Consonant duration and VOT as a function of syllable complexity and voicing in a sub-set of Spanish clusters

Mark Gibson 1, Ana María Fernández Planas 2, Adamantios Gafos 3, Emily Remirez 1
1 Rice University
2 Universitat de Barcelona
3 Universität Potsdam

mark.a.gibson@rice.edu

Abstract

A series of hypotheses are addressed vis-à-vis the effects of syllable complexity and voicing on consonant duration and VOT (voice onset time) in a subset of Spanish clusters. Electroglottographic (EPG) and acoustic signals were obtained for native-speakers of Standard Peninsular Spanish producing clusters within and across word boundaries. The results show that VOT patterns for singleton onsets also hold for clusters, VOT in word-final devoiced stops and voiceless stops exhibit a short lag/long lag paradigm similar to German and English onset clusters, C1 duration in clusters is a function of voicing and place but not of syllable complexity, and the duration of /l/ in complex onsets is conditioned by the phonological specifications of the preceding consonant, but linguo-palatal contact is not. Combined results suggest an intimate relationship between the temporal parameters of gestures in onset clusters, but not across word boundaries, and that the temporal and spatial parameters of gestures in onsets, even though related, operate relatively independently from one another.

Index Terms: VOT duration, Spanish clusters, temporal organization, consonant duration.

1. Introduction

Recent studies in a number of languages have shown that syllable complexity, constriction degree and constriction location, independently, affect the duration of consonants and patterns of VOT. In complex onsets, the duration of initial stops (C1) has been found to be a function of complexity, duration tending to reduce as syllable complexity increases [1-3]. Vowel-adjacent consonants in clusters show more reduction than word-initial consonants [1, 2], while liquids shorten more than stops [4, 5]. In simplex onsets, voiceless stops tend to be longer than voiced consonants [6], a pattern which has been robustly confirmed for Spanish [7-9]. Labials have been reported to be longer than coronals and velars in some languages [3, 10-13], though a recent study dealing with Spanish (Venezuelan) stops showed that duration of velars in simplex onsets was significantly longer than the duration of labials [9], a claim which is expected to hold for Peninsular Spanish as well. To date, however, no study has addressed the interaction of syllable complexity, voicing and place of articulation on the duration of consonants in Spanish clusters with C2 laterals. A propos prosodic boundaries, an increase in relative consonant duration has been found to be associated with the proximity of the consonant to initial domains [14]. EPG data show that domain-initial consonants are produced with more linguo-palatal contact than domain-internal consonants [15-22]. Further it was shown that this effect is gradient and cumulative in relation to the prosodic hierarchy. None of these studies, however, examine the interaction between prosodic/syllabic structure and the phonological specifications of the consonants in clusters.

As for VOT, two typological patterns have been found for bimodal voicing contrasts [23]. Languages such as French and Spanish contrast voiced and voiceless stops by way of a true voice/short lag paradigm [24-28], whereas many Germanic languages maintain the phonological contrast by way of a short-lag-long-lag pattern [29, 30, 3]. For voiceless stops in French, VOT patterns for singleton onsets were found to hold for onset clusters [3]. However, as yet, no study has addressed VOT duration in Spanish clusters, either within or across word boundaries.

Constriction location has long been reported to affect the temporal parameters of VOT, velar consonants having longer VOT’s than labials and coronals [24, 5, 31-36]. The biomechanical basis underlying this effect has been extensively discussed in [36, 25, 37, 3] and corroborated by studies examining the pre-speech babble of infants in a number of languages [38, 39]. For Spanish stops, [27] and later [28] found significantly shorter lag times for labials and apicals than for velars, though no data addressing VOT patterns in clusters has been reported.

In sum, whereas much is known about the duration and VOT of individual consonants in Spanish and other languages, the present study addresses these parameters in the context of clusters, both within and across word boundaries.

2. Experiments

2.1. Corpus

The corpus contained tokens with both simplex and complex onsets for voiced and voiceless labials and velars: /k, g, p, b/. Complex onsets consisted of uniform- and mixed-voice clusters with /U/: /kl/, /gl/, /pl/, /bl/. These same clusters were also examined across word boundaries: /k#l/, /g#l/, /p#l/, /b#l/. Target tokens were embedded in a carrier phrase, Diga ___, por favor, ‘Say ___, please’. Five repetitions were obtained per speaker (14 speakers x 5 repetitions x 136 tokens = 9520 total tokens, 1400 target tokens) for the acoustic analysis, while three repetitions per speaker were solicited for the EPG analysis (2 speakers x 3 repetitions x 136 tokens = 816 total tokens, 205 target tokens).

2.2. Hypotheses

In line with previous research, no effect on the duration of VOT for voiceless stops is expected for syllable complexity. Place of
articulation is expected to influence VOT duration but only for voiceless stops, velars exhibiting longer VOTs than labials. Since phonologically voiced consonants always emerge with negative VOT in Spanish onsets, voicing is expected to only be relevant to VOT duration in devoiced codas. In this case, two scenarios may emerge. First, it is possible that the voiced and voiceless consonants become neutralized. Conversely, it is conceivable that the underlying contrast between voiced and voiceless stops is maintained by way of disparate VOT patterns. In this case, it might be expected that the voiced and voiceless stops exhibit a paradigm similar to the short-lag/long-lag pattern attested in previous studies for English and German onsets, though any prediction to such effect will be withheld for now.

As regard the duration of the word-initial consonant (C1), syllable complexity is not expected to affect the length of the consonant, although differences based on place and voice are anticipated. Duration is expected to be greater for velars than for labials and voiceless consonants are expected to be longer than voiced stops.

As for the duration of the liquid (C2), a main effect of prosodic boundary is anticipated, duration being greater in word-initial position than in onset clusters, though the constriction location and degree of the preceding consonant is not expected to show any effect on the duration of the liquid at either boundary.

Finally, the degree of linguo-palatal contact is not expected to be influenced by the phonological specifications of the preceding consonant, but differences are expected based on prosodic boundary; word-initial liquids are expected to entail more linguo-palatal contact than /l/ in onset clusters.

2.3. Methodology

Acoustic signals were obtained for fourteen native speakers of Peninsular Spanish. Recordings were performed in a double attenuated sound proof chamber in the Phonetics Laboratory at the National Council for Scientific Research (CSIC) in Madrid, Spain and the Universitat de Barcelona in Barcelona, Spain. Sound files were annotated by hand using Praat by two Spanish-speaking phoneticians. Time stamps were placed at (1) the onset of C1, (2) the initial burst of VOT, (3) the onset of C2 in the case of complex onsets, (4) the onset of phonation, and (5) end of phonation.

EPG data were collected for two of the subjects at the Universitat de Barcelona using the WinEPG system, which records linguopalatal contact with an electropalate equipped with 62 electrodes. The present study focused on the point of maximum contact (PMC) of the alveolar lateral in order to establish its linguopalatal configuration, which has been studied from three points of view [40]: a) globally, using an index Q – which was obtained by dividing the number of contacted electrodes by the total amount (62) of electrodes, b) using a CAA index which measures the advance of linguopalatal contact in the first four rows of the artificial palate, covering the area between the upper teeth and gums and the alveolar ridge, and (c) using a CPA index that registers a lengthwise measure of the subsequent contacts in the same four rows.

2.4. Results

Linear mixed effects models were employed to analyze the data using the “lme4” package in R [41]. The variables under study included (1) VOT duration, (2) C1 duration, (3) C2 duration, (4) total linguo-palatal contact, and (5) contact in the alveolar region (CA and/or CPA). The predictors were (1) SYLLABLE COMPLEXITY (simplex onsets, complex onsets), (2) PROSODIC/WORD BOUNDARY (word-initial, word-final), (3) VOICE (voiced, voiceless), and (4) PLACE (labial, coronal, dorsal). To account for differences in speech rate and individual variation, repetition was nested within speaker and programmed as random effects. Maximum Likelihood Chi-squared tests based on the deviance statistics were performed in order to determine significance. Post-hoc univariate ANOVA were used to address contrasts at individual levels.

2.4.1. VOT duration

As with French, there is no contrast for positive VOT duration between word-initial voiced and voiceless stops due to the fact that phonologically specified voiced consonants in Spanish systematically emerge with negative VOT. Nevertheless, a main effect was found for voice on the duration of VOT (χ²[1, N=1,400] = 1399, p < 0.001) which was related to the word boundary. Of interest is the word-final context in which 16% of all phonologically voiced consonants emerge devoiced, which produced a three way interaction between voicing, place and word boundary (see Figure 1b). In word-final position, VOT for voiced stops was approximately 7 ms shorter than for voiceless stops (+voice = 18.6 ms, -voice = 26.1, F[1,00] = 38.03, p < 0.001 ), voiceless velars having a longer VOT than voiced labials by approximately 7 ms ([b] = 15.5 ms, [ɡ] = 22.3 ms, F[1,00] = 11.49, p < 0.001 ). Thus although word-initially the voicing contrast between voiced and voiceless stops is characterized by a true-voice/short-lag paradigm, word-finally this pattern shifts to a short-lag/long-lag paradigm ([b] = 15 ms, [p] = 22ms; [ɡ] = 22ms, [k] = 29 ms) similar to the voicing contrast in English and German.

![Figure 1: (a) VOT for simplex and complex onsets. (b) VOT for clusters across word boundaries.](image-url)

There was no main effect of structural complexity on VOT duration (χ²[2, N=1,400] = 1399, p = 0.074). In complex onsets (see Figure 1a), VOT for voiceless velars and labials showed some variation when compared to [CV] syllables (approximately 3 ms for labials and 0.5 ms for velars), though this difference was only significant for labials (labial, F[1,00] = 21.41, p < 0.001; velar, F[1,00] = 0.30, p = 0.53). For voiceless stops, contrasts in VOT duration based on prosodic boundary were significant overall (F[1,00] = 16.79.30, p < 0.001). For voiceless labials, VOT in word-final position (23 ms) was approximately 6 ms longer than VOT at the word-initial boundary (17 ms) (F[1,00] = 65.48, p < 0.001). This trend, however, did not hold for voiceless velars as VOT durations in word-initial and word-final positions were nearly identical (word-initial = 29.6 ms, word-final = 29.3 ms; F[1,00] = 0.18, p = 0.67).
Place of articulation was the most important predictor of VOT duration in the sense that significant differences were found between labial and velar gestures across all other factors ($\chi^2[1, N=1,400] = 1399, p < 0.001$). Pooled means show a difference of roughly 10 ms between VOT duration in velar and labial gestures, velars being longer in all contexts (velars = 29 ms, labials = 19 ms; $F[1,00] = 415.69, p<0.001$). In word-initial simplex onsets, VOT for voiceless velar stops was nearly twice as long (49%) as for voiceless labials (velar = 29.8 ms, labial = 15.3 ms; $F[1,00] = 381.43, p<0.001$).

As for the interaction of the predictors on VOT duration, significant effects were found for place of articulation and syllable complexity ($\chi^2[1, N=1,400] = 1399, p < 0.001$). Unsurprisingly, no interaction effect was found for C1 voice and syllable complexity, but this is due to the fact that the phonologically specified voiceless stops were realized with true voice where no VOT was detected in any word-initial contexts. In word-final context, however, a significant effect on the duration of VOT was found for voicing ($F[1,00] = 38.03, p<0.001$). In a similar way, no interaction effect was found between place and voice ($\chi^2[1, N=1,400] = 1399, p < 0.01$), again which is due to the true voice/short lag pattern for voiced versus voiceless consonants in Spanish.

2.4.2. C1 duration

Overall, C1 duration was more systematic than the duration of VOT. Pooled means show that velars are longer than labials by approximately 2.5 ms (velars = 79.2 ms, labials = 76.5 ms; $F[1,00] = 5.43, p=0.02$), rendering no significant main effect for place ($\chi^2[1, N=1,400] = 1399, p = 0.1$). There was, however, a main effect for voice ($\chi^2[1, N=1,400] = 1399, p < 0.001$) on C1 duration (see Figure 2). Pooled means of voiceless versus voiced stops show a difference of 20 ms (voiced = 69 ms, voiceless = 89 ms; $F[1,00] = 351.79, p<0.001$). In line with French and German, complexity has little effect on C1 duration ($F[1,00] = 1.31, p = 0.25$) for both voiced and voiceless stops (see Figure 2). The difference in duration between CV and CCV is less than 2 ms, yielding no main effect of syllable complexity on C1 duration ($\chi^2[1, N=1,400] = 1399, p < 0.1$). Nevertheless, the word boundary was found to have a significant effect on C1 duration ($F[1,00] = 12.97, p<0.001$), word-final stops being 5 ms longer than word-initial (word-initial stops = 76 ms, word-final stops = 81 ms). This result is surprising but will be explained in detail below when the interaction between voicing and word boundary is addressed.

2.4.3. Liquid duration and degree of linguo-palatal contact

As expected, the duration of the liquid is a function of syllable complexity ($\chi^2[1, N=1,153] = 1152, p < 0.001$). Mean duration of the liquid in complex onsets (66 ms) was nearly 23 ms shorter than in word-initial position (89 ms) ($F[1,00] = 242.1, p<0.001$). The effect of place of articulation of the C1 on the liquid was also significant ($\chi^2[1, N=1,153] = 1152, p < 0.001$), a longer C2 being related with a labial C1 (68 ms) than with velar (64 ms) ($F[1,00] = 4.98, p=0.03$) in complex onsets (see Figure 3a). Across word boundaries (C#C), however, the contrast based on place was greater (see Figure 3b). When preceded by a labial (89 ms), the liquid was nearly 9 ms longer than when preceded by a velar (80.5 ms) ($F[1,00] = 20.3, p<0.001$).

Figure 3: (a) Liquid duration following labial and velar C1 in complex onsets. (b) Liquid duration in word-initial position preceded by labial and velar consonants.

A main effect was discovered for voice as well ($\chi^2[1, N=1,153] = 1152, p < 0.001$). In complex onsets, liquids were approximately 7 ms longer when preceded by a voiced C1 (69 ms) than when preceded by a voiceless stop (62 ms) ($F[1,00] = 30.03, p<0.001$). Across word boundaries, however, there was no effect of voicing on the duration of the liquid ($F[1,00] = 0.57, p>0.45$).

No interaction effect was found for place and voicing on the duration of C2, but significant context-specific effects were found. In complex onsets duration of the C2 was approximately 7 ms longer following a voiced labial (71 ms) than a voiceless labial (64 ms) ($F[1,00] = 14.63, p<0.001$). The
same effect was also observed for velar stops. The duration of
the liquid following a voiced consonant (68 ms) was nearly 8
ms longer than following a voiceless stop (60 ms) ($F[1,00] =
15.7, p<0.001$). Nevertheless, no effect on the duration of the
liquid was found as a consequence of voice and place
interaction across word boundaries (voiced/voiceless labials, $F$
$[1,00] = 1.33, p=0.25$; voiced/voiceless labials, $F[1,00] = 0.02,
p<0.88$).

Concerning the degree of linguo-palatal contact for
C2, the only main effect was found for syllable complexity
($\chi^2[8, N=205] = 197, p<0.001$) (see Figure 4). There is
significantly more contact for the lateral in word-onset ($/\ell/$/:
Q value $[M=41], CPA [M=85]$) than in clusters ($/C/$: $Q$
value $[M=35], CPA [M=82], F[1,00] = 8.99, p=0.003$).

![Figure 4. Linguo-palatal contact for liquid in word-initial
onsets and clusters (C2)](image)

In onset clusters, there is more total linguo-palatal
contact (Q) following velars (Q value $[M=39.5]$) than labials
(Q value $[M=33.8]$) ($F[1,00] = 20.1, p<0.001$). Comparing
contact in the anterior region (CAa/CPA) only, however, there is
no difference in alveolar contact based on the place of
articulation of the preceding gesture (CAa, $Q$ value: labial
$[M=83.9]$; $Q; F[1,00] = 1.96, p=0.31$/CA: $F[1,00] =
1.25, p=0.27$). No main effect of voicing was found for linguo-
palatal contact in onset clusters ($Q$ value $[M=36.5]$ preceding
voiced C1; $Q$ value $[M=34.5]$ preceding voiceless C1; $F[1,00]
= 2.40, p=0.12$) or across word boundaries ($Q$ value $[M=40.7]
preceding voiced C1; $Q$ value $[M=40.9]$ preceding voiceless
C1; $F[1,00] = 0.007, p=0.93$). Concerning place and voice
interaction in onsets, no significant effect was found (voiced
labials: $Q$ value $[M=35]$, voiceless labials: $Q$ value $[M=32.7]$
$F[1,00] = 1.6, p=0.20$; voiced velars: $Q$ value $[M=35.8],
voiceless velars: Q value $[M=37.1]; F[1,00] = 0.76, p=0.39$).
Across word boundaries total linguo-palatal contact for laterals
following voiced labials was marginally greater (n.s.s.) than
those following voiceless labials ($Q$ value $[M=39.8]$, voiceless
labials: $Q$ value $[M=39.4]; F[1,00] = 0.01, p=0.92$). For liquids following velars, there was slightly more
linguo-palatal contact following voiceless stops than following
voiced consonants (voiceless velars: $Q$ value $[M=42.4]$),
voiced velars: $Q$ value $[M=41.7]; F[1,00] = 0.06, p=0.81$).

The results show that the spatial parameters of the C2
are influenced by the structure of the target task but not by
the phonological specifications of the preceding consonant gesture.
As for the correlation between the duration of C2 and linguo-
palatal contact, a simple linear regression model shows a weak
relationship ($p=0.03$, see Figure 5) between the temporal
and spatial parameters of the lateral such that longer duration
generally entails a more laminal articulation (shorter duration is
correlated with a more apical articulation).

Concerning the duration of VOT, the results of this study
confirm that patterns of VOT for singleton onsets generally hold
for clusters. As in German and English onsets, phonologically
voiced consonants exhibit a shorter VOT than voiceless stops in
Spanish codas.

Duration of the C1 was found to be susceptible to
place of articulation, voicing and prosodic context, but not to
syllable complexity. In line with past studies examining the
temporal organization of gestures in complex onsets, the shift
from simplex to complex onsets does not have an impact per se
on the duration of the word-initial gesture.

Predictably, the most striking effect on the duration of
C2 was motivated by syllable complexity. A marked
reduction in the duration of the lateral was found for all tokens
in complex onsets, as compared to the duration of word-initial
liquid tokens. Interestingly, however, it was found that the
degree of reduction is conditioned by the phonological
specifications of the preceding consonant. For complex onsets,
liths following velars were systematically shorter than those
following labials (surely due to the fact that labial articulation
does not involve any lingual gesture), while liquids preceded by
voiceless consonants were significantly shorter than those
preceded by voiced stops.

One possibility to explain liquid behavior in mixed-
and uniform-voice clusters could be posited based on the
perceptual saliency of the individual gestures. Since voicing in
uniform-voice clusters spans across both gestures, the
probability that listeners will not perceive the liquid would be
greater due to lack of an acoustic cue signalling a change in
voice modality across gestures. Thus, maintaining a longer
duration for the liquid, coupled with a more laminal tongue
posture, may be desirable in order to avoid perceptual ambiguity between simplex and complex onsets.

3. Discussion and Conclusions

This relation may play some role in accounting for
the overlap patterns in uniform- versus mixed-voice clusters
outlined in previous studies (for a thorough review see [3]).

AIG acknowledges support by ERC AdG 249440.

4. Acknowledgements
5. References


