Research Article

Phonetica 2019;76:448–478
DOI: 10.1159/000501508

Received: July 12, 2018
Accepted: June 17, 2019
Published online: August 27, 2019

Temporal Aspects of Word Initial Single Consonants and Consonants in Clusters in Spanish

1956

Mark Gibson[a] Stavroula Sotiropoulou[b] Stephen Tobin[b]
Adamantios I. Gafos[b]

[a]Universidad de Navarra, Pamplona, Spain; [b]Universität Potsdam, Potsdam, Germany

Abstract

We examined gestural coordination in C1C2 (C1 stop, C2 lateral or tap) word initial clusters using articulatory (electromagnetic articulometry) and acoustic data from six speakers of Standard Peninsular Spanish. We report on patterns of voice onset time (VOT), gestural plateau duration of C1, C2, and their overlap. For VOT, as expected, place of articulation is a major factor, with velars exhibiting longer VOTs than labials. Regarding C1 plateau duration, voice and place effects were found such that voiced consonants are significantly shorter than voiceless consonants, and velars show longer duration than labials. For C2 plateau duration, lateral duration was found to vary as a function of onset complexity (C vs. CC). As for overlap, unlike in French, where articulatory data for clusters have also been examined, clusters where both C1 and C2 are voiced show more overlap than where voicing differs. Further, overlap was affected by the C2 such that clusters where C2 is a tap show less overlap than clusters where C2 is a lateral. We discuss these results in the context of work aiming to uncover phonetic (e.g., articulatory or perceptual) and phonological forces (e.g., syllabic organization) on timing.

1 Introduction

It is a standard working hypothesis in the study of speech that phonological units (such as segments or syllables or higher structures such as feet or prosodic words) are embodied in actions of the speech gestures that constitute these units. Yet, variation in interarticulatory timing, in particular, has proven nettlesome in teasing apart the phonologically encoded constraints that act on the dynamic actions of the different articulators and the low level biomechanical constraints that may affect interarticulatory timing (Gafos & Goldstein, 2012). In the present study, we examine the timing patterns of Spanish word initial consonants and consonant clusters in an effort to add

Mark Gibson
C/Parque de los Olmos, 16 1º izq.
ES–31180 Zizur Mayor, Navarra (Spain)
E-Mail mgibson@unav.es

to the growing body of ongoing research which seeks systematicities underlying the language-specific ways oral and laryngeal gestures combine to form larger units.

Effects of voicing on the intrasegmental properties of oral and laryngeal gestural coordination as well as on durational properties of closures as a function of place of articulation constitute two of the longest and most fruitful areas of research in phonetics and phonology (see Abramson & Whalen, 2017, for a comprehensive review). More recently, with the greater availability of techniques for tracking articulation, researchers have increasingly also considered issues of intersegmental, as opposed to intrasegmental, timing and the extent to which these levels of timing interact with one another. For example, Bombien & Hoole (2013) examined the relationship between voicing and the relative timing of oral gestures using articulatory and acoustic data from German and French. Their results indicate that intersegmental oral timing in consonant clusters and specifically the timing of the oral constrictions of the two consonants in a CC cluster depend crucially on the voicing specifications of the two consonants. For instance, in German, a language where the voicing contrast involves, for voiceless stops, the presence of a laryngeal abduction gesture bound to the oral constriction gesture of the stop (intrasegmental timing), the oral gestures of the two consonants in /kl/ are farther apart in time than in /gl/. That is, intrasegmental timing (the timing between the laryngeal abduction gesture and the oral constriction for the stop) interacts with intersegmental timing (the lag between the two oral constrictions of the /k/ and the /l/). Following up on this line of work, our present study aims to synthesize these two strands of research on intrasegmental and intersegmental timing with acoustic and articulatory data from Spanish. Our objective with the present article, thus, is to advance current research dealing with how language-specific voicing implementation affects intersegmental timing patterns. This line of enquiry is significant in that it expands on what has been learned in recent years with regard to the temporal organization of articulatory gestures, taking us one step closer to understanding the language-specific intricacies of such organization.

This paper is organized as follows. Section 2 (Background) presents previous findings for VOT and consonant duration in Spanish word initial consonants or consonant clusters. Additionally, we review articulatory results from other languages which may inform hypotheses on overlap in CC clusters. From this review, we extract a set of hypotheses which we pursue with our Spanish data. Section 3 (Experimental methods) outlines our study design, experimental methodology and statistical analyses. Section 4 (Results) reports our findings for VOT, stop duration, C2 duration (where C2 is the nonvelarized or clear /l/ and the tap /ɾ/), overlap of the two consonants in CC, and distributional and durational properties of the intrusive vocoid which sometimes appears between the two consonants in CC clusters. In section 5 (Summary and Discussion), we place our results in the context of the literature which provided the basis for our hypotheses. Finally, in section 6 (Conclusion), we present a summary of the main results and point to some directions of future research.

2 Background and Hypotheses

In this section, and because our study is the first to address several phonetic factors that may play a role in overlap patterns of CC clusters in Spanish using both acoustic and articulatory data, we review previous results from Spanish and other languages that stand in the background of our investigation. On the basis of this past literature, we extract four broad hypotheses which will drive both the analyses and the discussion of our results.

Languages such as French and Spanish contrast voiced and voiceless stops by way of a true voicing/short lag voice onset time (VOT) paradigm (Maddieson, 2009). Thus, stops that are phonologically specified as voiced show (pre)voicing during the closure and stops that are (phonologically) voiceless surface with a short lag VOT (Caramazza & Yeni-Komshian, 1974; Lisker & Abramson, 1964; Rosner et al., 2000; Williams, 1977). This configuration is in contrast to many Germanic languages which express the phonological opposition by way of a short lag/long lag VOT pattern. In these languages, phonologically voiced stops (even if phonetically devoiced) are produced with a short lag VOT, and phonologically voiceless stops are produced with a long lag VOT (Bombien & Hoole, 2013; Hoole, Bombien, Kühnert, & Mooshammer, 2009; Klatt, 1975).

Constriction location (i.e., place of articulation) has long been reported to show effects cross-linguistically on patterns of VOT, velar consonants having longer VOTs than labials and coronals (Crystal & House, 1988; Docherty, 1992; Hoole, 2006; Lisker & Abramson, 1964; Maddieson, 1997; Nearey & Rochet, 1994; Weismer, 1980). The biomechanical and aerodynamic bases of this effect have been extensively discussed (see among others, Bombien & Hoole, 2013; Caramazza & Yeni-Komshian, 1974; Cho & Ladefoged, 1999; Maddieson, 2009; Ohala & Ohala, 1973; Stevens, 2000) and corroborated by studies examining the prespeech babble of infants in a number of languages (Eilers, Oller & Benito-Garcia, 1984). For Spanish stops specifically, Williams (1977), Rosner et al. (2000) and later Gibson et al. (2015) found significantly shorter lag times for labials and coronals (single consonant onsets followed by a vowel) than for velars. This result obtains regardless of whether the consonant appears in simplex or complex onsets (Gibson et al., 2015).

Less is understood about the effects of syllable complexity on the patterns of laryngeal timing for stops across languages. Bombien and Hoole (2013) found significant effects for VOT in different onset types in both French and German. However, the direction of the effects was conflicting. In German, VOT in complex onsets is shorter than in simplex onsets, while in French, VOT in complex onsets is longer than in simplex onsets. Once place of articulation was taken into account, though, the impact of syllable complexity on VOT was quite narrow. For Spanish, it has been found that onset complexity does not influence patterns of VOT (Gibson et al., 2015). In the current study we aim to confirm the results obtained in Gibson et al. (2015) with regard to the effects of constriction location on patterns of VOT in Spanish and extend this line of research to address the interaction between the voicing implementation for stops and oral-oral coordination patterns in Spanish clusters.

There is a long history of phonetic observations on Spanish and other languages addressing durational aspects of single consonants or consonants in clusters. Studies in a number of languages have shown that syllable complexity and constriction location (i.e., place of articulation), independently, affect the duration of consonant closures. For single consonants as onsets (simplex onsets), voiceless stops tend to be longer than voiced stops (Lisker & Abramson, 1964), a pattern that has been consistently confirmed for Spanish (Blecua, 2000; Gibson et al., 2015; Martinez Celdran, 1993, 1996; Villamizar, 2002). As for effects of place of articulation, previous studies have shown that velar closures in Spanish onsets tend to be longer than labials (Gibson et al., 2015; Villamizar, 2002).

In complex onsets (C1C2), the duration of initial stop closures (C1) has been found to decrease in comparison to their duration when the same stops appear unclustered (see among others, Bombien & Hoole, 2013; Haggard, 1973). Vowel-adjacent consonants (C2) in clusters show more reduction than word initial consonants (Haggard, 1973), while liquids shorten more than stops (Marin & Bučar Shigemori, 2014; O'Shaughnessy, 1974). In Spanish C1C2 clusters, Blecua and Machucas's (2000) acoustic study of stop + tap sequences revealed that the duration of the stop and tap are negatively correlated such that longer durations for the tap co-occur with shorter durations for the stop. Similarly, the acoustic study of Blecua et al. (2003) (also on Spanish) found a comparable effect for C2 laterals (which are nonvelarized or clear /l/s, in Spanish), such that lateral duration varied as a function of the duration of the preceding obstruent. The acoustic and electropalatographic study by Gibson et al. (2015) found that the closure duration of the C1 was susceptible to place of articulation and voicing as well as certain prosodic effects. A marked reduction in the duration of the lateral C2 was found for all tokens in complex onsets, as compared to the duration of laterals in the /lV/ context. Interestingly, it was shown that the degree of reduction is conditioned by the phonological specifications of the preceding consonant. For complex onsets, Gibson et al. (2015) found that laterals following velars were systematically shorter than those following labials (presumably due to the fact that labial articulation does not involve any lingual gesture), while laterals preceded by voiceless consonants were significantly shorter than those preceded by voiced stops.

Our review has so far focused on effects of voicing implementation and place of articulation on durational properties of consonants. We now turn to issues of relative timing or overlap between consonants in clusters. Here, there is a general absence of relevant prior works on Spanish onset clusters for the obvious reason that kinematic data have not been available. However, we can still extract with reasonable caution relevant predictions regarding the temporal arrangement of consonants in Spanish clusters from previous acoustic studies.

A rich tradition of phonetic observations for Spanish has dealt with the distribution and properties of a so-called intrusive vocoid that often appears between the two consonants in Spanish onset clusters (Bradley, 2006; Gili Gaya, 1921; Lenz, 1892; Malmberg, 1965; Menéndez Pidal, 1926).[1] Malmberg (1965) mentions that though vocoid intrusion in Spanish is more prevalent in stop + tap clusters, it is also attested in stop + lateral clusters. This claim is supported as well by the acoustic study of Blecua et al. (2003)of stop + lateral clusters in Spanish where it is reported that 16% of all stop + lateral clusters contained a vocalic element. Baltazani and Nicolaidis (2013), building on Baltazani and Nicolaidis' (2011) electropalatographic study of Greek clusters with rhotics, claim that the intrusive vocoid forms one of the two fundamental elements, along with the tongue tip constriction, of the rhotic (see Blecua, 2001, for a similar assertion for Spanish). In their analysis these authors envisage that "the rhotic is superimposed on a rhotic-specific vowel-to-vowel gesture" (p. 10). Baltazani (2009) states that "the Greek rhotic phoneme has at least two allophones: the tap in intervocalic position and the tap plus vowel-like transition in consonant clusters" (p. 94). Savu (2012) follows in this tradition, maintaining a "vocalic element – con-

[1] For intrusive vocoids in other languages, see Recasens & Espinosa (2007) on Catalan; Vago & Gósy (2007) on Hungarian; Baltazani & Nicolaidis (2011, 2013) on Modern Greek; Stolarski (2011) on Polish; Avram (1993) on Romanian; Gudurić & Petrović (2005) on Serbo-Croation; Pavlík (2008) on Slovak.

DOI: 10.1159/000501508

stricted interval – vocalic element" (p. 51) type structure of the tap in the presence of an adjacent consonant.

As Bradley (2006) notes, in order for an intrusive vocoid to occur in C1C2 clusters, the target plateaus of the two gestures that compose the sequence must be timed far enough apart such that phonation can be implemented prior to the target plateau of C2. This would imply relatively low overlap or what we will refer to as a nonzero interplateau interval (IPI, in short) between the main oral gestures of the two consonants in the cluster. If the appearance of the intrusive vocoid is indeed conditioned by low gestural overlap, then it is reasonable to expect that overlap patterns vary depending on whether C2 is a lateral or tap, given reports on differences in the distribution of the intrusive vocoid in stop + tap versus stop + lateral clusters (Blecua, 2001; Blecua & Machuca, 2000; Bradley, 2006; Colantoni & Steele, 2005, 2011; Gili Gaya, 1921; Lenz, 1892; Malmberg, 1965; Menéndez Pidal, 1926; Quilis, 1970, 1988; Quilis, Esgueva, Gutiérrez Araus & Cantarero, 1979, among others). Thus, since the intrusive vocoid appears more frequently in stop + tap clusters (Malmberg, 1965, among others), we expect to find lower overlap in stop + tap clusters than in stop + lateral clusters.[2]

Another area from which we can extract relevant hypotheses on overlap in CC clusters derives from previous studies in languages similar to Spanish. For instance, in their study of French and German stop + lateral clusters, Bombien and Hoole (2013) show that the underlying voice specification for C1 conditions articulatory overlap in German, but not in French. Whereas in German, voiced stop + lateral clusters are substantially more overlapped than voiceless stop + lateral clusters, in French, both onset types show low overlap (low in comparison with German voiced stop + lateral clusters). Specifically, the degree of overlap in both types of French clusters, voiced stop + lateral and voiceless stop + lateral, mirrors that of the German voiceless stop + lateral clusters. Bombien and Hoole (2013) attribute the low overlap in French clusters to the aerodynamic constraints that apply to the realization of the voicing contrast (true voice in voiced stops and short lag VOT in voiceless stops) in French. Assuming the temporal behavior of gestures in Spanish clusters can be likened to that of French, which is reasonable given the similarity in the implementation of the voicing contrast between these two languages, then one would expect low overlap in Spanish clusters (low in comparison to that in German voiced stop + lateral clusters) with no effect of C1 voicing on overlap patterns as has been reported for French.

Based on the reviewed literature, the hypotheses we assess in this study concern the patterning of VOT in singleton consonants versus consonants in clusters (complex onsets), plateau duration of stops in singleton and complex onsets (in C1 position of the onset when the onset is a C1C2 stop-sonorant cluster), plateau duration of laterals in word initial singleton onsets (henceforth /lV/) and complex onsets (in C2 position of the onset when the onset is a C1C2 stop + lateral cluster), and articulatory overlap and intrusive vocoids in Spanish stop + lateral and stop + tap clusters. In more explicit terms, our hypotheses are as follows.

For VOT (hypothesis 1 or H1 in short), on the basis of previous studies, we expect the usual effect of place of articulation on VOT, with values for velars greater

[2] For studies on the articulatory basis of such transitional vocoids, see Gafos (2002) and Gafos et al. (2010) on Moroccan Arabic as well as Hall (2006).

DOI: 10.1159/000501508

than for labials, and no effect of onset complexity (i.e., VOT should not change as consonants are added to the onset).

Regarding the durational properties of stop C1s and C2s (hypothesis 2, H2) in C1C2 clusters, effects of onset type, place of articulation, and voicing are all anticipated for stop C1s whereby gestural plateaus are expected to decrease as onset complexity increases for all place/voice contrasts. As for place and voicing effects, velars are expected to be longer than labials, while voiceless stops are expected to show longer durations than voiced stops. For C2 duration, we anticipate taps to exhibit shorter plateaus than laterals. A decrease for lateral duration is expected as a function of onset complexity such that laterals in /lV/ contexts exhibit longer durations than C2 laterals in clusters.

Concerning gestural overlap (hypothesis 3, H3), Spanish is expected to pattern like French (Bombien & Hoole, 2013), which exhibits relatively low overlap between C1 and C2. We expect some differentiation in C1C2 overlap on the basis of previous literature on intrusive vocoids which show a higher distribution of vocoids in stop + tap clusters as compared to stop + lateral (Malmberg, 1965, among others). Based on Bombien & Hoole (2013) on French, it is reasonable to expect no effect of voicing on overlap since Spanish and French have the same voicing instantiation patterns for voiced and voiceless stops. Finally, with regard to the intrusive vocoids (hypothesis 4, H4), vocoids are expected to occur more frequently in stop + tap clusters as compared to stop + lateral clusters.

3 Experimental Methods

3.1 Speakers and Corpus

Kinematic data were collected for six native speakers of Standard Peninsular Spanish. All subjects were between the ages of 18 and 35 years. Subjects were interviewed before data collection in order to evaluate fitness to participate in the study. The following criteria were established for the subject pool: (1) they had to be monolingual speakers of Standard Peninsular Spanish, (2) they could have no higher than a B2 level (upper intermediate according to the European Framework for Foreign Languages) in a foreign language, and (3) they must not show signs of nonstandard productions such as consonant lenition in syllable final position common to many regions of Spain. All subjects gave their informed consent to take part in the study and were paid for their participation.

Speakers were instructed to read a series of phrases that appeared on a computer screen. The corpus, which appears in Appendix A, consisted of native words that contained simplex and complex onsets involving both voiced/voiceless and labial/velar consonants in initial position. For simplex onsets consisting of voiced and voiceless stops (/b ɡ p k/), there were 14 total tokens: /p/ and /k/ each had 3 (1 with /a/, 1 with /e/, 1 with /o/), while 4 were included for /b/ and /ɡ/ (2 with /a/, 1 with /e/, 1 with /o/). The breakdown of tokens by cluster was the following: /bl/ and /ɡl/ had 4 tokens (2 with /a/, 1 with /e/, 1 with /o/) each (8 total), /kl/ and /pl/ had 3 tokens (1 with /a/, 1 with /e/, 1 with /o/) each (6 total), /kɾ/ and /pɾ/ had 3 tokens (1 with /a/, 1 with /e/, 1 with /o/) each (6 total)[3]. The usual constraints on finding tokens that are real words and fit the cluster-vowel combinations targeted in our study were the major determinants of these selections. Overall, then, this renders a total of 34 tokens (simplex stops had 14 total tokens, clusters had 20 tokens). Eight /lV/ stimuli were also included in order to compare lateral duration in complex onsets and /lV/ contexts (8 total /lV/

[3] For our analysis of the effects of C1 voicing on patterns of overlap, we used a subset of our data which only included stop+lateral clusters. We disregarded the stop+tap clusters for this particular analysis due to the fact that our corpus only contained stop+tap clusters with voiceless stops. However, we do examine overlap in the stop+tap clusters as a function of place of articulation.

tokens, 2 with /a/, 3 with /e/, 3 with /o/). This yields a total of 42 total tokens for our data set. Ten repetitions of all tokens were registered for all six speakers (42 × 10 × 6 = 2,520 total tokens).

Tokens were embedded in a carrier phrase *Di X, por favor*, "Say X, please." The vowel /i/ (in the carrier phrase *Di*, "Say") preceded all target onsets, while the vowel that followed the target onset was either /a/, /e/, or /o/, with each of these three vowels equally distributed across the different clusters (see corpus in Appendix A). High vowels, especially in the case of /i/, were excluded following the target onset in order to facilitate parsing the lateral or tap C2 from the following vowel. Finally, it is an empirical question whether the vowel following the cluster affects articulatory timing of the consonants. Some articulatory timing studies report results based on only one (such as Byrd, 1996, /a/) or two (Bombien & Hoole, 2013, /a/ and /i/) vowels. For the present study, we opted to include vowels of two heights (mid and low) as well as two positions in the horizontal dimension (front and back). In this way, we can assess whether and how robust timing in clusters is across the different vowel contexts.

3.2 Data Collection Procedures

A Carstens AG501 3-dimensional electromagnetic articulograph (EMA) was used to register the kinematic movements of the different articulators. The articulatory data were recorded with a sampling rate of 250 Hz. To record articulatory movement, sensors are adhered to the articulators, and the position of these sensors is subsequently recorded as they move through an electromagnetic field. For the current study, three sensors were adhered to midsagittal points on the speaker's tongue: a tongue tip sensor that was attached approximately 1 cm posterior to the tongue tip, a tongue mid sensor that was placed 2 cm posterior to the tongue tip sensor, and a tongue back sensor that was placed 2 cm posterior to the tongue mid sensor. Additional sensors were attached to the upper and lower lip and to the lower incisors (jaw). Reference sensors used to correct for head movement were located on the nose bridge, upper incisor, and behind the ears (left and right mastoid). A triangular bite plane was used with three sensors in order to obtain a reference for the occlusal plane, which is used for translation to a standardized orientation of the vocal tract. A hard palate trace was obtained for each speaker by instructing them to trace the top of their mouth (starting from the teeth and moving back toward the velum) with their tongue tip. Data were corrected for head movement by subtracting head movement captured from the reference sensors on the upper incisor and on the left and right mastoid from the movement of all the sensors. The reference sensors' data were filtered using a cutoff frequency of 5 Hz, while the rest of the sensors' data were filtered using a cutoff frequency of 20 Hz.

3.3 Measurements

Analysis of the kinematic data was carried out using a computer program, MView, developed at Haskins Laboratories by Mark Tiede. The MATLAB-based program displays the acoustic and positional trajectories together with the corresponding instantaneous velocity trajectories, which were calculated by differentiating the positional trajectories. Three-dimensional EMA provides information about vertical, anterior-posterior and lateral movement. For our analysis, we focus on the vertical and horizontal (i.e., up-down/anterior-posterior) movements within the midsagittal plane. The EMA sensor used to quantify the movements of a consonant which was the one that corresponds to the primary oral articulator for that consonant: for [l ɾ] the tongue tip sensor was used, for [k ɡ], the tongue back sensor was employed, and for [b p], lip aperture was used, which is calculated by means of the Euclidean distance between upper and lower lip sensor. To parse the articulatory landmarks for the target plateaus of the consonants, the tangential velocity was used. The onset of a consonant's plateau was defined by the timestamp at which velocity fell below a 20% threshold of the local velocity peak of the *to* movement (the movement towards the target plateau). The release landmark was defined by the timestamp at which velocity surpassed a 20% threshold of the local velocity peak of the *fro* movement (the movement away from the target plateau). Exceptionally, the vertical velocity signal was used instead of tangential velocity to parse velars from the preceding high vowel that appeared in the carrier phrase.

For our study, VOT measures were obtained from simultaneously (with EMA data) collected audio signals using a t.bone EM 9600 unidirectional microphone at a sampling rate of 48 kHz. VOT was defined as the lag between the acoustic burst (marked as c in Fig. 1) following the release of the

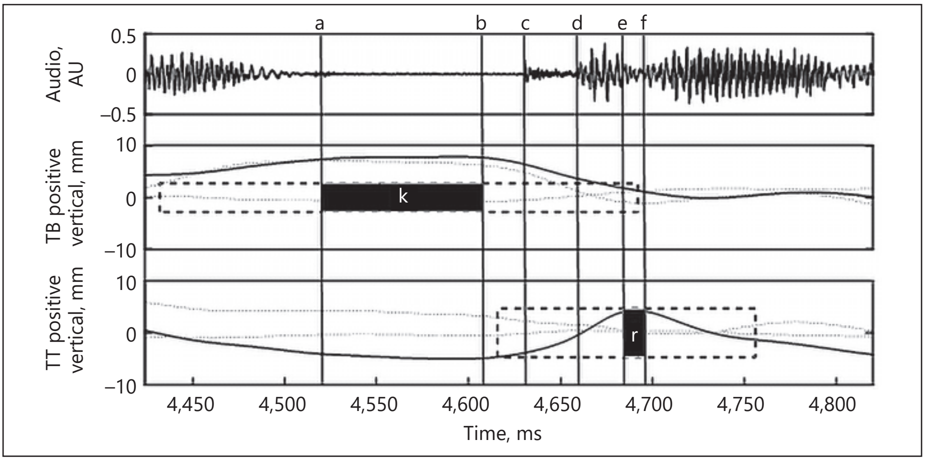


Fig. 1. Waveform (top panel), tongue back movement (middle panel), and tongue tip movement (bottom panel) for *crema* "cream," as produced by a native Spanish-speaking subject. The black rectangles in the middle (TB) and bottom (TT) panels indicate the gestural plateaus of /k/ (middle panel) and /ɾ/ (bottom panel), as generated by the automatic gestural parse algorithm based on tangential velocity signals. The dotted trajectories in the middle and bottom panels represent the anterior-posterior and lateral movement dimensions, while the solid trajectories in the middle and bottom panels represent the inferior-superior movement dimension in millimeters. The *x* axis represents time in milliseconds. Mnemonics "a, b, c, d, e, f" at the top indicate the timestamps used to derive variable values: a, target of [k]; b, release for [k]; c, onset of VOT; d, onset of voicing; e, target of [ɾ]; f, release of constriction for C2.

stop plateau and the instantiation of voicing (marked as d in Fig. 1), as determined using the voice bar in the acoustic signal (spectrogram, not included in Fig. 1 for space reasons). Gestural plateau duration was defined as the interval between the landmarks corresponding to target and release of any given gesture. The target for C1 in Figure 1 is marked as a and the target of C2 is marked as e. The release for C1 in Figure 1 is marked as b while the release for C2 is marked as f. Thus, C1 gestural plateau duration was calculated as (b – a) (= C1 plateau duration), and C2 plateau duration was calculated as (f – e) (= C2 plateau duration). Gestural overlap was defined as the latency between the release of C1 (marked as b in Fig. 1) and the target of the C2 (marked as e in Fig. 1). In measuring gestural overlap via the difference between these two landmarks, (e – b), negative values imply gestural plateau overlap and positive values indicate the existence of an IPI (this overlap measure refers strictly to gestural plateaus and not to the *to* and *fro* movements before and after the gestural plateaus of the two consonants; see Gafos et al., 2010, for relevant discussion).

For quantifying the duration of the intrusive vocoid which may (but does not always) appear between the two consonants in a CC, we need a variable which we term duration of the pre-plateau voicing phase. When voicing before the plateau of the second consonant occurred, our pre-plateau voicing measure was the lag between the timestamp of the acoustic start of voicing (marked as d in Fig. 1) and the articulatory target of the gestural plateau for C2 (marked as e in Fig. 1).

3.4 Statistical Analyses

The effects of the phonological specifications of interest (C1 place/voice, and complex/simplex onset) on the temporal variables were analyzed using the "lme4" package (Bates et al., 2015) in R Studio. C2 was also coded as a predictor in order to test for effects on overlap of the second consonant in clusters (i.e., whether overlap is modulated by the segment that occupies C2). Additionally, as one of our objectives was to test whether overlap patterns generalize across clusters

Table 1. Test variables, predictors, and random effects

Dependent variables	Description
C1 duration	ms
C2 duration	ms
VOT	ms
Overlap	ms
Intrusive vocoid duration	ms
Predictors	**Levels**
Place of articulation	labial, velar
Voice	+voice (voiced), –voice (voiceless)
Onset type	C, CC
Vowel	/a/, /e/, /o/ (only for overlap)
C2	/r/, /l/
Random effects	**Description**
Speaker	6 speakers
Repetition	10 repetitions/target token
Item	42 target tokens

Models were tested for main effects and all interactions.

that precede different vowels (making the results more robust), vowel was also considered as a predictor in our models dealing with gestural overlap. Mixed effects models were preferred over a repeated measures multilevel ANOVA design to avoid potential inflation of the *F* ratio, which results from violations of the sphericity assumption. Histograms were generated before modeling to test for the possible need to perform logit and/or square root transformations, and q-q plots for the residuals were used as well to evaluate normality assumptions of distributions. No transformation of the raw data was performed because it was not warranted to do so, as concluded from the histograms and q-q plot analyses. For certain variables (C1 and C2 duration), there is a very slight right-tail skew in our data. In that case, both square root and stronger log transformations were performed. Results were compared for each transformation with the results using the untransformed data. There was no discrepancy in significance levels for any of the variables. For that reason, results will be reported based on untransformed data.

In all cases, models were tested for main effects and all interactions. Speaker, repetition, and item were modeled as random effects, following usual practice. Both random intercepts by subject and random bysubject slopes were included.[4] Maximum likelihood χ^2 tests based on the de-

[4] Since all of the stimuli in our dataset were attested Spanish words, the frequency of the different stimuli may have an effect on the continuous variables under study (see Bybee, 1999, 2000, 2002; Pierrehumbert, 2002; Baese-Berk & Goldrick, 2009; Goldrick, Vaughn, & Murphy, 2013, among others). To address the possibility of a word frequency effect on the continuous variables during pre-modeling analyses (i.e., before building the linear mixed effects models), log-frequencies calculated from count data using an online database (EsPal) of over two million Spanish words from print, online and multimedia resources were coded for each lexical item. Regression analyses were performed using both the covariance of each pair (log-frequency and timing variable) and normalized Pearson product-moment correlation coefficients. No correlation between word frequency and any of the timing variables addressed in this study reached significance (using either the covariance of each pair or the normalized Pearson product-moment correlation coefficients), so frequency was not modeled as a predictor in the linear mixed effects models.

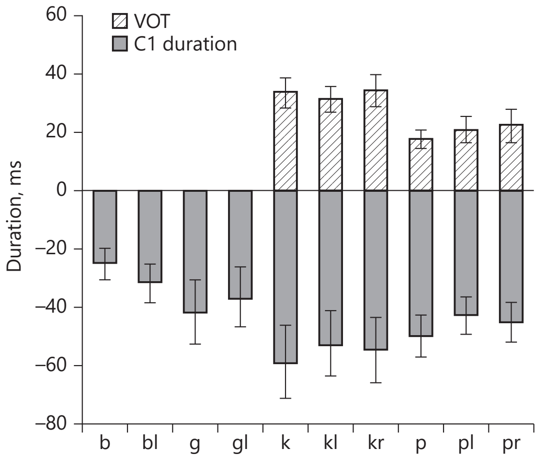


Fig. 2. Mean duration in milliseconds (*y* axis) of C1 constriction and VOT, where the height of the solid dark gray bars represents the duration of constriction for the consonants in the respective clusters (*x* axis) and the height of the bars filled with slanted lines represents VOT. For all clusters, the release of the stop is aligned to 0 in the graph. Recall that a landmark in the articulatory data is used to mark the end of the C1 plateau and a landmark in the acoustic data is used to mark the VOT (i.e., the end of constriction and acoustic releases are not isochronous, and articulatory release and VOT onset are often not abutting as they are in the figure).

viance statistics were performed in order to determine significance. For post hoc comparisons, significance was determined using the Tukey adjusted contrast using the "multcomp" package (Hothorn, Bretz & Westfall, 2008) in R Studio. To address correlations between continuous variables, simple linear regressions were used. Table 1 lists the dependent variables, predictors and random effects used in our statistical analyses of Spanish clusters.

4 Results

4.1 VOT in Singleton Consonants versus Consonants in Clusters

In our data, as expected, stops specified phonologically for voice were produced with (pre)voicing, and phonologically voiceless stops were produced with short lag VOTs. Figure 2 illustrates changes in VOT for voiceless stops across onset types and place categories and variation in C1 plateau duration across onset types, place and voicing (results for C1 plateau duration follow in section 4.2.1).

Strong main effects on VOT were found for place ($\chi^2[1, n = 718] = 38.17, p < 0.001$), but not onset type ($\chi^2[1, n = 718] = 1, p = 0.99$). Pooled means show that VOT for velars (mean = 33.4 ± 9.71 ms) (we use "±" throughout to report standard deviations) is approximately 13 ms longer than VOT for labials (mean = 20 ± 8.53 ms) across onset types. Means pooled across both labial and velar categories show a slight mean decrease for VOT in complex onsets (mean = 25.78 ± 11.24 ms) vis-à-vis simplex onsets (mean = 27.33 ± 11.36 ms), though as the results show for onset type, this difference is insignificant. Additionally, specifically for clusters, the second consonant (C2) was found to have a main effect on VOT ($\chi^2[1, n = 452] = 44.84, p < 0.001$), whereby voicing for laterals (mean = 25.84 ± 9.9 ms) commenced approximately 3 ms sooner than voicing for taps (mean = 29.0 ± 12.62 ms) (see Fig. 2 for an illustration of VOT by consonant and onset type).

Strong 2-way interactions were also found for place and onset type (χ^2[2, n = 718] = 44.58, $p < 0.001$) and place and C2 (χ^2[2, n = 452] = 55.61, $p < 0.001$). For individual onset types, VOT for labial simplex onsets (mean = 17.47 ± 5.45 ms) is nearly half that of VOT in velar simplex onsets (mean = 34.02 ± 9.26 ms). For complex onsets, the difference in VOT between labials (mean = 21.51 ± 9.62 ms) and velars (mean = 32.94 ± 9.97 ms) is slightly smaller than for simplex onsets, though still significant (Tukey HSD: $p < 0.001$). Within place categories in different onset types, VOT in labial simplex onsets (mean = 17.47 ± 5.45 ms) is slightly, though significantly (Tukey HSD: $p = 0.02$), shorter (4 ms) than VOT in labial complex onsets (mean = 21.51 ± 9.62 ms)[5]. This result is inconsistent with the pattern found for VOT in velars, where VOT in simplex onsets (mean = 34.17 ± 9.26 ms) is slightly, though not significantly (Tukey HSD: $p > 0.05$), longer (nearly 1.25 ms) than VOT in complex onsets (mean = 32.94 ± 9.97 ms).

Regarding the place/C2 interaction, for /pl/ clusters, mean VOT values (mean = 20.9 ± 8.3 ms) were only negligibly shorter than values for /pɾ/ (mean = 22.1 ± 11.1 ms) clusters (Tukey HSD: $p > 0.05$), though for velars the difference between VOT in /kl/ (mean = 31.4 ± 8.51 ms) and /kɾ/ (mean = 34.54 ± 11.0 ms) clusters reached significance (Tukey HSD: $p > 0.01$).

In summary, our hypothesis H1 regarding the effects of place of articulation on VOT in Spanish stops bore out: VOT in velars is greater than labials in all onset types. As predicted, no main effect for onset type was found for VOT. VOT for singleton stops was commensurate with VOT for clustered stops. Finally, a main effect was found for VOT based on the second consonant in clusters, whereby VOT before a lateral C2 was shorter than VOT before a tap. Strong interactions were found as well for place and onset type as well as for place and C2.

4.2 Gestural Plateau Duration

4.2.1 Duration of Stops in Simplex and Complex Onsets

This section reports results for the effects of onset type (C, CC), place of articulation, and voicing on singleton and clustered stop duration in Spanish onsets. As predicted in H2, main effects on consonant plateau duration were found for onset type (χ^2[1, n = 1,466] = 4.20, $p = 0.041$), place of articulation (χ^2[1, n = 1,466] = 28.59, $p < 0.001$), and voicing (χ^2[1, n = 1,466] = 52.33, $p < 0.001$). As for the effect based on onset type, singleton onset stops (mean = 44.3 ± 22.6 ms) were approximately 2.2 ms longer than stops in complex onsets (mean = 42.1 ± 20.88 ms). Regarding the effects of place, pooled means reveal that labials (mean = 38.05 ± 17.37 ms) are on average 11 ms shorter than velars (mean = 49.06 ± 14.56 ms). With respect to voicing, voiced consonants (mean = 33.73 ± 18.03 ms) exhibit shorter mean plateau durations than voiceless consonants (mean = 50.75 ± 21.29 ms) pooled across different place and cluster categories (see Fig. 2 for an illustration of plateau duration by consonant and onset type).

A strong 3-way interaction was found between place, voice, and onset type (χ^2[3, n = 1,466] = 97.25, $p < 0.001$). Duration of voiced labials increases as a function of

[5] The results for labials (i.e., an increase in VOT from simplex to complex onsets) seems contradictory to the results for the pooled means for VOT across different onset types presented in the previous paragraph. However, the results presented in the previous paragraph included both place categories. Thus the different directions of the trend (for velars)/effect (for labials) across different onset types cancelled each other out, hence no main effect for VOT across onset types.

Table 2. Means and standard deviations in milliseconds for lateral and tap duration in different C1C2 clusters

Cluster	Lateral duration	
	mean	SD
/gl/	40.55	20.86
/bl/	37.37	22.08
/kl/	37.41	14.91
/pl/	35.75	18.07
/kɾ/	19.50	6.38
/pɾ/	17.68	8.99

onset type, where voiced labials in simplex onsets show a mean plateau duration of 25.84 ms (SD ±10.11 ms) while plateau duration of C1 voiced labials in complex onsets has a slightly higher, though significant (Tukey HSD: $p < 0.05$), mean of 31.62 ms (SD ±14.54). For C1 involving voiceless labials, however, an opposite direction emerged, whereby plateau duration decreases by roughly 5 ms as syllable complexity increases (voiceless labials in C: mean = 49.89 ± 17.63 ms; voiceless labials in CC: mean = 44.09 ± 16.17 ms) (Tukey HSD: $p < 0.01$). The decrease in plateau duration as a result of onset complexity also characterizes the changes in duration for voiced and voiceless velars, plateau duration for both voiced and voiceless velars decreasing by approximately 5 ms as onset complexity increases (voiced velars in C: mean = 41.79 ± 21.39 ms; voiced velars in CC: mean = 37.04 ± 22.06 ms; voiceless velars in C: mean = 59.04 ± 24.39 ms; voiceless velars in CC: mean = 54.0 ± 23.99 ms) (Tukey HSD: $p < 0.01$ for all voiced/voiceless velar contrasts in C and CC).

To review, all three predictors (onset type, place, and voicing) were found to affect the duration of the stop as predicted in H2. Duration of the stop reduces in complex onsets (in all cases except with voiced labials), as compared to stops in singleton onsets, while labials are shorter than velars in all contexts. Finally, as predicted, voiced stops are shorter than their voiceless counterparts in both singleton and complex onsets.

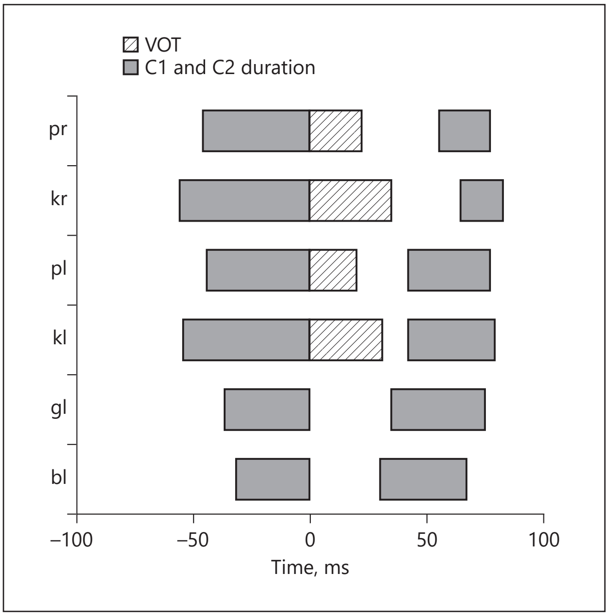
4.2.2 Duration of Sonorants in Simplex and Complex Onsets

Here we first examine the duration of C2 as a function of the segment that occupies this position (/ɾ/ or /l/) and then investigate whether the plateau duration of C2 changes as a function of the phonological specifications of C1 (voice and place of articulation) and onset complexity. Subsequently, we use a subset of the data (/bl, pl, gl, kl/) to examine effects of voicing, and to compare lateral duration in clusters to /lV/ onsets (i.e., thus, testing for an effect of onset complexity on the lateral). For effects of place, results of both stop + lateral and stop + tap clusters are presented.

The results reveal that duration of C2 is significantly accounted for by the segment that appears in this position ($\chi^2[1, n = 633] = 47.93$, $p < 0.001$), C2 duration being significantly longer for the lateral than for the tap. The tap (mean = 18.63 ± 7.89 ms) shows a mean duration nearly half of that of the lateral (mean = 37.84 ± 18.27 ms). Table 2 reports means and standard deviations for lateral and tap duration in different C1C2 clusters.

Concerning the prediction regarding the effects of the phonological specifications of C1 on C2 timing, no effect was found for the duration of C2 as a function of

Fig. 3. Mean duration in milliseconds (*x* axis) of C1 and C2 constriction (in dark gray) and VOT (slanted lines). The *x* axis represents time in milliseconds, while the *y* axis shows the individual clusters. IPI is the time between the release of C1 and the target of C2. For all clusters, the release of the C1 plateau is aligned to 0 in the graph. Recall that a landmark in the articulatory data is used to mark the end of the C1 plateau and a landmark in the acoustic data is used to mark the VOT (i.e., the end of constriction and acoustic releases are not isochronous, and articulatory release and VOT onset are often not abutting as they are in the figure).



C1 place for either stop + tap ($\chi^2[1, n = 235] = 1, p = 0.99$) or stop + lateral ($\chi^2[1, n = 898] = 3.45, p = 0.07$) clusters.

Figure 3 plots mean C1 and C2 duration and VOT for the voiceless stops by cluster. IPI (statistical results appear in the following section 4.3), it will be recalled, is defined as the latency from the release of C1 (left gray bars) to the onset of C2 (right gray bars).

In order to address the effects of C1 voicing on C2 duration, we only considered /bl, pl, gl, kl/ clusters since voiced stop + tap sequences were not included in our corpus. Voice was shown to reliably account for changes in C2 duration ($\chi^2[1, n = 898] = 7.16, p < 0.01$), whereby the lateral in voiced clusters (mean = 40.33 ± 20.48 ms) was approximately 4 ms longer than in voiceless clusters (mean = 36.40 ± 16.78 ms).

In order to address the effects of onset complexity on the duration of C2, we examined a subset of our data that included stop (both voiced and voiceless) + lateral clusters (/bl, pl, gl, kl), as well as /lV/ tokens. The main question we examined was whether there is lateral reduction in complex onsets as compared to word initial simplex onsets (for example, *globo* vs. *lobo*, "balloon" and "wolf," respectively). The clusters with taps were eliminated from this analysis since Spanish-specific phonotactics do not permit a straightforward analysis of changes in duration between word initial simplex onsets and C2 rhotics in clusters. The rule for Spanish is that in word initial position the rhotic must be realized as a trill, whereas when occupying the C2 in C1C2 clusters it must be realized as a tap. This, along with other distributional idiosyncrasies, serves as the basis of the argument that Spanish has two phonologically specified rhotics, a tap and a trill (Martinez Celdran, 2001). It is known that trill duration, independently of syllable position, is longer than the duration for taps (Quilis, 1993). Thus, comparing rhotics in different onset types is not warranted, since any dura-

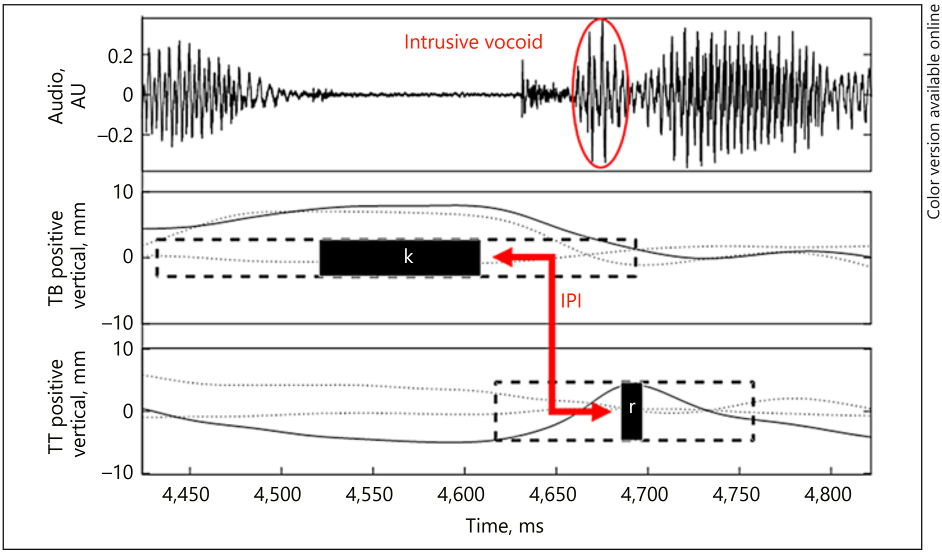


Fig. 4. From top to bottom are the waveform, tongue back (TB) and tongue tip (TT) signals for token *crema* "cream." In the waveform signal, the red ellipses delimit the acoustic realization of the intrusive vocoid. The black rectangles in the TB and TT signals represent the gestural plateaus for the different consonants in the onset cluster. Dotted trajectories in the middle (TB) and bottom (TT) panels represent the tongue back and tongue tip movement in the horizontal and lateral dimensions, while the solid trajectories in the middle and bottom panels represent movement in the vertical dimension in millimeters. The *x* axis represents time in milliseconds. The latency from the release of the plateau for /k/ in the middle panel to the target for /ɾ/ in the bottom panel is the interplateau interval, or IPI, as indicated in red.

tional differences between the two (trill vs. tap) cannot be solely attributable to changes in onset complexity.

The results reveal a main effect of onset complexity on lateral duration ($\chi^2[1, n = 929] = 77.95$ $p < 0.001$) whereby laterals in simplex onsets (mean = 47.78 ± 22.43 ms) show a mean of approximately 9 ms longer than laterals in complex onsets (mean = 38.64 ± 19.0 ms). This result did not obtain for all speakers, though it did for most.

Recall that Blecua and Machuca (2000) and Blecua et al. (2003) reported on the basis of acoustic data that C2 duration varies as a function of the duration of the C1 stop. To address this possibility, simple linear regressions were performed after observing q-q plots to ensure a normal distribution of the residuals. As the data were slightly right skewed, models were tested with both log-transformed and square root-transformed C1 and C2. However, there was no difference in significance levels between the raw and transformed variables. Results show no correlation in our data between the duration of C1 and C2 for either stop + lateral ($F = 0.95$, $p = 0.33$, $R^2 < 0.001$) or stop + tap ($F = 1.38$, $p = 0.24$, $R^2 = 0.001$) clusters.

In sum, as was predicted in H2, laterals in clusters were significantly longer than taps. For laterals in clusters, our hypotheses regarding the effects of place and voice did not bear out. That is, no significant effect of place or voice of C1 was found on the duration of laterals in clusters. As for voiceless stop + tap clusters, the place specification for C1 had no effect on tap duration, which is in line with the results we found for lateral duration in stop + lateral clusters. Likewise, we examined the possible correlation between stop duration in C1 and lateral duration in C2, but results were not significant.

4.3 Gestural Overlap in C1C2

We first turn to effects of place of articulation and voicing specifications of C1 on intergestural timing between C1 and C2, turning to address the possible effects of the following vowel on patterns of C1C2 timing after the basic patterns have been established. To recall, we expect Spanish to pattern like French (Bombien & Hoole, 2013), which exhibits relatively low overlap between C1 and C2. We also expect some differentiation in C1C2 overlap on the basis of previous literature on intrusive vocoids which show a more robust presence of vocoids in stop + tap clusters as compared to stop + lateral clusters (Malmberg, 1965, among others). Based on Bombien and Hoole (2013) on French, it is reasonable to expect no effect of voicing on overlap since Spanish and French have comparable voicing instantiation systems in their contrast between voiced and voiceless stops.

As was predicted, our results confirm that the Spanish consonant clusters in our study are typified by a robust IPI. Figure 4 offers an illustration of the kinematics where the target gestural plateaus for the different consonants do not overlap in time (although the "fro" and "to" phases of the gestural movements do overlap), resulting in an intrusive vocoid. The shaded black rectangles in the middle and bottom signals represent the gestural plateaus for the consonants /k/ and /ɾ/ in the Spanish word *crema* "cream." This basic configuration or mode of coordination is exhibited by all speakers. The latency from the end of the target gestural plateau for /k/ to the onset of the gestural plateau for /ɾ/ is the IPI.

With regard to differences in IPI as a function of the gestures that occupy C2, clusters with laterals (mean = 41.41 ± 21.22 ms) show significantly shorter IPIs between gestures than do clusters with taps (mean = 62.53 ± 18.33 ms) ($\chi^2[1, n = 445] = 48.13, p < 0.001$).

In the stop + lateral subset (/bl, pl, gl, kl/), a main effect of C1 place was found for overlap ($\chi^2[1, n = 609] = 8.98, p = 0.003$), whereby velar clusters (i.e., /kl/, /gl/) (mean = 47.59 ± 16.3 ms) exhibit lower overlap (higher IPIs) than clusters with C1 labials (i.e., /pl/, /bl/) (mean = 37.78 ± 13.46 ms).

As for effects of C1 voicing on overlap in stop + lateral clusters, a main effect of C1 voicing was found ($\chi^2[1, n = 609] = 20.50, p < 0.001$) such that IPI in sequences where there is no change in voicing across C1 and C2 (i.e., /bl/, /gl/) (mean = 31.55 ± 21.11 ms) exhibit systematically shorter IPI values (thus more gestural overlap) than clusters where there is a change in voicing categories across C1 and C2 (i.e., /kl/, /pl/) (mean = 41.82 ± 21.13 ms). A strong interaction between C1 place and voice on IPI obtained as well ($\chi^2[2, n = 609] = 31.35, p < 0.001$).

Mean values for IPI in clusters with voiced labials (mean = 29.74 ± 19.80 ms) are approximately 11 ms shorter than the IPI means in voiceless labial sequences (mean = 41.89 ± 18.78 ms). For velar + lateral sequences, the difference is not as dramatic, yet still significant (Tukey HSD: $p < 0.01$), clusters with voiced velars (mean = 34.61 ± 22.93 ms) showing a difference of roughly 7 ms compared with sequences involving voiceless velars (mean = 41.72 ± 24.03 ms). We note that in French consonant-lateral clusters, Bombien and Hoole (2013) report no effect of voicing on over-

DOI: 10.1159/000501508

Table 3. Means and standard deviations in milliseconds for IPI in stop + tap and stop + lateral clusters by C1 place

Place C1	IPI				Significance
	/Cr/		/Cl/		
	mean	SD	mean	SD	
Labial	55.0	19.1	41.58	18.48	***, /Cr/>/Cl/
Velar	66.4	14.6	41.21	24.11	***, /Cr/>/Cl/
Significance	***, Vel>Lab		ns		

*** p < 0.001.

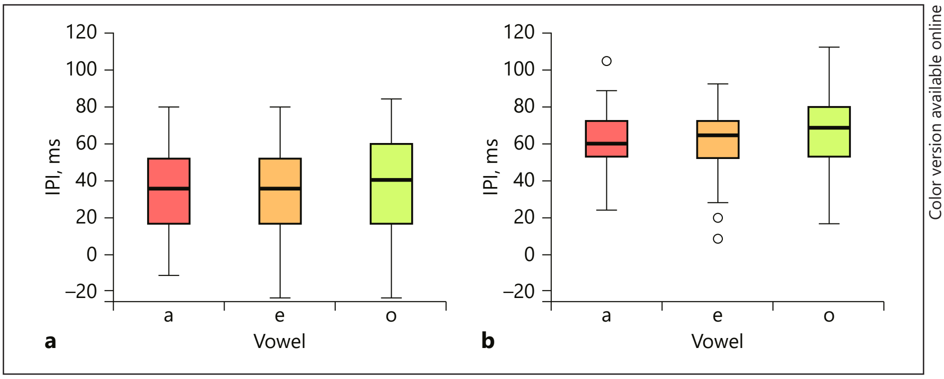


Fig. 5. Boxplots of IPI in milliseconds (*y* axis) for clusters appearing before /a/, /e/ and /o/ (*x* axis). **a** IPI in stop + lateral clusters. **b** IPI in stop + tap clusters.

lap patterns. This contrast between our results and those of Bombien and Hoole (2013) may appear surprising given that voicing is implemented in French in the same way as in Spanish (that is, by a true voicing vs. short lag VOT opposition). We take up this contrast further in the Discussion.

For voiceless stop + tap clusters, a main effect of place was found whereby IPI was systematically longer in velar + tap as compared to labial + tap clusters ($\chi^2[1, n = 445] = 13.88, p < 0.001$). IPI in velar stop + tap clusters (mean = 66.4 ± 14.6 ms) is approximately 12 ms longer than in labial stop + tap sequences (mean = 55.0 ± 19.91 ms).

Means and standard deviations for IPI in different C1C2 clusters by C2 segment and C1 place are provided in Table 3, along with significant codes and the direction of the effect.

Finally, concerning the possible effects of the following vowel on patterns of gestural overlap, the results of the linear mixed effects models reveal no effect of the following vowel on the timing between gestures in either stop + lateral ($\chi^2[1, n = 610] = 3.76, p = 0.11$) or stop + tap ($\chi^2[1, n = 239] = 2.13, p = 0.09$) clusters. IPI in millisec-

Phonetica 2019;76:448–478
DOI: 10.1159/000501508

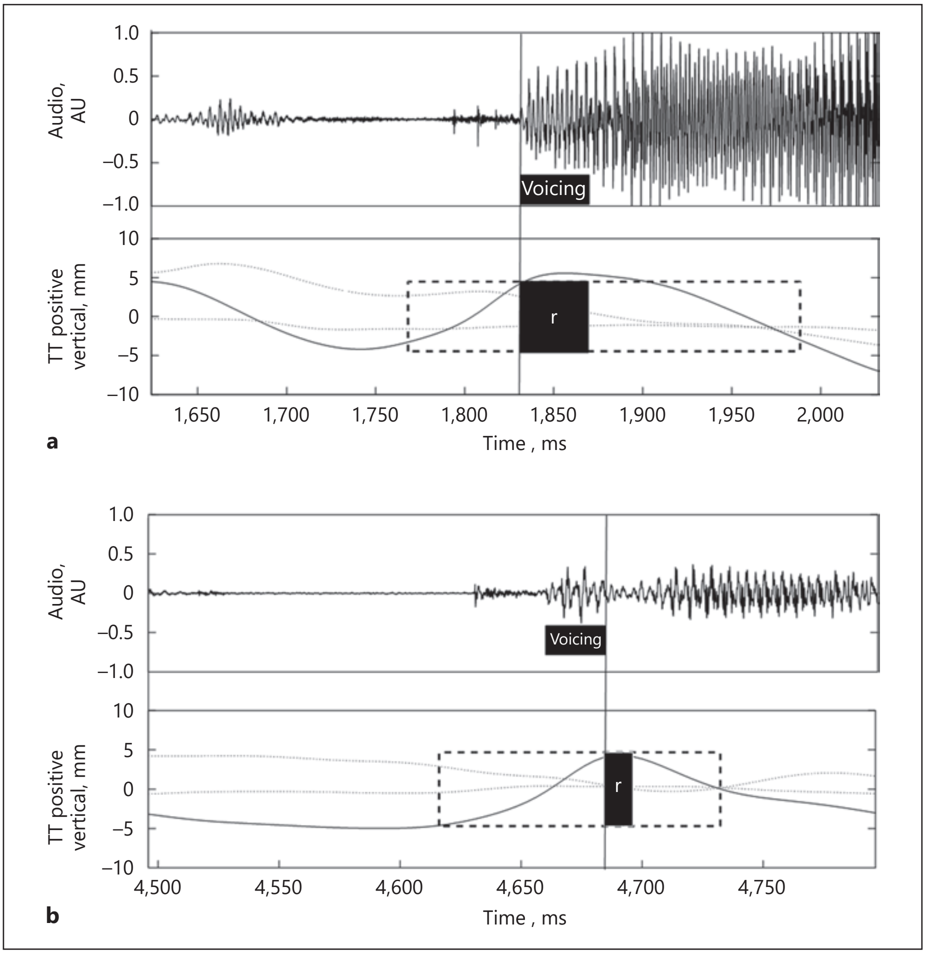


Fig. 6. Acoustic waveforms (top panels in **a** and **b**) and tongue tip (TT) articulatory signals (bottom panels in **a** and **b**) for the token *crema* "cream." Rectangles in the bottom (TT) panels represent the tongue tip segmentation for the tap. The dotted trajectories in the articulatory movement signals (bottom panels in **a** and **b**) represent the horizontal and lateral dimensions of movement, while the solid trajectories represent movement in the vertical dimension in millimeters. The *x* axis represents time in milliseconds. **a** Laryngeal action is aligned to the gestural plateau of the tap (no intrusive vocoid). **b** Laryngeal action precedes the gestural plateau of the tap, producing what has been called in the acoustic literature an intrusive vocoid.

onds (*y* axis) for clusters (stop + lateral on the left, stop + tap on the right) preceding each vowel (*x* axis) is illustrated in Figure 5.

For stop + lateral clusters, means for IPI were consistent across different vowels. Clusters preceding /a/ /e/, and /o/ showed similar means (/a/: mean = 34.2 ±19.9 ms; /e/: mean = 35.8 ± 21.2; /o/: mean = 38.6 ± 23.1). This trend held for stop + tap clusters as well, where means for IPI preceding /a/, /e/, and /o/ showed higher overall means than in stop-lateral clusters (/a/: mean = 60.2 ± 15.7 ms; /e/: mean = 61.3 ±

DOI: 10.1159/000501508

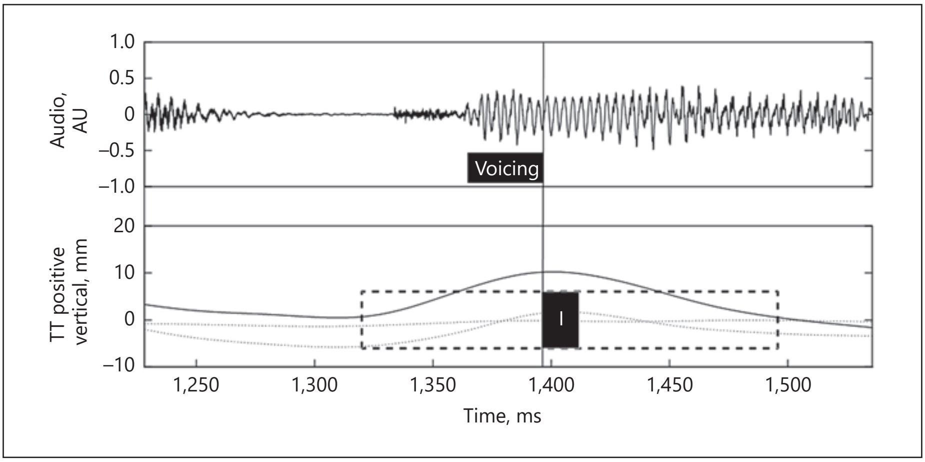


Fig. 7. Acoustic (top panel) and articulatory signal (bottom panel) of /l/ for the token *clema* "conduit piece used for an electrical connection." Dotted trajectories in the bottom panel represent the tongue tip (TT) movement in the anterior-posterior and lateral dimensions, while the solid trajectory represents the movement in the top-down dimension in millimeters. The *x* axis represents time in milliseconds. As can be observed in the top panel, laryngeal action precedes the gestural plateau of the lateral, producing what has been called in the acoustic literature an intrusive vocoid.

17.9; /o/: mean = 65.4 ± 18.3), though no significant differences were based on the following vowel.

In sum, our assessment of overlap in clusters reveals the following results. As predicted in H3, IPI was found to be modulated by the gesture that occupies the C2 position. IPI was significantly higher in clusters where the C2 is a tap as compared to laterals. C1 place was also found to affect the overlap of the gestural plateaus in both stop + lateral and voiceless stop + tap clusters. Additionally, we hypothesized in the Introduction that Spanish would pattern with French with regard to low gestural overlap for voiceless stop clusters (H3) and would show no effect of C1 voicing on intergestural timing in C1C2 clusters (H3). Our predictions were partially confirmed in that Spanish clusters are characterized by low gestural overlap. However, in contrast to French, this interval is additionally modulated by the voicing specification of C1, which resembles Bombien and Hoole's (2013) finding on German but not on French (we return to this finding in the Discussion section). Finally, the results indicate that overlap in our data is not modulated by the vowel that follows the consonant clusters.

4.3.1 Intrusive Vocoid (Pre-Plateau Voicing) Duration

A long series of studies report the existence of an intrusive vocoid in acoustic records of Spanish stop + tap and stop + lateral clusters (Blecua, 2001; Blecua et al., 2003; Bradley, 2006; Gili Gaya, 1921; Lenz, 1892; Malmberg, 1965; Menéndez Pidal, 1926, among others). Here we intend to show how oral-laryngeal and oral-oral coordination patterns give rise to the previously observed acoustics.

Figure 6 illustrates two different instances of the timing of voicing in relation to the gestural plateau for [ɾ] for a token *crema* "cream," spoken by the same speaker. In

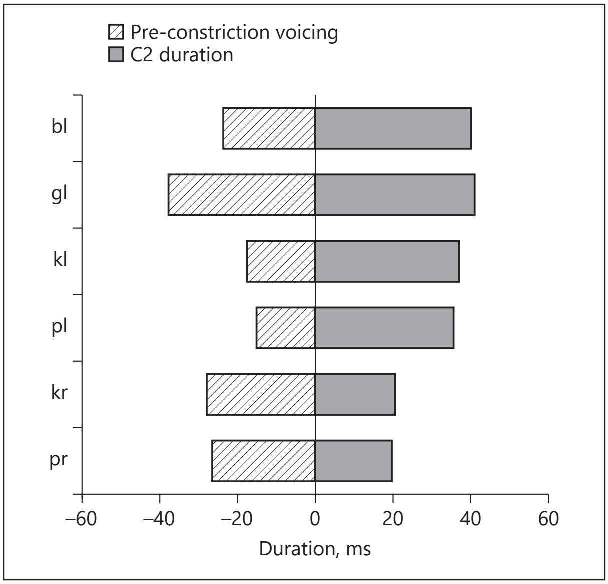


Fig. 8. Mean duration in milliseconds (*x* axis) of pre-plateau voicing and C2 plateau, where the bars with the slanted lines represent pre-plateau voicing and the solid dark gray bars represent the plateau duration of C2 per cluster (*y* axis). For all clusters, the target of the C2 plateau is aligned to 0 in the graph for ease of interpretation; however, this line-up at zero is for display purposes only as it conflates the acoustic and articulatory definitions of the zero point.

both panels, the gestural plateau of [ɾ] is delineated by the shaded black rectangles. In the top panel, voicing is roughly aligned with the gestural plateau of the tap where the intrusive vocoid does not appear, while in the bottom panel, voicing precedes the gestural plateau, producing an acoustic consequence reminiscent of a vowel.

Of the total number of C1C2 clusters (n = 764), pre-plateau voicing occurred in 255 of them. As expected, pre-plateau voicing was more prevalent in stop + tap clusters, where out of 272 total tokens, pre-plateau voicing occurred in 188 (or 69.1% of the total cases). For stop + lateral clusters, pre-plateau voicing occurred in 67 out of 492 C1C2 clusters (or 13.6% of the total cases). Previous studies have emphasized the presence of the intrusive vocoid in stop + tap clusters in Spanish, though little attention has been directed toward vocoid intrusion in stop + lateral clusters. For concreteness, Figure 7 gives a representative example of pre-plateau voicing in stop + lateral clusters using the token *clema* "conduit piece used for an electrical connection."

Concerning the duration of pre-plateau voicing across stop + lateral and stop + tap sequences, C2 had no effect on voicing duration ($\chi^2[1, n = 255] = 1, p = 0.99$), pre-plateau voicing in stop + lateral sequences (mean = 25.93 ± 16.17 ms) being roughly equal to the duration of voicing in stop + tap clusters (mean = 25.72 ± 10.04 ms). In stop + lateral clusters, main effects were found for C1 voice ($\chi^2[1, n = 67] = 11.20, p < 0.001$) and place ($\chi^2[1, n = 67] = 6.62, p = 0.01$) on the duration of pre-plateau voicing, whereby voicing following voiced C1 consonants (mean = 27.1 ± 13.2 ms) is significantly longer than voicing following voiceless C1 (mean = 16.64 ± 8.96 ms) for both place categories (/bl/: mean = 23.32 ± 12.22 ms; /pl/: mean = 15 ± 8.03 ms; /gl/: mean = 37.67 ± 18.92 ms; /kl/: mean = 17.56 ± 9.88 ms). This leads to a strong 2-way interaction between C1 place and voice on pre-plateau voicing ($\chi^2[2, n = 67] = 23.50, p < 0.001$) for stop + lateral clusters (Tukey HSD: $p < 0.001$ for all voice/place contrasts).

Fig. 9. Mean duration in milliseconds (x axis) of C1 and C2 duration (dark gray bars), VOT (left slanting lines) and pre-plateau voicing (right slanting lines) per cluster (y axis). The gray-blue area between VOT (or C1 duration for voiced stops, where there is no VOT) and pre-plateau voicing represents the range of variation (as determined by standard deviations for VOT and pre-plateau voicing) of these two variables. The time between the release of C1 (left gray bar) and the target of C2 (right gray bar) is the IPI. For all clusters, the release of the C1 plateau is aligned to 0 in the graph for ease of interpretation. Recall that a landmark in the articulatory data is used to mark the end of the C1 plateau and a landmark in the acoustic data is used to mark the VOT (i.e., the end of constriction and acoustic releases are not isochronous, and articulatory release and VOT onset are often not abutting as they are in the figure).

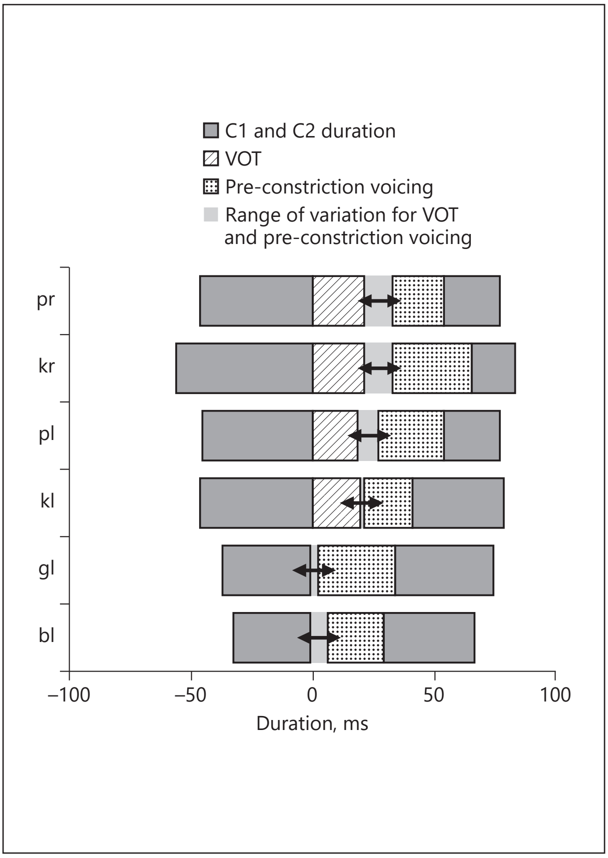


Figure 8 illustrates the duration of pre-plateau voicing in relation to the gestural plateau for C2.

For stop + tap clusters, no effect of C1 place was found on the duration of pre-plateau voicing in /pɾ/ (mean = 26.38 ± 10.4 ms) and /kɾ/ (mean = 27.65 ± 10.4 ms) clusters ($\chi^2[1, n = 188] = 1, p = 0.99$) (recall that our corpus did not contain /bɾ/ and /gɾ/ so we could not test for effects of voice in voiced stop + tap clusters).

For stop + lateral clusters, pre-plateau voicing was found to be negatively correlated (albeit, weakly) to C2 duration ($F = 26.92, p < 0.001$) with an R^2 of 0.30, such that pre-plateau voicing decreases as C2 duration increases. This result did not obtain for stop + tap clusters ($F = 1.55, p = 0.21, R^2 = 0.003$), where the duration of pre-plateau voicing was fairly impervious to changes in C2 duration. This is consistent with the result from the overlap analysis on the substantially longer duration of IPI in stop + tap in comparison to stop + lateral clusters.

Figure 9 illustrates mean duration of the gestural plateaus for C1 and C2, VOT and pre-plateau voicing. The lag between the right edge of C1 and the left edge of C2 represents mean IPI in Spanish clusters. The shaded area between VOT and pre-plateau voicing represents the temporal range where the end of VOT and the start of the pre-plateau voicing fall (in any given token, there is of course no latency between the end of VOT and the start of pre-plateau voicing, but such a latency may exist between means, depicted by the rectangles, across tokens). This shaded area thus indicates the temporal range where the two measures (VOT and pre-plateau voicing) fall in.

As predicted in H4, pre-plateau voicing was found to be more prevalent in stop + tap clusters than in stop + lateral clusters, though C2 did not affect the duration of voicing. For stop + lateral clusters, it was found that pre-plateau voicing was modulated by the place and voicing of the C1; pre-plateau voicing duration was longer in clusters where the C1 was voiced across both place categories. For stop + tap clusters, no effect of place was found. Finally, and in line with H4 for stop + lateral clusters, a weak correlation was found between C2 duration such that pre-plateau voicing decreases as the duration of C2 increases.

5 Summary and Discussion

We examined oral-laryngeal and oral-oral coordination patterns in C1C2 clusters using articulatory and acoustic data from six native speakers of Standard Peninsular Spanish. In this section, we summarize our results and relate the obtained articulatory and acoustic patterns for Spanish to previous studies dealing with gestural timing in onsets.

We begin with the ground-level results for VOT (H1) and consonant duration (H2) which establish continuity between previous studies and our present work. For VOT, as predicted in H1, place of articulation showed the strongest effect, velars exhibiting longer VOTs than labials. Also in line with H1 and the results reported in Gibson et al. (2015), VOT does not vary as a function of syllable complexity in our data, singleton voiceless stops showing commensurate VOT to clustered voiceless stops. However, the results do indicate significant interactions between syllable complexity and place, as well as syllable complexity and the identity of the C2 in clusters. It is difficult to speculate as to the motivations of these interactions, since the differences between the different categories are quite small and could be attributable to small fluctuations in other parameters which were not controlled in our design (e.g., speech rate).

As for stop duration (H2), both voice and place showed strong effects, as predicted in H2, voiced consonants being significantly shorter than voiceless consonants, and velars consistently exhibiting longer duration than labials. A significant effect of onset complexity was found for C1 duration in all place/voice contrasts except voiced labials, where the direction of the effect, though still significant, was opposite to that found for voiced and voiceless velars and voiceless labials (where duration of C1 reduces as onset complexity increases). For voiced labials, C1 duration in complex onsets was approximately 6 ms longer than for voiced labials in simplex onsets.

Likewise, and also in accordance with H2, for C2 duration a strong effect of syllable complexity was found for laterals, such that laterals in complex onsets are significantly shorter than laterals appearing as a word initial simplex onset. No main effects were found for the place of articulation of C1 on lateral duration in either stop + lateral or stop + tap clusters, though in stop + lateral clusters a main effect was found for the voice specification of C1, supporting previous acoustic (Blecua et al., 2003) and electropalatographic studies (Gibson et al., 2015) dealing with Spanish clusters.

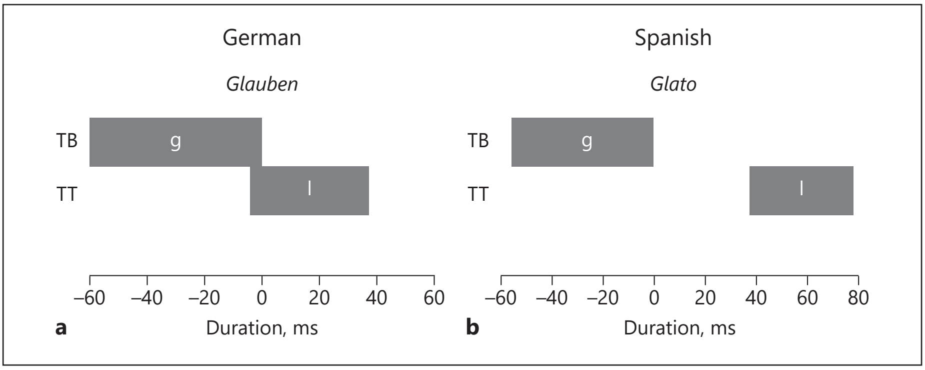


Fig. 10. Representative schemas of gestural overlap for the cluster /gl/ in German (**a**, from the word *glauben* "to believe") and Spanish (**b**, from the word *Glato* "proper name of town in northern Italy"). Spanish shows a substantial IPI or time lag between the plateaus (constriction phases) of the two consonants, as shown by the separation between the dark rectangles in the right-hand schematic, whereas in German the plateaus overlap in time.

The absence of a main effect of C1 place on lateral duration in complex onsets is in contrast to previous studies dealing with timing in Spanish clusters (Blecua et al., 2003; Gibson et al., 2015), though this could be attributed to methodological and labeling differences between the studies (acoustic labeling vs. articulatory labeling using velocity thresholds) and the nature of the different sources of data. For example, the study of Ying et al. (2017) tracked lateral channel formation by placing sensors parasagittally on the tongue blade and found that tongue tip raising, tongue mid lowering and tongue back retraction vary across environments in order to maintain consistent timing of lateral channel formation. Thus, to the extent that these results generalize to peninsular Spanish, variation in tongue tip timing may follow from tradeoffs between multiple articulators that are coordinated to maintain stable timing of lateral channel formation in /l/, and not solely from the tongue tip gesture, which is the trajectory measured in our study. Further exploration of these issues must be left to a follow-up study using more appropriate sensor placements and more focused stimuli than our present study which aimed at establishing a first empirical basis for guiding such further studies.

With regard to gestural overlap (H3), the predictions in H3 of low overlap were borne out by the results. The Spanish clusters in our data sets are characterized by a positive IPI. To illustrate, Figure 10 (from Sotiropoulou, Gibson, Tobin, & Gafos, 2018) juxtaposes an example from /gl/ in *Glato* (name of town in northern Italy) in our Spanish data and the corresponding cluster in *glauben* "to believe" from German data we have collected independently. As shown in the figure, the gestural plateaus in German voiced clusters overlap temporally, whereas in Spanish, the two plateaus are separated by a substantial lag.

In line with our hypothesis in H3, overlap was modulated by the consonant which occupies C2: overlap in stop + tap clusters is significantly lower than in stop + lateral clusters. Our data indicate an articulatory basis for this difference in overlap. The tap is a sequence of two gestures; a tongue back-lowering gesture followed by a

Phonetica 2019;76:448–478
DOI: 10.1159/000501508

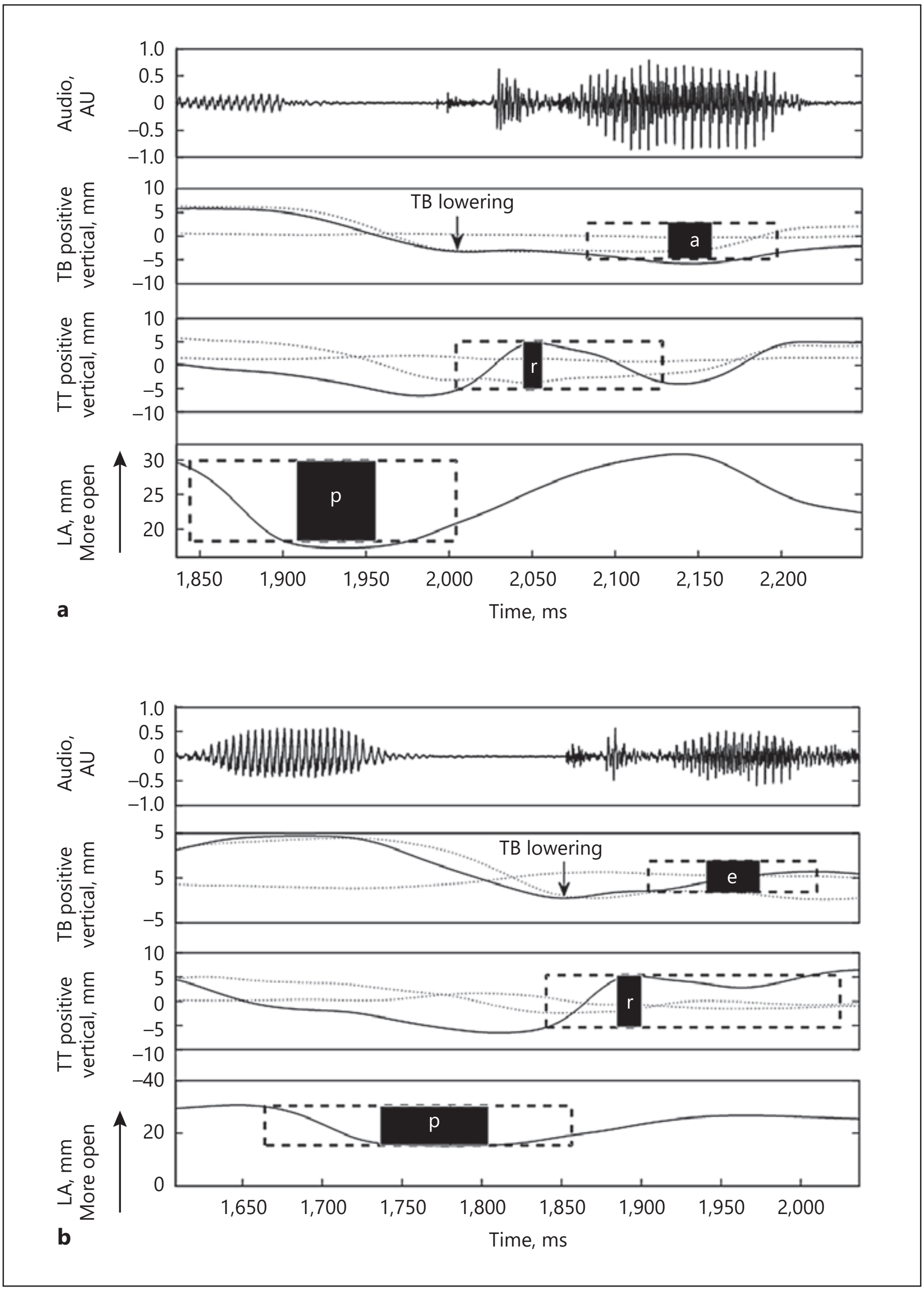
forward and up gesture of the tongue tip. We illustrate with examples from the sequences /pɾe/ and /pɾa/ shown in Figure 11a and b. In Figure 11a, /pɾe/, even though a mid vowel /e/ follows the cluster, the tongue back first lowers after the release of the labial and before the tongue tip gesture for /ɾ/ which is in turn followed by the raising of the tongue back for /e/. The same sequencing of a tongue back gesture preceding the tongue tip gesture can also be seen in Figure 11b, /pɾa/. Between the lips and tongue tip plateaus, the extra tongue back gesture for the tap can be observed. The spatial target for that tongue back gesture for the tap is different from that of the following low vowel /a/ (specifically, in the top-down dimension, which is the lower of the two lines in the tongue back panel (the other line being the front-back dimension).

The tap is thus a sequence of two gestures; a first tongue back-lowering gesture followed by a forward gesture of the tongue tip. This sequencing between the tongue back gesture and the subsequent tongue tip gesture for the /ɾ/ is what results in the lower overlap for /Cɾ/ than for /Cl/. This characterization is in line with Baltazani (2009) and Savu (2012) who report results of the tap in Greek clusters and in intervocalic contexts. The results of their cluster analyses suggest (just as in our case) a complex tongue body behavior for the tap which inhibits overlap with the C1. To recall, these authors describe the tap in clusters as a short and fast supralaryngeal movement that is imposed over a longer laryngeal gesture, which produces an intrusive vocoid. In their justification, the intrusive vocoid forms part of the tap, though our data show that the vocoid is an acoustic output with a nonunique articulatory source, meaning it can appear in both stop + tap and stop + lateral clusters. Thus, while our articulatory and acoustic data coincide with Baltazani (2009) and Savu (2012) in that the tap is a complex gesture consisting of a tongue back movement followed by a short and fast movement of the tongue tip, our data show the intrusive vocoid is not exclusive to clusters with taps.

Fig. 11. a Kinematics of sequence [pɾa]. From top to bottom are the audio, tongue back, tongue tip, and lip aperture signals. Gestural plateaus are indicated by black rectangles. The second (tongue back, TB) and third (tongue tip, TT) panels from the top show the tongue back and tongue tip movement trajectories in the anterior-posterior and lateral dimensions. The solid trajectory in the second and third panels from the top represent tongue back and tongue tip movement in the top-down dimension in millimeters. The solid trajectory in the bottom panel (lip aperture) represents the degree of lip aperture. The *x* axis represents time in milliseconds. Between the lip and tongue tip plateaus for the /p/ and /ɾ/ respectively, the extra tongue back gesture for the tap can be observed (see superimposed "TB lowering" landmark in the trajectory). The positional target for that tongue back gesture is different from that of the following low vowel /a/. As can be seen in the top-down dimension (solid trajectory), the positional target for the tongue back gesture for the tap is higher than that for the vowel /a/. **b** Kinematics of a [pɾe]. From top to bottom are the audio, tongue back (TB), tongue tip (TT), and lip aperture signals. Gestural plateaus are indicated by black rectangles. Dotted trajectories in the second (tongue back) and third (tongue tip) panels from the top represent the tongue back and tongue tip movement in the anterior-posterior and lateral dimensions. The solid trajectories in the second and third panels from the top represent tongue back and tongue tip movement in the top-down dimension in millimeters. The trajectory in the bottom panel (lip aperture) represents the degree of lip aperture. The *x* axis represents time in milliseconds. Between the lip and tongue tip plateaus, the extra tongue back gesture for the tap can be observed (see superimposed "TB lowering" landmark in the trajectory). The positional target for that tongue back gesture is different from that of the following mid vowel /e/. As can be seen in the top-down dimension (solid trajectory), the positional target for the tongue back gesture for the tap is lower than that for the vowel /e/.

(For figure see next page.)

In contradiction to H3, which was motivated on the basis of results from French in Bombien and Hoole (2013), our data indicate that C1 voicing modulates overlap patterns in Spanish clusters. In our Spanish voiced stop + lateral clusters, the IPI is shorter (mean: 10 ms) than in voiceless stop + lateral clusters. In a study of the effects of voicing



11

of C1 on overlap in stop + lateral (but not stop + rhotic) clusters of French, Bombien and Hoole (2013) report no such effect of voicing. This may seem surprising given the similarities between French and Spanish in terms of their voicing implementation systems. However, as anticipated in the Introduction, how intrasegmental timing (the timing between the laryngeal abduction gesture and the oral constriction for the stop) interacts with intersegmental timing (the lag between the two oral constrictions in a cluster) is a topic where much empirical terrain needs to be charted before differences of this sort can be evaluated properly. At this stage, at least two possible (not mutually exclusive) sources of the observed difference should be kept in mind and pursued in future work. First, differences in supralaryngeal properties of the gestures in Spanish and French may be involved in accounting for minor differences in overlap between etymologically related languages. For example, in Spanish, the nonlateral sonorant occupying the C2 position in our clusters is an alveolar tap, whereas it is uvular in French. This difference, along with distinctions in the durations, constriction degree and location of the involved consonants, may be partially implicated in deriving the minor differences in overlap between French and Spanish. Secondly, minor differences in overlap patterns between French and Spanish clusters may be expected if the aerodynamic requirements to produce the voicing contrast were different in each language, which is conceivable given the stop/approximant opposition that exists for Spanish (voiced stops in Spanish can undergo fricativization, i.e., /b, ɡ/ → [β, ɣ]), but not for French voiced stops.

Recall that, as mentioned in section 3.1, one of our objectives was to assess how robust timing patterns between consonants in clusters are across different vowels. Specifically, we used vowels of two different height categories (low and mid) as well as two different frontness categories (front and back). Our results indicate no significant difference in overlap in C1C2 as a function of these differences in vowel category. However, given how little is known about such effects, we mention two trends in the results which though not statistically significant hint at potential IPI differences in C1C2 clusters as a function of subsequent vowel height and/or horizontal contrasts. For both cluster sets (stop + lateral and stop + tap), IPI was shortest in clusters preceding /a/ (stop + lateral: mean = 34.2 ± 19.9 ms; stop + tap: mean = 60.2 ± 15.7 ms) and was longest in clusters preceding /o/ (stop + lateral: mean = 38.6 ± 23.1; stop + tap: mean = 65.4 ± 18.3), with clusters preceding /e/ being in the middle (stop + lateral: mean = 35.8 ± 21.2; stop + tap: mean = 61.3 ± 17.9). Thus, for future studies it may be fruitful to address the potential influences of the following vowel on intergestural timing in clusters more fully by including all five peripheral vowels, especially in relation to the effects of voicing on C1C2 clusters, since vowels show different aerodynamic properties as a function of height (Demolin, Gerbers & Haddid, 2008) and have been shown to vary in their coarticulation with surrounding segments (Recasens & Espinosa, 2009).

In agreement with predictions made in H4, the results for pre-plateau voicing (the name adopted in our study to refer to the intrusive vocoid) show that pre-plateau voicing is more common preceding taps (stop + /ɾ/ clusters) than laterals (stop + /l/ clusters), though no difference in the duration of pre-plateau voicing was found as a function of C2. For stop + lateral clusters there does seem to be a compensation effect whereby the duration of pre-plateau voicing is conditioned by the duration of the C2, such that as C2 reduces, pre-plateau voicing lengthens. This compensation effect was not present in stop + tap clusters. Conversely, the amount of gestural overlap has no effect on the duration of pre-plateau voicing in stop + lateral clusters, while there is a positive correlation between IPI and the duration of pre-plateau voicing in stop + tap clusters.

6 Conclusion

Temporal properties in the consonant + lateral and consonant + tap clusters /bl, gl, kl, pl, kɾ, pɾ/ of Spanish were studied using EMA and acoustics. Our results establish continuity with previous mainly acoustic results on Spanish and offer an articulatory basis on which these results, as well as new patterns observed in our data, can be characterized and rationalized.

We found that, in accordance with past work, VOT of the initial stop was modulated by its place of articulation, and the duration of the stop was strongly affected by both its voice and place specification. Additionally, lateral duration was found to be significantly shorter in clusters than in word initial singleton onsets. Among our new findings, we highlight two, the effect of voicing on overlap clusters and, in considering prospects for future work, the presence of a positive IPI between the two consonantal constrictions in the clusters.

Specifically, one of the main findings of our study is that voicing modulates the intergestural timing in Spanish onset clusters. Overlap in clusters in which the two consonants share the same voice specification (voiced) exhibit more overlap than in clusters in which C1 is voiceless and C2 is voiced. This is significant in light of the results for French in Bombien and Hoole (2013), where voicing of C1 did not affect patterns of overlap. It thus follows that even etymologically related languages and languages that share, on the surface, similar voicing implementations, can exhibit language-specific differences in intergestural timing.

A second main finding is that Spanish clusters are characterized by a positive IPI, that is, a lag between the release of the first consonant and the achievement of target of the second consonant. This finding is of interest in the context of studies on the question of how prosody and more specifically syllables are expressed in terms of intersegmental timing patterns. In a landmark study on this topic, Browman and Goldstein (1988, 2000) argued that English syllables are characterized by certain syllable-position-specific timing patterns. Specifically, for onset clusters they found evidence that the cluster as a whole enters into a global timing relation with its tautosyllabic vowel. In a theoretical interpretation of this pattern, Browman and Goldstein (2000) proposed that each of the consonants in the onset is required to be in an in-phase relation to the vowel. However, the consonants are also required to be in a sequential or antiphase relation to each other. These requirements are in conflict and thus a compromise is found whereby the consonants satisfy the in-phase relation to the vowel to the extent possible while still being perceptually recoverable (see Gafos, 2002, for an optimality theoretic analysis). The resulting configuration is one where a time point at the center of the cluster (known as the c-center) enters into a stable relation with the vowel. In subsequent investigations of similar patterns in different languages, a generalization has emerged where languages with low overlap in C1C2 clusters tend not to show the same timing arrangement as English; for Moroccan Arabic and Tashlhiyt Berber, see Shaw, Gafos, Hoole & Zeroual (2009) and Hermes, Ridouane, Mücke, and Grice (2011), respectively. In these languages, it appears that the vowel following a consonant cluster in C1C2V enters into a local relation with the immediately prevocalic consonant C2 to the exclusion of the C1 in the C1C2 cluster. A possible explanation for this result may be that the large IPIs characteristic of C1C2 timing in these languages preclude satisfaction of the in-phase relation between C1-V, which in turn results in maximal satisfaction of the C2-V in-phase relation (giving rise to the observed local temporal stability patterns reported in these studies).

Our Spanish results indicate the presence of low overlap or a substantial IPI in C1C2. However, Spanish is a language which like English is assumed to admit clusters as onsets in syllables, in contrast to Moroccan Arabic which is claimed not to admit clusters as onsets. Thus, the results obtained in the present study indicate that Spanish clusters exhibit C1C2 timing patterns of the sort that characterize languages which do not admit clusters as syllable onsets (in that there is a robust interplateau lag between the two consonants in the cluster), but the language is prototypically of the complex onset type. In this way, our results are similar to Romanian (Marin & Pouplier, 2014) in that Romanian onset clusters were generally found to have a temporal organization like that found for English and other languages that permit complex onsets; yet, overlap between the consonants in clusters was constrained by the C2 (rhotic trill in the case of Romanian). In this study, stop + trill clusters were shown to have lower overlap than stop + lateral clusters. This difference in overlap is similar to the pattern found in our study for Spanish. The overlap results presented here for Spanish are also consistent with results reported on Polish (Pastätter & Pouplier, 2015; Hermes, Mücke & Auris, 2017), which shows low overlap in clusters (especially in stop + rhotic clusters) and is generally regarded as allowing complex onsets. Further analyses should thus focus on addressing the relationship between overlap and the global timing of the onset cluster with the following vowel, and how the global relationship between the cluster and vowel is modulated by changes in overlap as a result of articulatory constraints on the individual gestures.

Acknowledgments

A.I.G. gratefully acknowledges support from DFG SFB 1287 and ERC AdG 249440. M.G. gratefully acknowledges support from PIUNA 11338701.

Statement of Ethics

All subjects gave their informed consent to take part in the study and were paid for their participation.

Disclosure Statement

The authors declare no conflicts of interest.

Appendix A

Consonant	Word	English	log frequency
/p/	*pato*	duck	1.188991
	plato	plate	1.290773
	prato	town in Italy	0.000397
	pomo	knob	0.310954
	plomo	lead	1.128012
	promo	from *promoción*, promotion	0.002856
	pena	pity	1.882075
	plena	full	1.652334
	presa	prisoner	1.488705
/k/	*cata*	tasting, as in "wine tasting"	0.417983
	clapas	stretch of barren land	0.000591
	crapa	proper name of restaurant in Valencia, Spain, and a city in Haiti	0.000046
	como	as or like, 1st pers. singular of verb *comer*, to eat	3.718704
	clono	neurological alteration	0.000257
	cromo	sports cards	0.419597
	quema	3rd pers. sing. of *quemar*, to burn	0.916907
	clema	piece used for electrical connection	0.001942
	crema	cream	0.844365
/b/	*banda*	band	2.256765
	blanda	soft	0.747241
	bata	robe	0.684890
	blata	brand name of motorcycle	0.000094
	beque	urinal	0.006999
	bleque	excess of activity or force	0.000483
	bota	boot	0.646512
	bloque	block, as in apartment block	1.445043
/g/	*gana*	3rd pers. sing. of *ganar*, to win	1.592388
	glana	proper surname from Galicia, Spain	0.001932
	gato	cat	1.597231
	glato	town in Italy	0.000764
	gueto	ghetto	0.442094
	gleba	piece of land that can be cultivated	0.137877
	gobo	perforated metallic disc that is resistant to high temperatures	0.000062
	globo	balloon	1.253927
/l/	*lapa*	of Lap (from Lapland) origin, feminine	0.355484
	lato	dilated or stretched	0.207053
	leco	international company with franchises throughout Spain, scream (colloquial from Venezuela)	0.000091
	lema	motto	1.266025
	lena	blabbermouth	0.531207
	loco	crazy	1.527539
	lomo	loin	0.820499
	loto	lotus (plant) and short form of *lotería*, lottery, as in *bono-loto* (lottery bonus)	0.719074

References

Abramson, A. S., & Whalen, D. H. (2017). Voice Onset Time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. *Journal of Phonetics, 63*, 75–86. https://doi.org/10.1016/j.wocn.2017.05.002

Avram, A. (1993). Cercetări experimentale asupra consoanelor lichide din limba română [Experimental research on liquid consonants in Romanian]. *Fonetică şi dialectologie, 12*, 8–20.

Baese-Berk, M., & Goldrick, M. (2009). Mechanisms of interaction in speech production. *Language and Cognitive Processes, 24*(4), 527–554. https://doi.org/10.1080/01690960802299378

Baltazani, M. (2009). Acoustic characterization of the Greek rhotic in clusters. *Proceedings of the 18th International Symposium on Theoretical and Applied Linguistics* (Vol. 1, pp. 87–95). Thessaloniki, Greece: Monochromia.

Baltazani, M., & Nicolaidis, K. (2011). Production of the Greek rhotic in initial and intervocalic position: An acoustic and electropalatographic study. In: Z. Gavriilidou, A. Efthymiou, E. Thomadaki, & P. Kambakis-Vougiouklis (Eds.), *Selected papers of the 10th International Conference on Greek Linguistics* (pp. 141–152). Komotini, Greece: Democritus University of Thrace.

Baltazani, M., & Nicolaidis, K. (2013). The many faces of /r. In L. Spreafico & A. Vietti (Eds.), *Rhotics: New data and perspectives* (pp. 125–144). Bozen: University of Bozen-Bolzano.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Blecua, B. (2000). Compensación temporal en los elementos del ataque silábico. *Proceedings from the II Congress of Experimental Phonetics*, March 5–7, Sevilla, Spain (pp. 101–108).

Blecua, B. (2001). *Las vibrantes del español: Manifestaciones acústicas y procesos fonéticos* (PhD dissertation), Universitat Autònoma de Barcelona, Spain. Retrieved from http://hdl.handle.net/10803/4859

Blecua B, Boadas R, Bofill M, Rost A (2003): La consonante lateral en los grupos consonánticos prenucleares del español. *Proceedings of the XXXIII Simposio de la Sociedad Española de Lingüística*. December 16-19, Girona, Spain.

Blecua, B., & Machuca, M. (2000). Los grupos consonánticos en español: Manifestaciones acústicas en dos estilos de habla. *Proceedings of the IV Congreso de Lingüística General*, April 3–6, Cádiz, Spain (pp. 283–297).

Bombien, L., & Hoole, P. (2013). Articulatory overlap as a function of voicing in French and German consonant clusters. *The Journal of the Acoustical Society of America, 134*(1), 539–550. https://doi.org/10.1121/1.4807510

Bradley, T. (2006). Spanish complex onsets and the phonetics-phonology interface. In F. Martinez Gil & S. Colina (Eds.), *Optimality-Theoretic Studies in Spanish Phonology* (pp. 15–38). Amsterdam: John Benjamins.

Browman, C. P., & Goldstein, L. (1988). Some notes on syllable structure in articulatory phonology. *Phonetica, 45*(2-4), 140–155. https://doi.org/10.1159/000261823

Browman, C., & Goldstein, L. (2000). Competing constraints on inter-gestural coordination and self-organization of phonological structures. *Bulletin de la Communication Parlée, 5*, 25–34.

Bybee, J. (1999). Usage-based phonology. In: M. Darnell, E. Moravcsik, F. Newmeyer, M. Noonan, & K. Wheatley (Eds.), *Functionalism and formalism in linguistics* (Vol. I, pp. 211–242). Amsterdam, the Netherlands: Benjamins.

Bybee, J. (2000). Lexicalization of sound change and alternating environments. In: M. Broe & J. Pierrehumbert (Eds.), *Laboratory phonology V: Acquisition and the lexicon* (pp. 250–268). Cambridge, UK: Cambridge University Press.

Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change, 14*(3), 261–290. https://doi.org/10.1017/S0954394502143018

Byrd, D. (1996). Influences on articulatory timing in consonant sequences. *Journal of Phonetics, 24*(2), 209–244. https://doi.org/10.1006/jpho.1996.0012

Caramazza, A., & Yeni-Komshian, G. H. (1979). Voice onset time in two French dialects. *Journal of Phonetics, 2*, 239–245.

Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics, 27*(2), 207–229. https://doi.org/10.1006/jpho.1999.0094

Colantoni, L., & Steele, J. (2005). Phonetically-driven epenthesis asymmetries in French and Spanish obstruent-liquid clusters. In R. Gess & E. Rubin (Eds.), *Experimental and theoretical approaches to Romance linguistics* (pp. 77–96). Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.272.06col

Colantoni, L., & Steele, J. (2011). Synchronic evidence of a diachronic change: Voicing and duration in French and Spanish stop-liquid clusters. *Canadian Journal of Linguistics/Revue canadienne de linguistique, 56*(2), 147–177.

Crystal, T. H., & House, A. S. (1988). Segmental durations in connected-speech signals: Current results. *The Journal of the Acoustical Society of America, 83*(4), 1553–1573. https://doi.org/10.1121/1.395911
Demolin, D., Gerbers, S., & Haddid, S. (2014). Vowel and consonant effects on subglottal pressure. *The Journal of the Acoustical Society of America, 135*(4), 2294–2295. https://doi.org/10.1121/1.4877542
Docherty, G. J. (1992). *The Timing of Voicing in British English Obstruents*. Berlin, New York: Walter de Gruyter. https://doi.org/10.1515/9783110872637
Eilers, R. E., Oller, D. K., & Benito-Garcia, C. R. (1984). The acquisition of voicing contrasts in Spanish and English learning infants and children: A longitudinal study. *Journal of Child Language, 11*(2), 313–336. https://doi.org/10.1017/S0305000900005791
Gafos, A. (2002). A grammar of gestural coordination. *Natural Language and Linguistic Theory, 20*(2), 269–337.
Gafos, A., & Goldstein, L. (2012). Articulatory representation and phonological organization. In: A. C. Cohn, C. Fougeron, & M. K. Huffman (Eds.), *The Oxford handbook of laboratory phonology*. Oxford, UK: Oxford University Press.
Gafos, A., Hoole, P., Roon, K., & Zeroual, C. (2010). Variation in timing and phonological grammar in Moroccan Arabic clusters. In: C. Fougeron (Ed.), *Laboratory phonology 10: Variation, detail and representation*. Berlin, Germany: Mouton de Gruyter.
Gibson M, Fernández Planas, A M, Gafos A, Remirez E (2015): Consonant duration and VOT as a function of syllable complexity and voicing in a subset of Spanish clusters (pp. 1690–1994). *Proceedings of the International Speech Communication Association (INTERSPEECH)*, Dresden, Germany.
Gili Gaya, S. (1921). La r simple en la pronunciación española. *Revista de Filología Española, 8*, 271–280.
Goldrick, M., Vaughn, C., & Murphy, A. (2013). The effects of lexical neighbors on stop consonant articulation. *The Journal of the Acoustical Society of America, 134*(2), EL172–EL177. https://doi.org/10.1121/1.4812821
Gudurić, S., & Petrović, D. (2005). О природи гласа /r/ у српском језику [The nature of the sound [r] in the Serbian language]. *Зборник Матице српске за филологију и лингвистику 48*(1–2): 135–150.
Haggard, M. (1973). Correlations between successive segment durations: Values in clusters. *Journal of Phonetics, 1*, 111–116.
Hall, N. (2006). Cross-linguistic patterns of vowel intrusion. *Phonology, 23*(3), 387–429. https://doi.org/10.1017/S0952675706000996
Hermes, A., Mücke, D., & Auris, B. (2017). The variability of syllable patterns in Tashlhiyt Berber and Polish. *Journal of Phonetics, 64*, 127–144. https://doi.org/10.1016/j.wocn.2017.05.004
Hermes, A., Ridouane, R., Mücke, D., & Grice, M. (2011). Kinematics of syllable structure in Tashlhiyt Berber: The case of vocalic and consonantal nuclei. *9th International Seminar on Speech Production, June 20–23, Montréal, Canada*.
Hoole, P. (2006). *Experimental studies of laryngeal articulation. II. Laryngeal-oral coordination in consonant sequences* (Habilitation thesis), Ludwig–Maximilians–Universität, Munich, Germany. http://www.phonetik.uni-muenchen.de/
Hoole, P., Bombien, L., Kühnert, B., & Mooshammer, C. (2009). Intrinsic and prosodic effects on articulatory coordination in initial consonant clusters. In G. Fant, H. Fujisaki, & J. Shen (Eds.), *Frontiers in Phonetics and Speech Science* (pp. 275–286). Beijing: The Commercial Press.
Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal/Biometrische Zeitschrift, 50*(3), 346–363. https://doi.org/10.1002/bimj.200810425
Klatt, D. H. (1975). Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research, 18*(4), 686–706. https://doi.org/10.1044/jshr.1804.686
Lenz, R. (1892). Chilenische Studien. *Phonetische Studien, 5*, 272–293.
Lisker, L., & Abramson, A. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word, 20*(3), 384–422. https://doi.org/10.1080/00437956.1964.11659830
Maddieson, I. (1997). Phonetic universals. In: W. J. Hardcastle, F. E. Gibbon, & J. Laver (Eds.), *The Handbook of Phonetic Sciences* (pp. 619–639). Oxford, Cambridge: Wiley-Blackwell.
Maddieson, I. (2009). *Patterns of Sounds*. Cambridge: Cambridge University Press.
Malmberg, B. (1965). *Estudios de fonética hispánica (Collectanea Phonetica, I.)*. Madrid: Consejo Superior de Investigaciones Científicas.
Marin, S., & Bučar Shigemori, L. S. (2014). *Vowel compensatory shortening in Romanian* [online]. Retrieved from http://phonetik.unimuenchen.de/universals/pub/ISSP2014_VCS_Marin_BucarShigemori.pdf
Marin, S., & Pouplier, M. (2014). Articulatory synergies in the temporal organization of liquid clusters in Romanian. *Journal of Phonetics, 42*, 24–36. https://doi.org/10.1016/j.wocn.2013.11.001

Martinez Celdran, E. (1993). Un mismo parámetro fonético en el fondo de la lenición protorromance: la duración. In J. A. Bartol Hernández, J. F. García Santos, & J. De Santiago Guervós (Eds.), *Estudios Filológicos en Homenaje a Eugenio de Bustos Tovar. Volumen 2* (pp. 621–640). Salamanca, Spain: Ediciones Universidad de Salamanca.
Martinez Celdran, E. (1996). *El sonido en la comunición humana: introducción a la fonética*. Barcelona: Octaedro.
Martinez Celdran, E. (2001). Cuestiones problemàticas en los fonemas sonantes del español. *Lingüística española actual, 23*(2), 159–171.
Menéndez Pidal, R. (1926). *Orígenes del español*. Madrid: Espasa Calpe.
Navarro Tomás, T. (1918). *Manual de pronunciación española*. Madrid: Centro de Estudios Históricos.
Nearey, T., & Rochet, B. (1994). Effects of place of articulation and vowel context on VOT production and perception for French and English stops. *Journal of the International Phonetic Association, 24*(1), 1–18. https://doi.org/10.1017/S0025100300004965
Ohala, M., & Ohala, J. J. (1973). The problem of aspiration in Hindi phonetics. In: Central Hindi Directorate (Ed.), Bhashavigyan Ank Hindi (pp. 67–72). Delhi, India: Central Hindi Directorate, Ministry of Education and Social Welfare, Government of India. Hindi.
O'Shaughnessy, D. (1974). Consonant durations in clusters. *IEEE Transactions on Acoustics, Speech, and Signal Processing, 22*(4), 282–295. https://doi.org/10.1109/TASSP.1974.1162588
Pastätter M, Pouplier M (2015): Onset–vowel timing as a function of coarticulation resistance: Evidence from articulatory data. *Proceedings of the 18th International Congress of Phonetic Sciences*, August 10-14, Glasgow, UK.
Pavlík, R. (2008). K niektorým otázkam kvalitatívnych a kvantitatívnych vlastností slovenských vibrant [To the question of qualitative and quantitative characteristics of the Slovak vibrant consonants]. *Jazykovedný časopis, 52*(1–2), 65–97.
Pierrehumbert, J. (2002). Word-specific phonetics. In: C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology VII* (pp. 101–140). Berlin, Germany: Mouton de Gruyter. https://doi.org/10.1515/9783110197105.101
Quilis, A. (1970). El elemento esvarabático en los grupos [pr, br, tr]. *Phonétique et Linguistique Romaines: Mélanges offerts à M. Georges Straka* (pp. 99–104). Lyon-Strasbourg, France: Société de Linguistique Romane.
Quilis, A. (1988). *Fonética acústica de la lengua española*. Madrid: Editorial Gredos.
Quilis, A. (1993). *Tratado de fonética y fonología españolas*. Madrid: Editorial Gredos.
Quilis, A., Esgueva, M., Gutiérrez Araus, M. L., & Cantarero, M. (1979). Características acústicas de las consonantes laterales españolas. *Lingüística española actual, 1*(2), 233–344.
Recasens, D., & Espinosa, A. (2009). An articulatory investigation of lingual coarticulatory resistance and aggressiveness for consonants and vowels in Catalan. *The Journal of the Acoustical Society of America, 125*(4), 2288–2298. https://doi.org/10.1121/1.3089222
Rosner, B., López-Bascuas, L. E., García-Albea, J. E., & Fahey, R. P. (2000). Voice onset times for Castilian Spanish initial stops. *Journal of Phonetics, 28*(2), 217–224. https://doi.org/10.1006/jpho.2000.0113
Savu, C. F. (2012). *On the phonetic structure of the rhotic tap and its phonological implications* (Unpublished Master's thesis), University of Bucharest, Sofia, Romania.
Shaw, J., Gafos, A., Hoole, P., & Zeroual, C. (2009). Syllabification in Moroccan Arabic: Evidence from patterns of temporal stability. *Phonology, 26*(1), 187–215. https://doi.org/10.1017/S0952675709001754
Sotiropoulou, S., Gibson, M., Tobin, S., & Gafos, A. (2018). Temporal stability patterns of stop-liquid and stop-rhotic clusters in Spanish. *LabPhon, 16*(June), 19–22.
Stevens, K. N. (2000). *Acoustic Phonetics*. Cambridge, MA: MIT.
Stolarski, Ł (2011). *Vocalic elements in the articulation of the Polish and English /r/*. Paper presented at Languages in Contact, University of Wrocław, Poland.
Vago R, Gósy M (2007): Schwa vocalization in the realization of /r/. *Proceedings of the 16th International Congress of Phonetic Sciences*: 505–509.
Villamizar, T. (2002). Caracterización acústica de las consonantes oclusivas en el español de Venezuela. El parámetro duración (silencio – barra de explosión – VOT). *Lengua y Habla, 7*(1), 140–157.
Weismer, G. (1980). Control of the voicing distinction for intervocalic stops and frictives: Some data and theoretical considerations. *Journal of Phonetics, 8*, 427–438.
Williams, L. (1977). The voicing contrast in Spanish. *Journal of Phonetics, 5*, 169–184.
Ying J, Carignan C, Shaw J A, Proctor M I, Derrick D, Best C T (2017): Temporal Dynamics of Lateral Channel Formation in /l: 3D EMA Data from Australian English. *Proceedings of the International Speech Communication Association (INTERSPEECH)*, Stokholm, Sweden: 2978-2982. https://doi.org/10.21437/Interspeech.2017-765