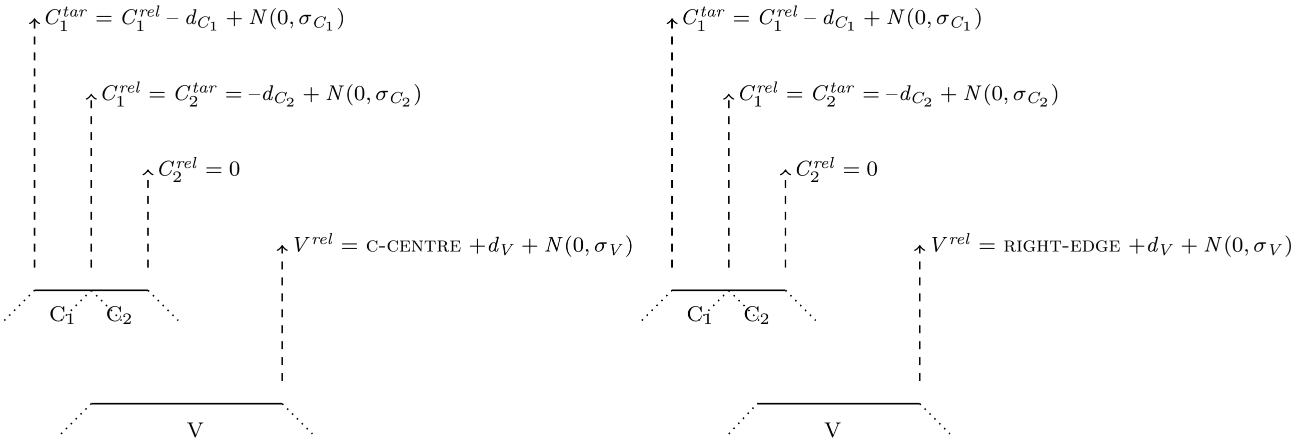
To address the above two concerns, we undertook modelling and simulations that include poly-constituent shortening, based on parameter estimates from our own results. We took inspiration from Shaw and Gafos (2015) and Shaw et al. (2009), who modeled C-CENTRE and RIGHT-EDGE alignment using estimates from articulatory data. But, since we used acoustic measurements, we had to modify their original model for our purposes. In order to estimate parameter values, we used average acoustic measurements from word-medial /fl∼l/ pairs.

In Figure 13, we show the specifics of the modeling. For modelling purposes, we follow Durvasula et al. (2021) in making the assumption that the consonantal acoustic intervals identified during annotation represent the achievement of the articulatory plateau (target to release).

We modelled the target achievement of the pre-vocalic consonant (C_2 in the figure) as the zero point. To generate the release of C_2, the average pre-vocalic (acoustic) consonant duration in VCV words was used as an estimate (d_{C_2}=79ms), with normally-distributed variation estimated as the standard deviation of such durations (σ_{C_2}=17ms). Given that in our acoustic measurements, in a CC sequence, the end of a consonant necessarily marks the onset of the following consonant, there is no inter-consonantal duration in our simulations (in contrast to Shaw et al. (2009)). Therefore, the release of C_1 corresponded to the 0 point; and to generate the target achievement of C_1, the average pre-consonantal consonant duration in CCV words was used as an estimate (d_{C_1}=130ms), with normally-distributed variation estimated as the standard deviation of such durations (σ_{C_1}=27ms).

The vowel duration estimation was a bit more complicated as both the RIGHT-EDGE and C-CENTRE alignment theories posit that a part of the vowel plateau is masked by the acoustics of consonants. Consequently, the acoustic onset of the vowel in our measurements cannot represent the true target achievement of the vowel. In order to estimate the vowel duration, we used the acoustic duration of the vowel in the VCV words, and then added half the average duration of the consonant in such words (since, the RIGHT-EDGE and C-CENTRE point coincide in such words, this is appropriate). For RIGHT-EDGE alignment, this estimated vowel duration (d_V=192ms) was added from the RIGHT-EDGE point along with some normally-distributed variation estimated as the standard deviation of the acoustic vowel durations (σ_V=32ms). For C-CENTRE alignment, the same estimated vowel duration (d_V=192ms) was added from the C-CENTRE point along with some normally-distributed variation estimated as the standard deviation of the acoustic vowel durations (σ_V=32ms).

Beyond this, we also estimated poly-constituent shortening on the relevant segments based on the proportional change from VCV to VCCV words: pre-vocalic consonant proportional change (0.88), and vowel duration proportional change (0.8).



C-CENTRE alignment

RIGHT-EDGE alignment

Figure 12: Schematic representations of time stamp generation for C-CENTRE alignment (left) and RIGHT-EDGE alignment (right). The x-axis in the figure represents time. The V marks the vowel, and C_1-C_2 represent word-initial consonants.

We first wanted to verify that we are able to reproduce the main insights in Shaw and Gafos (2015) and Shaw et al. (2009), even when we factor in poly-constituent shortening. In Figure 12, we vary the standard deviation of the vowel duration (σ_V) from 0 to 20ms, in steps of 1. For each standard deviation value, we simulated a 1000 word pairs, and then calculated the RSD values over them.

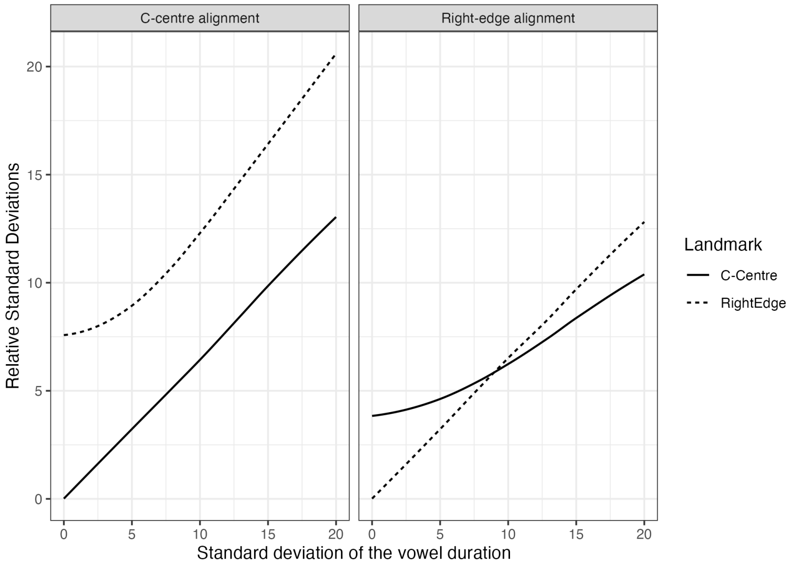


Figure 13: Simulated Relative Standard Deviations values for C-CENTRE and RIGHT-EDGE alignment with increasing standard deviation of vowel duration (replicating the simulation results in Shaw and Gafos (2015) and Shaw et al. (2009))

As can be seen in the figure, for RIGHT-EDGE alignment (right), after the vowel duration standard deviation goes past about 9, the RSD for C-CENTRE-TO-ANCHOR interval is lower. In contrast, for C-CENTRE alignment (left), the RSD for C-CENTRE-TO-ANCHOR interval is always lower.

Given that the estimated standard deviation of the vowel plateau duration (σ_V) was about 32ms in our data, the simulated results in Figure 13 already suggest that the RSD measurements for C-CENTRE-TO-ANCHOR interval are expected to be lower than the RSD measurements for RIGHT-EDGE-TO-ANCHOR interval, irrespective of the underlying alignment for word-medial /fl/ sequences in Spanish. To further establish this fact, we simulated word-pairs with across word-pair variation that we observed in our own data. We estimated the standard deviations of the average duration of each segment in each word pair [pre-consonantal consonant (sd=8ms), pre-vocalic consonant (sd=1.7ms), and the vowel (sd=7.4ms)], and used these values to simulate a 1000 word-pairs separately for RIGHT-EDGE and C-CENTRE alignments. These results are shown in Figure 14. As gleaned from the previous simulation, for our data, independent of the underlying organisation, the RSD for the C-CENTRE-TO-ANCHOR interval duration is expected to be generally lower than the RSD for the RIGHT-EDGE-TO-ANCHOR interval duration.

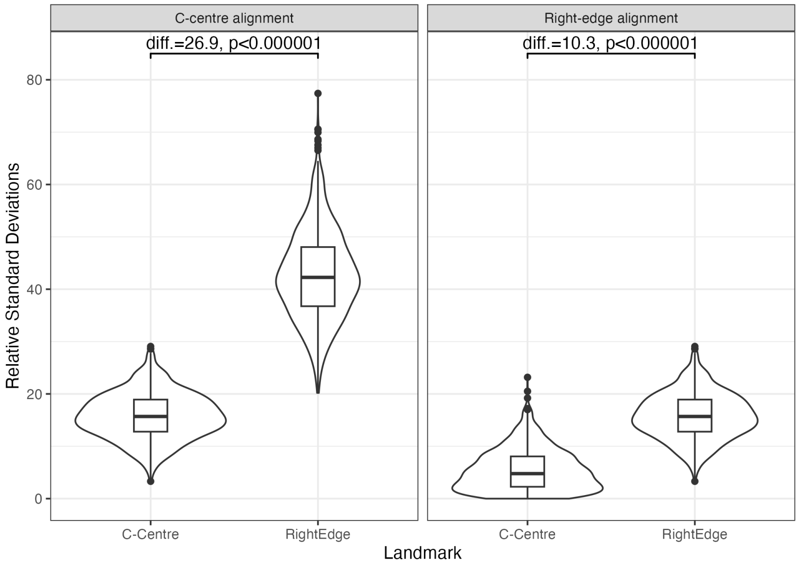


Figure 14: Simulated Relative Standard Deviations values based on observed parameter values in our data

To summarise, we found that for our data, the RSD for C-CENTRE-TO-ANCHOR interval duration was lower; however, our modelling results suggest that, when we include an estimate of poly-constituent shortening, both complex and simplex onsets are expected to show generally lower RSD values for C-CENTRE-TO-ANCHOR interval duration. The above results suggest that, *even when we factor in poly-constituent shortening*, RSD measurements are difficult to interpret when the value corresponding to C-CENTRE-TO-ANCHOR interval duration is lower. Note, in contrast, when the RSD measurements are generally lower for the RIGHT-EDGE-TO-ANCHOR interval duration, then one can in fact infer a RIGHT-EDGE alignment (as is true in Moroccan and Jazani Arabic in prior research).

Above we say "difficult", not impossible, since there is some information above to suggest that word-medial /fl/ sequences may have a simplex organisation. Note, in Figure 14, the simulated RSD values for both the C-CENTRE-TO-ANCHOR and RIGHT-EDGE-TO-ANCHOR interval durations under a RIGHT-EDGE alignment scenario are in the same range as the observed values in our data. In contrast, for the C-CENTRE alignment scenario, the simulated RSD values for RIGHT-EDGE-TO-ANCHOR interval duration are far above our observed values, suggesting that if the underlying organisation were indeed C-CENTRE alignment for word-medial /fl/ sequences, then we would have observed far more overlap of the pre-vocalic consonant and the vowel (and subsequent shortening of RIGHT-EDGE-TO-ANCHOR interval duration).

9 Conclusion

In the current paper, we presented two experiments and explored two different types of measures to examine C-CENTRE-TO-ANCHOR vs. RIGHT-EDGE-TO-ANCHOR interval stability for word-initial and word-medial consonant sequences in Spanish: (a) interval stability measured in terms of RSDs, and (b) pre-vocalic consonant duration.

The findings in both experiments showed that the C-CENTRE-TO-ANCHOR interval was more stable that RIGHT-EDGE-TO-ANCHOR for both word-initial and word-medial /fl/ and word-medial /sl/ and /sm/ sequences. Additionally, acoustic durations of the prevocalic consonants (that is, /l/ in /fl/ and /sl/ and /m/ in /sm/) exhibited shortening when part of a consonantal sequence. Before discussing the potential implications of these findings, it is worth noting that we have no a priori reasons to believe that the results pertain only to the dialect under study, namely, north-central Peninsular Spanish. There is no independent evidence that we are aware of that the observed patterns are part of a larger systematic sound change in this dialect area. This suggests to us that our findings are very likely generalisable to other dialects of Spanish.

At first blush, our findings raise intriguing implications regarding the traditional analysis of the syllabic affiliation of consonant sequences in Spanish. If we assume that there is a universal mapping of phonetic signatures to syllabic structure and that the proposed phonetic diagnostics allow us to infer the latter, our results suggest that word-medial consonant sequences (*e.g.*, /sm sl/) that are traditionally assumed to be heterosyllabic may in fact be tautosyllabic. This interpretation is also suggested by Aldrich and Simonet (2019), who highlighted the possibility of such syllabic reanalysis in their analysis of vowel compression. Their investigation found no clear differences in the durations of initial vowels in stimuli containing various consonant sequences (*p*V*pa*, *p*VC*pa*, and *p*VCC*pa*) — *i.e.*, the relevant coda consonants (C) and coda clusters (CC), as per standard analyses, did not result in shorter initial vowels, when compared to the initial vowel in an open syllable. This stands in contrast to observations in English by Katz (2012) and Munhall et al. (1992), where substantial vowel shortening occurs in the presence of a coda consonant. Moreover, it is crucial to acknowledge that much of the evidence regarding syllabic affiliation in Spanish stems from speaker intuitions, which have generally exhibited consistency. However, reliance solely on meta-linguistic tasks, such as speaker intuitions, may introduce confounding factors, particularly concerning word-edge and morpheme-edge judgments across languages. Furthermore, the reliance on speaker intuitions prompts consideration of the extent to which these judgments are informed by orthographic knowledge and other word-edge phonotactics as opposed to their knowledge of syllable structure *per se*. This distinction is particularly relevant for a language like Spanish, where orthographic representations, extensively emphasized in school for literacy purposes, may influence speaker intuitions about syllabic structure — to our knowledge, no study has systematically explored this potential confound.

Despite the logical possibility raised above, following the general advice of Shaw and Gafos (2015), we caution against interpreting our experimental results as evidence for a particular onset structure analysis, without precise quantitative modelling or appropriate controls. There are two main reasons for this. First, the prevocalic consonant (C_2) shortening we observed could actually be from poly-constituent shortening, which extends beyond the syllable. As a reminder, we not only found that the target pre-vocalic consonant shortened in the presence of a preceding consonant (C_1C_2, which would be standardly taken as evidence of a complex syllable), but we also observed similar shortening outside the syllable. Therefore, the shortening cannot be attributed to syllable structure without more precise specification of the effects. Second, we showed with simulations, using estimated parameter values from our own data, that a lower RSD for the C-CENTRE-TO-ANCHOR intervals is possible when the actual underlying temporal organisation is one of either C-CENTRE alignment or RIGHT-EDGE alignment. RIGHT-EDGE alignment languages, under the right circumstances, can be observed to have more C-CENTRE-TO-ANCHOR interval stability. This generally replicates the findings of Shaw and Gafos (2015) and Shaw et al. (2009), but extends them to show that the issue persists even when we factor in poly-constituent shortening in the simulations. Both the above reasons instead suggest that there is a need for quantitatively more precise theoretical understanding of the phenomenon of poly-constituent shortening in order to make progress on understanding the connection between syllable structure and temporal organisation. The challenge for future work is that poly-constituent shortening has been observed to be different across different consonant sequences in the same language (Katz 2012). Therefore, a more general statement of the effect is not possible given current knowledge; instead, appropriate controls need to be used with the experiment to estimate the effect of poly-constituent shortening as relevant to the specific stimuli used.

When we looked at prevocalic consonant (C_2) shortening effects, there was one case where the effect size was larger than the rest, namely the word-initial /fl/ case. As can be seen in Table 5, the shortening effect on the /l/ in /fl/ sequences is about 42 ms — in post-hoc tests, this was statistically clearly larger than any of the other shortening effects. In contrast, the shortening effects observed in the post-hoc comparisons probing the effect of poly-constituent shortening (Table 10), which are all less than 18ms. So, one could argue that while the word-medial cases may be due to poly-constituent shortening, the word-initial case constitutes some evidence of a complex onset related shortening effect. However, even here, caution is warranted. In the word-initial case, there is the possibility of another well-studied factor, namely, domain-initial lengthening. One could argue that the word-initial /l/ in words such as ⟨lato⟩ is subject to a domain-initial lengthening effect (Fougeron and Keating 1997). Consequently, the additional shortening of /l/ observed in the case of word-initial /fl/ could simply be the result of the /l/ being non-initial. As a result, the larger effect size of C_2 shortening seen in the word-initial /fl/ context could simply be result of not adjusting for domain-initial lengthening and poly-constituent shortening. In short, all the observed patterns in our studies can be explained as an interaction between poly-constituent shortening and domain-initial lengthening, while maintaining that all the consonant sequences studied have RIGHT-EDGE alignment. Despite the observation of a C-CENTRE-TO-ANCHOR stability, there is no need (and in fact some evidence against) the possibility of the consonant sequences having an underlying C-CENTRE alignment.

An important observation stems from our results and the above discussion — poly-constituent shortening and domain-initial lengthening themselves appear to be language-specific (or minimally, in need of further study). While it is possible to account for some of the C-CENTRE-TO-ANCHOR stability patterns as a result of not adjusting for poly-constituent shortening or domain-initial lengthening, crucially, it can't be the case that such effects are there in all languages (at least, not within the same phonological domain). Crucially, as noted earlier, there are consonant sequences, in Moroccan Arabic, Jazani Arabic and Italian, where there is RIGHT-EDGE-TO-ANCHOR stability — *i.e.*, the addition of a preceding consonant to a word-initial C_2V sequence to create $\#C_1C_2V$ sequences does not result in any shortening of the pre-vocalic consonant (C_2) or for that matter the vowel, which is the reason there is RIGHT-EDGE-TO-ANCHOR stability. If poly-constituent shortening and domain-initial lengthening played a role in these two languages as they do in Spanish, then the RIGHT-EDGE-TO-ANCHOR stability effect would not have been possible. In fact, Durvasula (2023) shows that there is no clear change in the pre-vocalic consonant duration due to the addition of consonants word-initially in Jazani Arabic. We point this issue out here because if indeed poly-constituent shortening and domain-initial lengthening have to be understood better in their own right, the quest to probe for consistent temporal stability patterns related to syllable structure is further complicated.

There are two important implications that stem from the discussion in the previous paragraphs. First, we have a new possibility for why such stability patterns have been inconsistently observed in specific languages and segmental contexts — it is possible that other phonetic factors (including poly-constituent shortening and domain-initial lengthening) have not been appropriately adjusted for. Second, and more importantly, it is possible to view all observations of C-CENTRE alignment in the literature as an interaction of the language-specific effects of poly-constituent shortening and domain-initial lengthening; if so, it is possible that C-CENTRE alignment is always simply a mirage that stems from such an interaction, and that RIGHT-EDGE alignment is the only alignment that is underlying present. This latter implication is a possibility that we think is particularly exciting given that it suggests uniform temporal stability pattern across languages, once other factors are controlled for — we leave to explore in future work.

10 Acknowledgements

Author contributions: The corresponding author Karthik Durvasula and co-author Silvina Bongiovanni are responsible for the conceptualization and design of the study, and carried out participant recruitment, data collection and analysis and writing of the manuscript. Both authors have contributed equally to this manuscript.

Competing interests: The authors have no conflicts of interest to declare.

Ethics statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Review Board at Michigan State University (study 00005619). Informed consent was obtained from all subjects involved in the study.

References

Aldrich, Alexander C and Miquel Simonet (2019). "Duration of syllable nuclei in Spanish." *Studies in Hispanic and Lusophone Linguistics* 12.2, pp. 247–280.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker (2015). "Fitting linear mixed-effects models using `lme4`." *Journal of Statistical Software* 67.1, pp. 1–48.

Boersma, Paul and David Weenink (2023). *Praat: doing Phonetics by Computer [Computer program]*. Version 6.3.09, retrieved 2 March 2023 from from http://www.praat.org/.

Bolozky, Shmuel (1997). "Israeli Hebrew phonology." In: *Phonologies of Asia and Africa (including the Caucasus). Vol. 1.* Ed. by A. S. Kaye. Winona Lake, Ind.: Eisenbrauns, pp. 287–311.

Bradley, Travis G (2006). "Spanish complex onsets and the phonetics-phonology interface." In: *Optimality-theoretic studies in Spanish phonology*. Ed. by Fernando Martinez Gil and Sonia Colina. Amsterdam: John Benjamins, pp. 15–38.

Browman, Catherine P. and Louis H. Goldstein (1988). "Some Notes on Syllable Structure in Articulatory Phonology." *Phonetica* 45, pp. 140–155. DOI: `10.1159/000261823`.

Brunner, Jana, Christian Geng, Stavroula Sotiropoulou, and Adamantios I. Gafos (2014). "Timing of German onset and word boundary clusters." *Laboratory Phonology* 5, pp. 403–454.

Cho, Taehong, Patricia Keating, Cécile Fougeron, and Chai-shune Hsu (2003). "Domain-initial strengthening in four languages." In: *Laboratory Phonology 6*. Cambridge University Press, pp. 145–163.

Colantoni, Laura and Jeffrey Steele (2005). “Phonetically-driven epenthesis asymmetries in French and Spanish obstruent-liquid clusters.” In: *Experimental and Theoretical Approaches to Romance Linguistics.* Ed. by R. Gess and E. Rubin. Amsterdam; Philadelphia: John Benjamins Pub. Co, pp. 77–96.

Colina, Sonia (2009). *Spanish phonology: A syllabic perspective.* Georgetown University Press.

Colina, Sonia (2012). “Syllable structure.” In: *The handbook of Hispanic linguistics.* Ed. by Erin O’Rourke José Ignacio Hualde Antxon Olarrea. Wiley Online Library, pp. 133–151.

Colina, Sonia (2016). “On onset clusters in Spanish: Voiced obstruent underspecification and/f.” In: *The syllable and stress: Studies in honor of James W. Harris.* Ed. by Rafael A. Nuñez Cedeño. Boston: Mouton de Gruyter, pp. 107–137.

Cuenca, María Heliodora (1997). “Análisis instrumental de la duración de las vocales en español.” *Philologia hispalensis, 11 (1), 295-307.*

Davis, Stuart Michael (1987). “Italian Onset Structure and the Distribution of “il” and “lo”.” *Proceedings of the Annual Eastern States Conference on Linguistics*, pp. 64–74.

Dell, François (1995). “Consonant clusters and phonological syllables in French.” *Lingua* 95.1. French Phonology, pp. 5–26. ISSN: 0024-3841. URL: `http://www.sciencedirect.com/science/article/pii/0024384195900993`.

Durvasula, Karthik (2023). “A simple acoustic measure of onset complexity.” In: *Proceedings of the 20th International Congress of Phonetic, Sciences (ICPhS 2023).* Ed. by Radek Skarnitzl and Jan Volín. Prague, The Czech Republic: Guarant International, pp. 2010–2013. URL: `https://drive.google.com/file/d/15U2l2y4_-9lyZAgmiccQYXYj9zBi_CAu/view`.

Durvasula, Karthik, Mohammed Qasem Ruthan, Sarah Heidenreich, and Yen-Hwei Lin (2021). “Probing syllable structure through acoustic measurements: case studies on American English and Jazani Arabic.” *Phonology* 38.2, pp. 173–202. DOI: `10.1017/S0952675721000142`.

Dushoff, Jonathan, Morgan P. Kain, and Benjamin M. Bolker (2019). “I can see clearly now: Reinterpreting statistical significance.” *Methods in Ecology and Evolution* 10.6, pp. 756–759. DOI: `10.1111/2041-210X.13159`. eprint: `https://besjournals.onlinelibrary.wiley.com/doi/pdf/10.1111/2041-210X.13159`. URL: `https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.13159`.

Farnetani, Edda and Shiro Kori (1986). “Effects of syllable and word structure on segmental durations in spoken Italian.” *Speech communication* 5.1, pp. 17–34.

Fougeron, Cécile and Patricia Keating (1997). “Articulatory strengthening at edges of prosodic domains.” *Journal of the Acoustic Society of America* 101.6, pp. 3728–3740. DOI: `https://doi.org/10.1121/1.418332`.

Fowler, Carol A (1981). “A relationship between coarticulation and compensatory shortening.” *Phonetica* 38.1-3, pp. 35–50.

Franke, Mona, Philip Hoole, and Simone Falk (2023). “Temporal organization of syllables in paced and unpaced speech in children and adolescents who stutter.” *Journal of Fluency Disorders* 76, p. 105975.

Gafos, Adamantios I., Simon Charlow, Jason A. Shaw, and Philip Hoole (2014). “Stochastic time analysis of syllable-referential intervals and simplex onsets.” *Journal of Phonetics* 44. Dynamics of Articulation and Prosodic Structure, pp. 152–166. ISSN: 0095-4470. DOI: `https://doi.org/10.1016/j.wocn.2013.11.007`. URL: `https://www.sciencedirect.com/science/article/pii/S0095447013000934`.

Gelman, Andrew and John Carlin (2014). “Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors.” *Perspectives on Psychological Science* 9.6. PMID: 26186114, pp. 641–651. DOI: `10.1177/1745691614551642`. eprint: `https://doi.org/10.1177/1745691614551642`. URL: `https://doi.org/10.1177/1745691614551642`.

Gili y Gaya, Samuel (1921). “La r simple en la pronunciación espanola [The tap in Spanish pronunciation].” *Revista de Filologıa Espanola* 8, pp. 271–280.

Goldstein, Louis H., Ioana Chitoran, and Elisabeth Selkirk (2007). “Syllable Structure as Coupled Oscilator Models: Evidence from Georgian vs. Tashlhiyt Berber.” *Proceedings of the 16th International Congress of Phonetic Sciences*, pp. 241–244.

Harris, James Wesley (1983). *Syllable structure and stress in Spanish. A nonlinear analysis.* Cambridge, MA: MIT Press.

Hermes, Anne, Doris Mücke, and Bastian Auris (2017). “The variability of syllable patterns in Tashlhiyt Berber and Polish.” *Journal of Phonetics* 64, pp. 127–144.

Hermes, Anne, Doris Mücke, and Martine Grice (2013). “Gestural coordination of Italian word-initial clusters: The case of ’impure s’.” *Phonology* 30.1, pp. 1–25. ISSN: 09526757. DOI: `10.1017/S095267571300002X`.

Hlavac, Marek (2018). *stargazer: Well-Formatted Regression and Summary Statistics Tables.* R package version 5.2.2. Central European Labour Studies Institute (CELSI). Bratislava, Slovakia. URL: `https://CRAN.R-project.org/package=stargazer`.

Hualde, José Ignacio (1991). “On Spanish syllabification.” *Current studies in Spanish linguistics* 475.493, pp. 182–198.

Hualde, José Ignacio (2005). *The sounds of Spanish.* Cambridge University Press.

Katz, Jonah (2012). “Compression effects in English.” *Journal of Phonetics* 40.3, pp. 390–402. ISSN: 0095-4470. DOI: `https://doi.org/10.1016/j.wocn.2012.02.004`. URL: `https://www.sciencedirect.com/science/article/pii/S0095447012000149`.

Kuznetsova, Alexandra, Per B. Brockhoff, and Rune H. B. Christensen (2017). "lmerTest Package: Tests in Linear Mixed Effects Models." *Journal of Statistical Software* 82.13, pp. 1–26. DOI: `10.18637/jss.v082.i13`.

Marchini, Gilly and Michael Ramsammy (2022a). "Dialect-specific Acoustic Correlates of Stress in Spanish: The Role of Vowel Compression and Syllable Structure." *University of Pennsylvania Working Papers in Linguistics* 28.1, p. 12.

Marchini, Gilly and Michael Ramsammy (2022b). "Vowel Compression in Altiplateau Mexican Spanish." *Isogloss. Open Journal of Romance Linguistics* 8.4, pp. 1–27.

Marin, Stefania and 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, pp. 380–407.

Marin, Stefania and Marianne Pouplier (2014). "Articulatory synergies in the temporal organization of liquid clusters in Romanian." *Journal of Phonetics* 42, pp. 24–36.

Martínez-Gil, Fernando (2001). "Sonority as a primitive phonological feature." In: *Features and interfaces in Romance: Essays in Honor of Heles Contreras*. Ed. by K. Zagona J. Herschensohn E. Mallen. Amsterdam & Philadelphia: John Benjamins, pp. 203–222.

McAuliffe, Michael, Michaela Socolof, Sarah Mihuc, Michael Wagner, and Morgan Sonderegger (2017). "Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi." In: *Proc. Interspeech 2017*, pp. 498–502. DOI: `10.21437/Interspeech.2017-1386`.

Morales-Front, Alfonso (2018). "The Spanish Syllable." In: *The Cambridge handbook of Spanish linguistics.* Ed. by Kimberly L Geeslin. Cambridge University Press, pp. 190–210.

Mücke, Doris, Anne Hermes, and Sam Tilsen (2020). "Incongruencies between phonological theory and phonetic measurement." *Phonology* 37.1, pp. 133–170. DOI: `10.1017/S0952675720000068`.

Munhall, Kevin, Carol Fowler, Sarah Hawkins, and Elliot Saltzman (1992). ""Compensatory shortening" in monosyllables of spoken English." *Journal of Phonetics* 20.2, pp. 225–239. ISSN: 0095-4470. DOI: `https://doi.org/10.1016/S0095-4470(19)30624-2`. URL: `https://www.sciencedirect.com/science/article/pii/S0095447019306242`.

Pouplier, Marianne (2012). "The gestural approach to syllable structure: universal, language- and cluster-specific aspects." Ed. by Susanne Fuchs, Melanie Weirich, Daniel Pape, and Pascal Perrier, pp. 63–96.

Prieto, Monica (2002). "Acoustic correlates of the syllable: Evidence from Spanish." PhD dissertation. University of Illinois at Urbana-Champaign.

R Core Team (2021). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing. Vienna, Austria. URL: `http://www.R-project.org`.

Real Academia Española (2011). *Nueva gramática de la lengua española. Vol 3: Fonética y fonología.* Madrid: Espasa Calpe.

RStudio Team (2020). *RStudio: Integrated Development Environment for R.* RStudio, PBC. Boston, MA. URL: `http://www.rstudio.com/`.

Saporta, Sol and Heles Contreras (1962). *A Phonological Grammar of Spanish.* Seattle: University of Washington Press.

Selkirk, Elliot and Karthik Durvasula (2013). *Acoustic correlates of consonant gesture timing in English.* Paper presented at the 166th Meeting of the Acoustical Society of America, San Francisco, USA.

Shaw, Jason A. and Adamantios I. Gafos (2015). "Stochastic Time Models of Syllable Structure." *PLoS One* 10.5, e0124714.

Shaw, Jason A., Adamantios I. Gafos, Philip Hoole, and Chakir Zeroual (2009). "Syllabification in Moroccan Arabic: evidence from patterns of temporal stability in articulation." *Phonology* 26.1, pp. 187–215. DOI: `10.1017/S0952675709001754`.

Shaw, Jason A., Adamantios I. Gafos, Philip Hoole, and Chakir Zeroual (2011). "Dynamic invariance in the phonetic expression of syllable structure: a case study of Moroccan Arabic consonant clusters." *Phonology* 28.3, pp. 455–490.

Sotiropoulou, Stavroula, Mark Gibson, and Adamantios I. Gafos (2020). "Global organization in Spanish onsets." *Journal of Phonetics* 82, p. 100995. ISSN: 0095-4470. URL: `http://www.sciencedirect.com/science/article/pii/S0095447020300863`.

Tilsen, Sam, Draga Zec, Christina Bjorndahl, Becky Butler, Marie-Josee L'Esperance, Alison Fisher, Linda Heimisdottir, Margaret Renwick, and Chelsea Sanker (2012). "A cross-linguistic investigation of articulatory coordination in word-initial consonant clusters." *Cornell Working Papers in Phonetics and Phonology 2012*, pp. 51–81.

Wickham, Hadley (2017). *tidyverse: Easily Install and Load the 'Tidyverse'.* R package version 1.2.1. URL: `https://CRAN.R-project.org/package=tidyverse`.

Wiese, Richard (1996). *The Phonology of German.* Oxford: Clarendon.

11 Appendix

11.1 Fillers, experiment 1

Real word pair	Glosses	Nonce word pair
<leches, leyes>	milk.MASC.PL, leyes.MASC.PL	<chefa, yesa>
<macho, mayo>	male.MASC.SG, May.MASC.SG	<chasa, yasa>
<pena, peina>	pity.FEM.SG, comb.3SG	<fena, feina>
<reno, reino>	reindeer.MASC.SG, kingdom.MASC.SG	<teso, teiso>
<paso>	step.MASC.SG	<dape>
<pozo>	well.MASC.SG	<dapo>
<silo>	silo.MASC.SG	<pile>
<celo>	zeal.MASC.SG	<pale>

Table 12: Fillers, experiment 1

11.2 Fillers, experiment 2

Real word pair	Glosses
<casado, cazado>	married.MASC.SG, hunted.MASC.SG
<haz, has>	do.IMP.2SG, have.2SG
<roza, rosa>	graze.3SG, pink.SG
<taza, tasa>	mug.FEM.SG, rate.FEM.SG
<sede, cede>	seat.MASC.SG, relinguish.3SG
<sito, cito>	located.MASC.SG, make an appointment.1SG
<sien, cien>	temple.MASC.SG, one hundred.SG
<voz, vos>	voice.FEM.SG, you.PRO.2SG
<cauce, cause>	riverbed.MASC.SG, cause.SUBJ.1SG
<reciente, resiente>	recently.ADJ.SG, resent.3SG
<cierra, sierra>	close.3SG, mountain.FEM.SG
<hacia, Asia>	towards.PREP, Asia
<vez, ves>	time.FEM.SG, see.2SG
<cepa, sepa>	strain.FEM.SG, know.SUBJ.3SG
<ciento, siento>	one hundred.SG, feel.1SG

Table 13: Fillers, experiment 2