

Dynamical account of how /b, d, g/ differ from /p, t, k/ in Spanish: Evidence from labials

BENJAMIN PARRELL

Department of Linguistics, University of Southern California

Abstract

This study examines articulatory lenition of intervocalic stops in Spanish and tests the theories that 1) /b, d, g/ have an intended target for closure equal to that of /p, t, k/ and 2) spirantization of /b, d, g/ is caused by undershoot due to their short duration phrase medially. Consistent with past acoustic studies, subjects produce /b/ with incomplete closure phrase medially and complete closure phrase initially. Additionally, /b/ is shorter than /p/ phrase medially though not initially. For /b/, though not for /p/, there is a correlation between constriction degree and duration, consistent with the theory of dynamical undershoot. The results from the study are accurately modeled with a virtual target for /b/ slightly beyond the point of articulator contact. Such a target results in full closure at long durations (such as found phrase initially) and incomplete closure at shorter durations. Based on this evidence, it is proposed that /b, d, g/ differ from /p, t, k/ in three ways: they are shorter, lack a devoicing gesture, and have a target closer to – but still beyond – the point of articulator contact.

1. Introduction

1.1. Production of Spanish stops

All dialects of Spanish are characterized by a process of spirantization, wherein the voiced stops /b, d, g/ are produced with full occlusion only phrase initially, after a homorganic nasal, or, for /d/ only, after /l/. In all other positions they commonly are realized as the voiced approximants [β, ð, ɣ].

The first treatment of this alternation in the generative tradition was Harris (1969). That study analyzed these sounds as underlying stops that undergo a rule of lenition except after a homorganic, non-strident obstruent or a phrase boundary. This proposal has been widely adopted, in general terms, as the correct analysis. A number of later works in both linear and autosegmental phonology have adopted the same or similar analyses (Goldsmith 1981; Mascaró 1984; Hualde 1988). One

recent approach in Optimality Theory also specifies these sounds as voiced stops, and relies on markedness constraints and articulatory effort (after Kirchner [1998]) to drive spirantization (Piñeros 2002).

Another view is that these sounds are underlying approximants that undergo a process of fortition to derive the full stops. The first suggestion along these lines is found in Lozano (1979) who proposed that these sounds are unspecified for the feature [\pm continuant], with a rule that fills in [+continuant] where appropriate and a complementary rule that fills in [-continuant] elsewhere. González (2003) extends this interpretation to Optimality Theory, arguing for a constraint ranking that derives the correct output regardless of whether stops or spirants are specified in the input. Other proposals in Optimality Theory have gone farther, explicitly specifying the input for OT analysis as spirants (Baković 1994; Barlow 2003).

The assumption underlying most of these proposals is that there exists a clear alternation between the stop and approximant allophones. However, there is a large and growing body of evidence that the production of these segments is variable in all positions, and can be influenced by a large number of factors. A number of studies have used amplitude ratio – that is, the ratio of the amplitude minimum during a consonant to the maximum of the following vowel – as a measure of constriction degree. Stress seems to play a significant role in the realization of these segments, with longer and more constricted productions at the onset of a stressed syllable compared with an unstressed one (Cole et al. 1999; Ortega-Llebaria 2004; Eddington 2010). The preceding and following segments have also been shown to influence production, though the results are somewhat inconsistent across studies (Cole et al. 1999; Ortega-Llebaria 2004; Carrasco and Hualde 2009; Hualde et al. 2010). Speech rate also plays a role, with fast speech leading to more open productions (Soler and Romero 1999). There is also evidence that, for the dialect spoken near Toledo, Spain, /b, d, g/ can be realized as voiceless stops in syllable initial position while /p, t, k/ often surface as voiced (Torreblanca 1976). In general, these studies show productions that range from wide approximants to full stops both in contexts where, in the traditional description, we expect spirantization to occur and where we expect full occlusion.

In addition to the variability found in the production of the voiced stops, recent work has shown that the voiceless stops themselves are highly variable. Machuca (1997), in her examination of casual speech in Barcelona, finds that roughly 40% of productions of phonologically voiceless stops show some degree of voicing. Additionally, about 9% of the voiceless stops in her corpus were produced as approximants, the majority voiced. Other studies have found similar results, though there were large differences in the frequency with which the two processes occur between different dialects (Lewis 2001; Martínez Celdrán 2009; Hualde et al. 2011).

We are left with a situation where /b, d, g/ can be realized as voiced stops, voiced approximants, and perhaps even voiceless stops, while /p, t, k/ can be produced as voiceless stops, voiced stops, or even voiced approximants. The question

we must ask, then, is how Spanish speakers reliably distinguish these sounds. There is experimental evidence to support that they are, in fact, reliably distinguished in both production and perception (Romero et al. 2007). That study found that, even though there was some overlap in intensity ratio (used as a measure of constriction degree and voicing) between /p, t, k/ and /b, d, g/, they were nonetheless significantly different in production. Additionally, listeners correctly identified even the voiceless stops with the highest intensity ratio (most voiced/approximant production). Hualde et al. (2011) also found even voiced productions of /p, t, k/ to be produced with greater constriction than /b, d, g/.

The large variability in production also raises the undecided question of the phonological representation of /b, d, g/. Are these stops that spirantize, approximants that undergo fortition, or perhaps something else entirely?

1.2. *The nature of /b, d, g/ and how they are distinguished from /p, t, k/*

One possible difference between /b, d, g/ and /p, t, k/ is their duration. There is ample evidence that the voiceless series is generally longer than the voiced series (Herrera [1997]; Lavoie [2001]; and Recasens [1986] for Catalan, which shows the same pattern of spirantization). Additionally, there is evidence that listeners use duration to distinguish the voiced and voiceless stops. Reducing the period of silence associated with the closure of a voiceless stop causes it to be perceived as voiced (Martínez Celdrán 1991a, 1993). Machuca (1997) found that voiced, spirantized /p, t, k/ are still longer than phonological /b, d, g/ produced in the same manner. Based on this, Hualde (2005) posits that, in addition to a voicing distinction that may be neutralized in some contexts, duration still reflects the phonological contrast between /p, t, k/ and /b, d, g/.

There is some evidence, however, that /p, t, k/ can be realized with the same duration as /b, d, g/ (Martínez Celdrán 2008, 2009). Some authors take the fact that these two sets of sounds generally differ in voice, constriction degree, and duration, but in no category consistently, as evidence for a phonological distinction by the feature [\pm tense] (Martínez Celdrán 1991a, 1991b, 1993, 2008, 2009; Herrera 1997; Martínez Celdrán and Fernández Planas 2007). However, most of the evidence used in support of phonological tension comes from durational differences between voiced and voiceless stops. Martínez Celdrán writes of short voiceless stops, “[i]n fact, if attention is paid to the sound in question, it is perceived as voiced. By contrast, if we hear the whole word, the sound is perceived as voiceless, based on the use that we make of our knowledge of the word and the context.” (Martínez Celdrán 2008: 43). He takes this as evidence that the feature [\pm tense] must distinguish the stops here, but it can easily be taken as evidence that the stops in fact do not differ absolutely, but that contextual knowledge (e.g., all segments are short because the talker is speaking at a fast rate) conditions our perception of the stops. The question of the articulatory or acoustic causes or consequences of tension are left unresolved.

An issue with both proposals discussed here is that, in focusing on the phonological distinction between /p, t, k/ and /b, d, g/, they fail to account for the large amount of variation in production. Why, for example, do both tense and lax stops surface as stops after a nasal or phrase initially but lax stops spirantize elsewhere? It is not made clear what the relationship between phonological representation and attested phonetic variability is.

Two relatively recent proposals have attempted to do just that. Lavoie (2001), in her book on lenition, finds the voiced stops are consistently shorter than and have a higher intensity ratio than voiceless stops. She concludes that /b, d, g/ are therefore underlying approximants (in contrast, one assumes, with the voiceless stops) that undergo strengthening, which “may simply be an articulatory error, an overshoot closing gesture” (Lavoie 2001: 169). Additionally, the realization of stops as approximants is due to either mistiming of the oral and nasal gestures or a repair to a sequence of nasal + continuant segments that are articulatorily incompatible. Carrasco and Hualde’s approach is similar in general terms, though it makes explicit reference to articulatory gestures (Carrasco and Hualde 2009). These authors argue that, while spirantization of voiced stops most likely started as gestural reduction, since intervocally it is nearly universal, the constriction target for the stops in that position must be that of an approximant. They rely on an allophonic distribution of articulatory gestures (full stops after nasals and at a phrase boundary, approximants elsewhere) to derive the attested patterns.

It is not clear from these proposals, however, why /b, d, g/ should surface as stops after nasals. Honorof (1999) showed that nasals assimilate to both the place and constriction degree of following non-continuant, so that in an /ns/ sequence [n] is produced with the same stricture as [s]. Given that these nasal + continuant sequences are possible, and the fact that /b, d, g/ have the same surface constriction degree as fricatives (Romero 1995), we would expect /mb/ to be realized as [mβ], rather than the attested [mb]. Carrasco and Hualde’s solution of allophony, while explanatorily adequate, does not address any larger causes behind the distribution of allophones. Nor does either theory account for the differences seen in constriction degree due to stress, speech rate, or segmental context.

1.3. *Proposal*

This paper argues for an analysis of Spanish stop spirantization based in Articulatory Phonology (Browman and Goldstein 1992). In this theory, the basic units of abstract phonological contrast are gestures, which are also taken to be the units involved in the control of articulator motion. These gestures are goal-directed actions with particular dynamical parameters set for stiffness (implemented via a mass-spring model), constriction degree, and constriction location. Importantly, and different from most other theories of phonology, the duration of gestures is also explicitly part of the specification for each gesture.

The argument here leverages the fact that temporal differences can serve as a contrast between two gestures, and is based on two findings from previous studies. First, Spanish voiced and voiceless stops show significant durational differences. Second, there is ample evidence that duration and constriction degree of voiced stops in Spanish are related, and that productions of these segments with longer temporal durations have more spatial occlusion. It follows, then, that the shorter duration of voiced stops may lead to their less constricted productions, and that the difference between voiced and voiceless stops (in addition to the presence of a glottal spreading gesture for the voiceless stops) may lie not in their target constriction degree but solely in their duration. Constriction differences, as well as the widespread variation in constriction within each class, can be attributed to the temporal differences.

From the relatively simple assumption that voiced and voiceless stops differ in duration, a number of consequences fall out. First, the same gestural target as that for voiceless stops will result in greater undershoot, and systematic spirantization, for voiced stops due to their shorter active duration. Second, this account provides a straightforward explanation of phrase-initial strengthening of voiced stops. It has been shown that increased duration leads to a closer approximation/achievement of the target for gestures at a prosodic boundary (Byrd et al. 2000; Cho and Keating 2001; Byrd and Saltzman 2003; Byrd et al. 2006; Cho 2006) Given this extended duration, it follows that gestures at a phrase boundary will achieve their target. This extended duration, however, will result in full closure (as is the case in Spanish) only if the target is full closure; adding duration to a gesture for an approximant, for example, will not result in closure. By explaining the surface alternation in constriction degree as the dynamic consequences of interactions of a single, invariant spatial target and variable gesture duration, this approach avoids needing to posit different allophones at the level of gestural control. Last, this theory predicts (in agreement with previous work) that the amount of spirantization will vary with duration; shorter productions will also be less constricted.

An articulatory study was conducted to test the above hypothesis. The study aims to examine the duration of the relevant constriction gestures, their constriction degree, and the relationship between constriction degree and duration. Results from that experiment are reported in Section 2, followed by a computational model of that study's results in Section 3.

2. Articulatory study

2.1. Methods

2.1.1. Stimuli and subjects For this study, data were collected from two subjects (A and B), both native speakers of northern peninsular Spanish. The subjects lived in Spain until attending school in the US and since then have lived on-and-off in

Table 1. Target words divided by independent variables.

Boundary	Stress	/p/	/b/ ¹
Phrase boundary	Stressed σ	<i>pága</i> (/paɡa/) '(s)he pays'	<i>vága</i> (/baɡa/) '(s)he wanders'
	Unstressed σ	<i>pagába</i> (/paɡaba/) '(s)he payed'	<i>vagába</i> (/baɡaba/) '(s)he wandered'
Word boundary	Stressed σ	<i>pánta</i> (/panta/) 'ribbon'	<i>bánda</i> (/banda/) 'band'
	Unstressed σ	<i>pantálla</i> (/pantaja/) 'screen'	<i>bandáda</i> (/bandada/) 'flock'
No boundary	Stressed σ	<i>tapádo</i> (/tapado/) 'covered'	<i>chavál</i> (/tʃabal/) 'kid'
	Unstressed σ	<i>tápa</i> (/tapa/) 'lid'	<i>faltába</i> (/faltaba/) '(s)he delayed'

the US and Spain. The subjects, even while in the US, use Spanish on a daily basis. Neither subject reported any previous history of speech or hearing impairment.

Stimuli were designed in order to examine articulatory differences in production of voiced and voiceless stop consonants in a variety of prosodic conditions. Due to difficulty in measuring the constriction degree for velar and coronal consonants (Romero 1995), stimuli were limited to the labial stops (/p/ and /b/). Measuring coronal constriction for Spanish is particularly difficult as the very tip of the tongue, rather than the blade, is used. A pilot conducted for this study showed very unreliable measurements for tongue tip constriction degree with a sensor approximately 8 mm dorsal to the tongue apex. In all cases, the stimulus contained the sequence /aCa/, where C is one of the target stops. In addition to voicing (voiced vs. voiceless), conditions were included to test the effect of prosodic boundaries on the production of intervocalic stops. The prosodic boundaries examined included: phrase boundary, word boundary, and no boundary (word internal) conditions. Lexical stress for the target words was balanced, with target stops occurring in the onset position of both stressed and unstressed syllables. These variables, along with the stimulus word used in the experiment, are presented in Table 1 (stress, indicated by ' , is shown even where not normally marked in the orthography). Target words were chosen to minimize the effects of coarticulation by alternating labial and coronal/velar stops when possible. To create each stimulus, the target words were embedded in a carrier phrase. The carrier phrases used for each condition, presented in Table 2, were varied in order to avoid any possible repetition effects.

Due to an error in the presentation of the stimuli, no data were collected for /b/ in word-internal position. Data for this condition were taken from other sentences collected during the same experimental session. For *chaval*, relevant data were taken from all repetitions of the sentences *El chaval tardaba demasiado* | *El chaval*

Table 2. Carrier phrases for different prosodic boundary and stress conditions.

	Stress	Carrier phrase
Phrase boundary	Stressed σ	<i>La chica juega. ____ también.</i> ‘The girl plays. She ____ as well.’ <i>El niño salta. ____ también.</i> ‘The boy jumps. He ____ as well.’
	Unstressed σ	<i>El chico jugaba. ____ también.</i> ‘The boy played. He ____ as well.’ <i>La niña saltaba. ____ también.</i> ‘The girl jumped. She ____ as well.’
Word boundary	Stressed/Unstressed σ	<i>El chico canta ____ dos veces.</i> ‘The boy sings ____ twice.’ <i>La niña canta ____ muchas veces.</i> ‘The girl sings ____ many times.’
No boundary	Stressed σ /Unstressed σ	<i>Las chicas cantan ____ dos veces.</i> ‘The girls sing ____ twice.’ <i>Los niños cantan ____ muchas veces.</i> ‘The boys sing ____ many times.’

danzaba demasiado; for *faltaba*, from all repetitions of the sentences *El chico canta faltaba dos veces* / *La niña canta faltaba muchas veces*.

The stimuli were randomized across variables in 6 blocks. For each target word, a different carrier phrase was used in odd- and even-numbered blocks so that 1) each block contained one and only one stimulus per target word and 2) adjacent blocks contained different stimuli. The stimuli were presented on a computer monitor positioned roughly 1 m away from the subject. When cued by an auditory click, the subject read the sentence on the screen; after the subject read each sentence, the monitor changed to show the next sentence in the block and there was a roughly two second pause before the subject was cued again. There was a break of a few minutes at the end of every block. While it was intended to collect six blocks per subject, only four could be collected for subject A due to time limitations. This gave four repetitions of each target for subject A and six for subject B. Prior to data collection, the subjects were instructed to speak in a casual, relaxed manner, as if speaking with close friends.

2.1.2. *Data collection* Articulatory data was collected using an electromagnetic articulometer (Carstens AG500). This device allows three-dimensional tracking of transducers glued to various points in the subject’s vocal tract. For this study, transducers were attached at the vermilion border of both the upper and lower lips, the tongue tip (for data collected concurrently for a separate study), a point on the tongue dorsum approximately 2 cm posterior to the tongue tip sensor (as above), and the lower jaw. Additionally, reference sensors were attached to the bridge of the nose and behind each ear; a sample of the subject’s occlusal plane was also

taken. Articulatory data were collected at 200 Hz, and acoustic data at 16 kHz. After collection, the articulatory data was smoothed with a 9th-order Butterworth low pass filter, rotated to match the subject's occlusal plane and corrected for head movement using the reference sensors.

2.1.3. Data analysis In order to measure Lip Aperture (LA), the Euclidean distance in the sagittal plane between the sensors on the upper and lower lip was calculated. This derived variable was used for all subsequent analysis.

Gestural identification was conducted using the MVIEW software package, developed by Mark Tiede at Haskins Laboratories. The identification algorithm used takes as input a manually located estimate of the midpoint of constriction of one EMA sensor or derived variable. Using the velocity of that sensor or variable (the absolute value of the first difference of the signal), it then locates the velocity minimum crossing closest to the input point (measurement point: Time of maximum constriction). It then finds the peak velocity between that point and both the preceding and following velocity minima (measurement point: time of peak velocity). It then locates the onset of gestural motion by locating a point where the velocity signal from the preceding minimum to the first time of peak velocity crosses some arbitrary threshold of the velocity difference between the two points. Gestural offset is defined as the point where the velocity falls below the same threshold from the second time of peak velocity to the velocity minimum following the point of maximum constriction. Onset and offset of the constriction proper are also defined by the points where the velocity crosses a threshold between the times of peak velocity and the point of maximum constriction. A representative example is shown in Figure 1.

Thresholds for identification of constriction onset and offset were set at 30%, and those for gesture onset and offset at 20%. These thresholds are consistent with past work (e.g., Byrd et al. 2008) and allowed for consistent measurements across subjects, with small fluctuations in velocity during the constriction captured between the points of constriction onset and offset. All locations were checked by hand. For one trial of word-internal /b/, the algorithm incorrectly identified the points of maximum constriction, constriction release, peak offset velocity, and gesture offset (these points were labeled after the following vowel). This trial was excluded from further analysis.

From these measurements, a number of derived variables were calculated. First, total duration of the gesture was defined as the time between the gesture onset and the constriction release. This latter point was chosen as it coincides with the theoretical end of active control of the gesture (Browman and Goldstein 1989; Saltzman 1995), and has been used in previous work as the location of the end of a gesture (Pouplier and Goldstein 2010). Movement after this point is heavily influenced by the following gestures. Total duration was further broken down into constriction duration (time between onset and release of constriction) and movement duration (time from gesture onset to constriction onset). In order to better compare LA

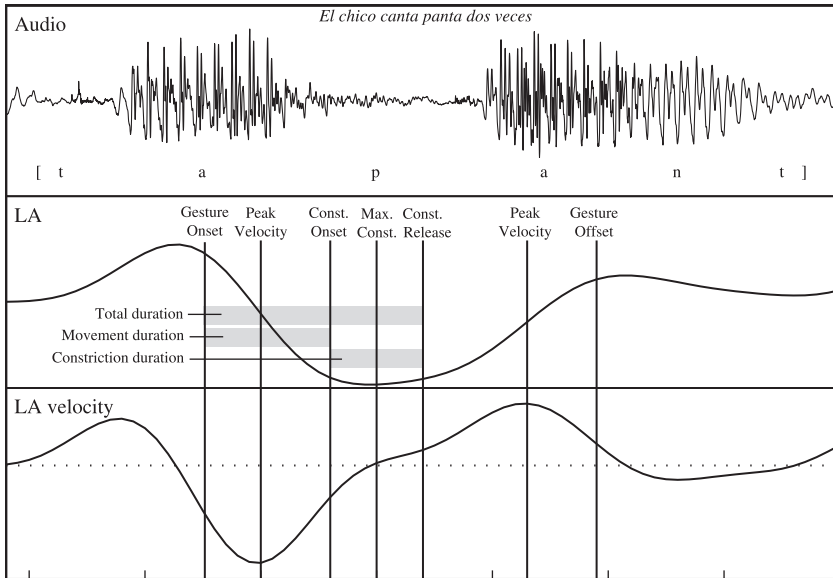


Figure 1. Representative example of measurement of a constriction gesture. The example here is *[p]* taken from “El chico canta panta dos veces.” LA in the figure refers to Lip Aperture, the Euclidean distance in the sagittal plane between the sensors on the upper and lower lips.

across both subjects, normalized constriction degree was also calculated. This was done by measuring the minimum value (most constricted) across all tokens at the point of maximum constriction and subtracting that value from the measurement for each individual token, giving a scale where 0 mm is the most constricted, and higher values correspond to more open productions.

Statistical analyses were conducted using linear mixed models implemented in the ‘lme4’ package in R (Bates 2005). For all tests, fixed factors were boundary (phrase boundary, word boundary, or no boundary), segment (*/b/* or */p/*), and an interaction between boundary and segment. Subject was included as a random intercept. Markov Chain Monte Carlo sampling based on the *t* statistic was used to estimate the *p* values (Baayen et al. 2008). Post-hoc comparison of factor levels was done with paired *t*-tests using a Bonferroni correction for multiple comparisons with an experiment-wise alpha of 0.05. Regressions were conducted with the linear regression function ‘regress’ in MATLAB.

2.2. Results

2.2.1. *Duration* For duration, there was a significant effect of prosodic boundary ($t = -11.40$, $p < 0.0001$), though not of voicing ($t = -0.86$, n.s.). There was a

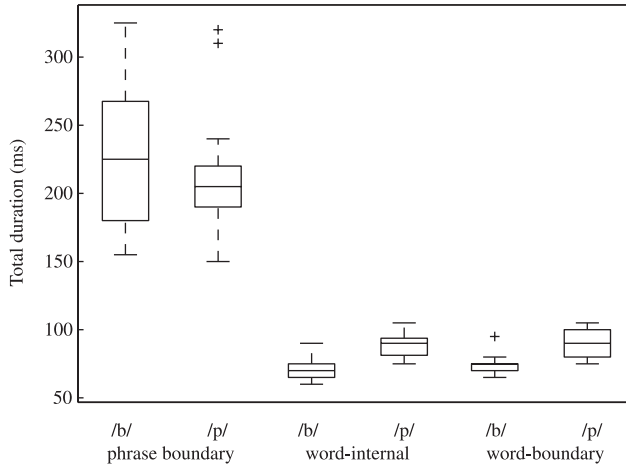


Figure 2. Total duration by prosodic boundary and voicing. For this and all following boxplots, the center mark represents the median, the edges of the box are the 25th and 75th percentiles, and the whiskers extend to those values not considered outliers (more than $\sim 2.7\sigma$ from the mean). /b/ is shorter than /p/, except at a phrase boundary. Durations of both /b/ and /p/ are longer phrase-initially than in either phrase-medial position.

marginally significant interaction between the two factors ($t = 1.76$, $p < 0.08$). A box plot showing the durations at the different conditions is shown in Figure 2. Given that post-hoc tests reveal a significant difference between phrase boundary ($M = 219$ ms) and non-phrase boundary ($M_{\text{word internal}} = 79$ ms, $M_{\text{word boundary}} = 83$ ms) conditions and no difference between the non-phrase boundaries, a separate model was constructed combining the two phrase-internal positions. In this case, there was a significant effect of boundary ($t = 19.48$, $p < 0.0001$), voicing ($t = 2.64$, $p < 0.01$), and their interaction ($t = -2.90$, $p < 0.01$). At the non-phrase boundary level, voiceless stops ($M = 90$ ms) are significantly longer than voiced stops ($M = 72$ ms); there is no difference at the phrase boundary level.

The pattern for constriction duration closely mirrors that for total duration. Figure 3 shows the constriction duration split by prosodic boundary and voicing. There is a significant effect of prosodic boundary ($t = -11.70$, $p < 0.0001$), though not for voicing ($t = -1.15$, n.s.). There is, however, a significant interaction between the factors ($t = 2.41$, $p < 0.05$). As for total duration, there is a difference between phrase boundary ($M = 131$ ms) and non-phrase boundary conditions ($M_{\text{word internal}} = 30$ ms, $M_{\text{word boundary}} = 31$ ms), with no difference between word internal and word boundary conditions. Again, a second model combining phrase-medial shows significant effects of all factors ($t_{\text{boundary}} = 20.32$, $p < 0.0001$; $t_{\text{voicing}} = 3.76$, $p < 0.0001$; $t_{\text{interaction}} = -4.10$, $p < 0.0001$). Post-hoc tests reveal a significant difference between voiced ($M = 22$ ms) and voiceless ($M = 40$ ms) phrase-internally but not at a phrase boundary.

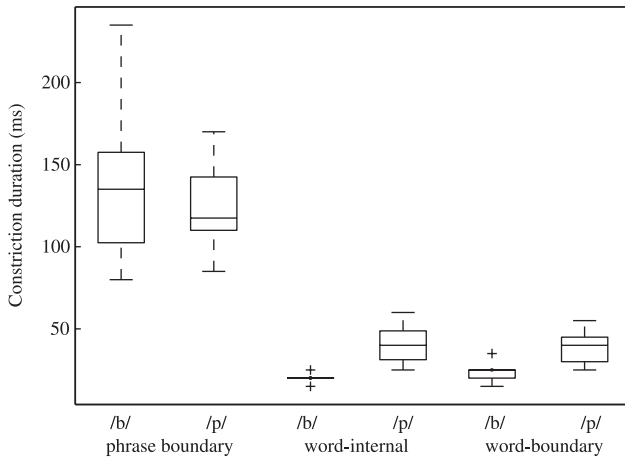


Figure 3. Constriction duration by prosodic boundary and voicing. *[b]* is shorter than *[p]*, except at a phrase boundary. Durations of both *[b]* and *[p]* are longer phrase-initially than in either phrase-medial position.

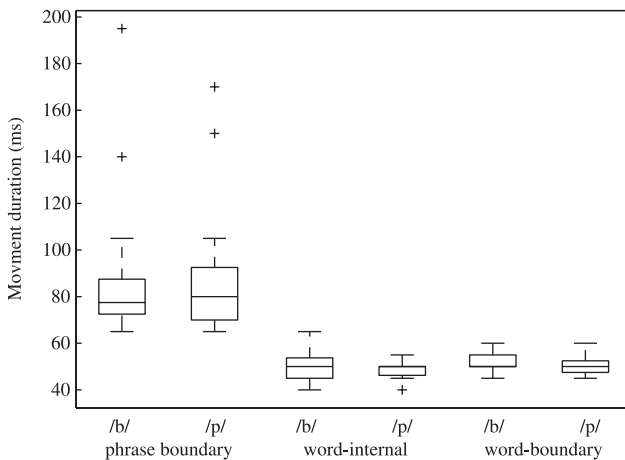


Figure 4. Movement duration by prosodic boundary and voicing. There is no significant difference in duration between *[b]* and *[p]* for any prosodic position.

Movement duration, like constriction and total duration, shows a significant main effect of prosodic boundary ($t = -6.45, p < 0.0001$), but differs in that it shows no significant effect of voicing and no interaction between the factors. A model combining word- and no-boundary conditions shows the same result. Movement duration is shown in Figure 4. For prosodic boundary, there is a significant difference between phrase boundary (M = 88 ms) and non-phrase boundary conditions

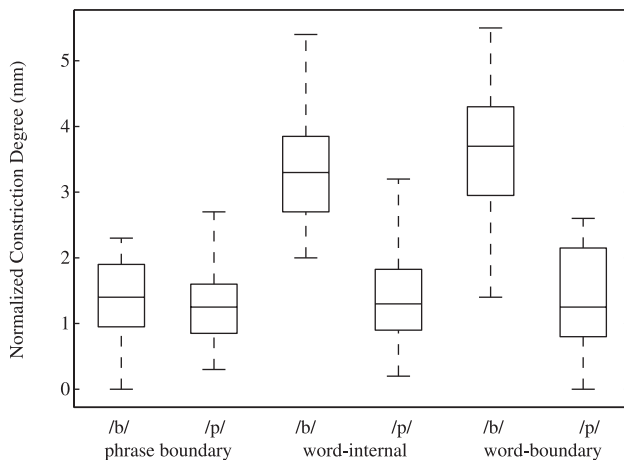


Figure 5. *Lip Aperture by prosodic position and voicing. /b/ is more open phrase medially than phrase initial /b/ and /p/ in all phrasal positions.*

($M_{\text{word internal}} = 49$ ms, $M_{\text{word boundary}} = 51$ ms). As for total duration, the difference between word internal and word boundary conditions is not significant. There is no difference between /b/ and /p/ for any boundary condition.

2.2.2. Constriction degree For Lip Aperture, there was a significant effect of prosodic boundary ($t = 7.13$, $p < 0.0001$) and a marginally significant effect of voicing ($t = -1.75$, $p < 0.08$), as well as a significant interaction between the two factors ($t = -4.87$, $p < 0.0001$). Post hoc tests reveal a significant difference between phrase medial /b/ ($M_{\text{word internal}} = 3.4$ mm, $M_{\text{word boundary}} = 3.5$ mm), on the one hand, and phrase initial /b/ ($M = 1.4$ mm) and /p/ in all prosodic positions ($M_{\text{word internal}} = 1.3$ mm, $M_{\text{word boundary}} = 1.4$ mm, $M_{\text{phrase boundary}} = 1.3$ mm), on the other. Lip Aperture is shown by prosodic boundary and voicing in Figure 5.

2.2.3. Relationship between constriction degree and duration In order to test the initial hypothesis that the reduced constriction degree of voiced stops was due to durational differences between voiced and voiceless stops, constriction degree was regressed onto the duration measures. Additionally, since there was no difference in duration or constriction degree between the two phrase internal boundary conditions (word boundary and word internal), they have been collapsed and treated as a single condition; phrase-initial and -medial conditions are treated separately. Correlations are presented for each subject individually.

For the phrase initial condition, there is no relationship between any measure of duration and constriction degree for either /p/ or /b/ or for either subject. As both are produced with full closure, this is not unexpected – once a gesture reaches the

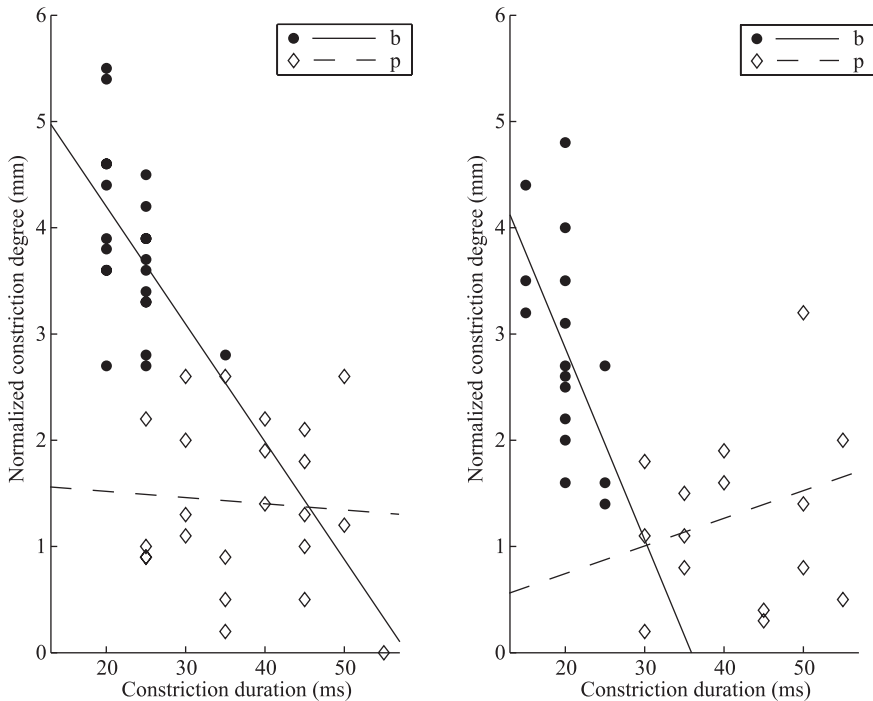


Figure 6. Regression analysis between constriction degree and constriction duration for *[p]* and *[b]*, as measured by LA. Subject A is on the left; subject B, on the right. Correlations are significant for *[b]* for both subjects; neither subject shows a significant correlation for *[p]*.

point of closure, little change will be shown in Lip Aperture even with increased duration. For the phrase medial condition, there is a significant negative relationship between constriction degree and constriction duration measures for *[b]* for both subjects (A: $R^2 = 0.26$, $p < 0.02$; B: $R^2 = 0.32$, $p < 0.03$). Additionally, subject A shows significant correlations of constriction degree with both movement ($R^2 = 0.21$, $p < 0.05$) and total duration ($R^2 = 0.31$, $p < 0.005$), though subject B does not. There is no significant correlation of constriction degree with any duration measure for *[p]* for either subject (again, these are produced with full closure so we do not expect to see an effect of duration). A plot of constriction degree and constriction duration, along with the relevant regression lines, is shown in Figure 6.

2.2.4. *Acoustics* While the primary focus of this paper is an articulatory analysis, a brief overview of the acoustic results may be informative. The stops produced by the subjects in this study followed the well-attested pattern in Spanish. Voiceless stops in phrase medial position were categorically realized with full closure, as

indicated by the consistent presence of a silent period followed by a release burst. They were unvoiced, with 5–20 ms of VOT. Voiced stops in the same context were always realized as spirants, with robust formant values visible throughout the consonant period and no release burst. In phrase initial position, voiceless stops again were produced with full closure and a short VOT. Voiced stops were generally produced as full stops with a period of prevoicing ranging from 0 to 103 ms. These results are consistent with the articulatory findings.

2.3. Discussion

The initial hypothesis had 3 main predictions: 1) voiced and voiceless stops differ in duration 2) voiced and voiceless stops differ in their produced constriction degree, at least phrase medially, and 3) the durational differences between the two classes of stops underlie the constriction differences. We can clearly see that voiced and voiceless stops do indeed differ in duration, with the voiceless stops roughly 15–20 ms longer than their voiced counterparts in phrase medial position, though this difference is not present at a phrase boundary. This difference is attributable to the duration of the constriction of the gesture, and not the movement prior to constriction.² Notably, this durational difference is somewhat smaller than what has been reported in previous literature. Lavoie (2001), for example, reports a difference of 41 ms between /p/ and /b/, and 51.5 ms for /t/ and /d/, though the finding here is similar to the 18 ms difference reported for /k/ and /g/. The differences between the current study and previous work may be due to the inherent inaccuracies in measuring the duration of a non-stop consonant without clear breaks in the visible formants from the acoustic signal and comparing those measurements to more easily measurable stops. The approach taken in this paper, measuring the duration of the gestures directly from articulator movement, seems a more accurate method of comparing gestures with full occlusion to those without. Though the thresholds used for gestural identification here are somewhat arbitrary, they are consistent between /p/ and /b/.

The data here also clearly support the traditional description of phrase medial spirantization. Voiced stops in phrase medial position have a more open posture at their point of maximum constriction when compared to voiceless stops, but phrase initial voiced and voiceless stops have the same constriction degree, equal to that of voiceless stops phrase medially.

Because /b/ and /p/ differ in both constriction degree and duration, this paper initially proposed the hypothesis that the two do not differ in their target constriction degree, but that constriction differences between voiced and voiceless stops are due to shorter duration for the former, which leads to a greater degree of undershoot. However, the data from this study do not support this proposal. If we look to Figure 7, we can see that for any measure of duration, there is significant overlap in the duration of /p/ and /b/, but little to no overlap in their respective constriction

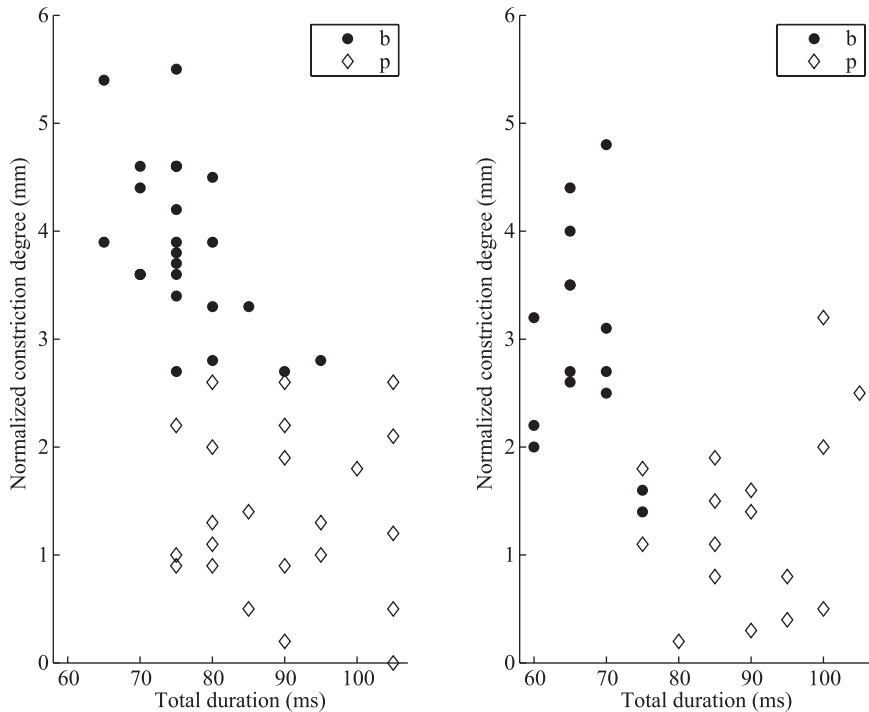


Figure 7. Duration and constriction degree of *[p]* and *[b]*. Data from subject A is on the left; data from subject B, on the right. Durations for subject A overlap around 75–80 ms but repetitions show much less overlap in constriction degree.

degrees. For example, when stops have a duration of 75–80 ms, the voiceless stops are consistently realized with a LA roughly 2 mm less than the voiced stops for subject A.

It would seem that we must conclude that voiced and voiceless stops have different targets for constriction degree. However, there is one additional possibility that must be ruled out. There is ample evidence that it is possible to vary the speed of articulator movement independently of changes in the magnitude and duration of that movement. This has been referred to as the *stiffness* of a gesture (Edwards et al. 1991; Beckman and Edwards 1992; Byrd and Saltzman 1998; Roon et al. 2007). A gesture with lower stiffness will take longer to reach its target position than the same gesture with a higher stiffness. If voiced stops were to have a lower stiffness than voiceless stops, then, we would expect the voiceless stops to more closely approximate their target than the voiced stops given the same duration of gestural activation. This hypothesis was tested from the experimental data, again including only labial stops due to the unreliable constriction degree data for coronals. Stiffness was calculated in two separate ways. First, following Roon et al.

(2007), as the peak velocity of the closing gesture (taken as the velocity measurement at the point of peak velocity between gestural onset and constriction onset), divided by the magnitude of the closing gesture (calculated as the difference in LA between the point of maximum constriction and gesture onset). Second, after Byrd and Saltzman (1998), stiffness was measured by the time from gestural onset to the peak velocity of the closing movement.

For the velocity/magnitude measure, there is a significant effect of prosodic boundary ($t = -10.72, p < 0.0001$). Post hoc tests reveal that stiffness at a prosodic boundary ($M = 1.50$) is lower than that phrase medially ($M_{\text{word internal}} = 2.35, M_{\text{word boundary}} = 2.41$). This difference is expected, as there is a well-documented effect of lower effective stiffness of gestures at a prosodic boundary (Edwards et al. 1991; Beckman and Edwards 1992, Byrd and Saltzman 1998). Importantly, however, there is no difference at all between voiced and voiceless stops at any prosodic level. For time to peak velocity, we find the same pattern. There is a main effect of prosodic boundary ($t = 5.14, p < 0.0001$) and no effect of voicing nor an interaction between the two factors. Post hoc tests indicate longer time to peak velocity at a phrase boundary ($M = 49$ ms) than phrase medially ($M_{\text{word internal}} = 23$ ms, $M_{\text{word boundary}} = 25$ ms). Based on the evidence from both measures of stiffness, we must conclude that stiffness cannot be driving the observed difference in constriction degree between voiced and voiceless stops.

We must conclude, given that we have excluded stiffness differences, that voiced and voiceless stops differ in their target constriction degree, which is reflected by constriction differences at the same total gestural and constriction durations. Does this imply that the voiced stops have a target constriction degree similar to that for an approximant? Not necessarily. Recall that the voiced stops in phrase medial position showed significant effects of duration on constriction degree. As the duration increased, stops became more constricted. Additionally, at very long durations (i.e., at a phrase boundary) the stops are realized with full closure. This is consistent with a target constriction degree that results in full closure, at least at a long duration.

There is evidence to support the fact that stops in general have targets for constriction degree that are actually beyond the point of contact between the articulators (Löfqvist and Gracco 1997; Löfqvist 2005). We know that voiced stop targets must be less constricted than that for voiceless stops. We can hypothesize, then, that the goal for voiced stops is simply closer to, but still beyond, the point of articulator contact than that for voiceless stops. If the short duration of both stops causes roughly the same amount of undershoot (if, of course, both stops have equal stiffness), it might be possible that this same absolute degree of undershoot would still result in full contact for the voiceless stops, but lead to incomplete closure and spirantization for the voiced. However, at long durations (such as that at a phrase boundary), both stops would have time to reach their targets, resulting in complete occlusion for both. In order to test this hypothesis, a gestural simulation was conducted, the results of which are presented in Section 3.

3. Gestural simulation study

3.1. *The Task Dynamics model*

The simulation study was conducted using the Task Dynamic Application (TaDA) developed at Haskins Laboratories to produce both articulatory and acoustic output from an input of gestural parameters (Nam et al. 2004; Saltzman et al. 2008). The model is an implementation of the theories of Articulatory Phonology (Browman and Goldstein 1992) and Task Dynamics (Saltzman and Munhall 1989). Within these theories, articulatory constriction actions are the basic compositional units of speech. These context-invariant gestures' temporal patterning is modeled by means of intergestural coupling relations, represented in a coupling graph for a given utterance (Browman and Goldstein 2000; Goldstein et al. 2006; Goldstein et al. 2008). The coupling graph in turn both represents an utterance's syllabic phonological structure and determines the coordination of the gestures in that utterance. For any given arbitrary input from English or Spanish (Parrell et al. 2010), gestural information is accessed from a dictionary and a coupling graph is constructed. From the coupling graph, a gestural score is created with the activation times and durations of the various gestures.

3.2. *Model input for Spanish stops*

Within the TaDA model gestures are specified by their constriction target, duration, and stiffness. For this study, the words *cava* (/kaba/) and *capa* (/kapa/) were used as inputs to the model. Word-initial /k/ and vowels /a/ were chosen in order to a) minimize coarticulatory influences on the lips and b) accurately model the words used in the EMA experiment in Section 2. The two words differed in the target constriction degree for the labial stop as well as in the presence or absence of a glottal spreading gesture to generate the voiceless /p/. For /p/, target constriction was set to -2 mm – that is, 2 mm beyond the point of contact between the upper and lower lips. This target is the standard in the English version of the model, and accords with articulatory evidence for hyper-articulated targets. The target for /b/ was set to -.5 mm, less than the target for /p/ but still beyond the point of contact for the lips. An additional utterance was created with the target for /b/ set to 0 mm to test the hypothesis that the target for /b/ must, like that for /p/, be beyond the point of articulatory closure. After the model generated the utterances, the durations of the lip aperture gestures were modified directly in the gestural score. To model the phrase medial condition, duration was set to 80 ms. This reflects a relatively short /p/ and a relatively long /b/, but there are numerous occurrences of both stops with this total duration in the EMA study. For the phrase initial condition, duration was set to 200 ms, again based on the durations found in the articulatory study. For both /b/ and /p/, stiffness was set to the model default.

Table 3. *Lip Aperture at the point of maximum constriction for /p/ and /b/ with two different constriction degree targets generated by TaDA. The undershoot is the difference between target and maximum achieved constriction degree.*

duration	/p/ (target: -2 mm)		/b/ _{-0.5} (target: -0.5 mm)		/b/ ₀ (target: 0 mm)	
	Max LA	undershoot	Max LA	undershoot	Max LA	undershoot
80 ms	-0.8 mm	1.2 mm	0.6 mm	1.1 mm	1.0 mm	1.0 mm
200 ms	-2 mm	0 mm	-0.5 mm	0 mm	0 mm	0 mm

3.3. Results

The articulatory output from TaDA was analyzed using the same MVIEW software and settings as described in Section 2.1.3. Gestural landmark identification was conducted on the LA signal created by TaDA, which is generated via the same calculation of Euclidean distance between the upper and lower lips as was used previously. LA at the point of maximum constriction is compared in Table 3. In order to test the hypothesis about different consequences of undershoot for voiced and voiceless stops, the amount of undershoot (in mm) was calculated as the maximum constriction less the target constriction degree.

As predicted, at each duration, /p/, /b/_{-0.5}, and /b/₀ show the same amount of undershoot. At 80 ms, all undershoot their targets by roughly more than 1 mm. At 200 ms, all reach their targets. Though the amount of undershoot for the three gestures is equal, the consequences of that undershoot are not. For /p/ undershooting the target still results in closure (any LA of 0 or less is complete closure of the lips). For both /b/s, however, the same absolute amount of undershoot results in an incomplete closure, with a maximum LA of 0.6 mm/1.0 mm. At 200 ms, all the gestures reach their target. For /p/ and /b/_{-0.5}, this results in complete closure with a negative constriction degree (which translates to compression of the relevant articulators in real-world speech). For /b/₀ the result is the articulators touching briefly but without any pass-through/compression.

These differences are visible in the acoustic signal generated from the articulatory patterns by the model, shown in Figure 8. TaDA generates these acoustic signals by tracking the changing aerodynamic conditions generated by the model articulators and generating acoustic output from those conditions via HLSyn (Hanson and Stevens 2002). There is a complete closure for /p/ at both durations, with no formants visible and a strong release burst. For /b/_{-0.5} and /b/₀, on the other hand, there are visible formants throughout the closure period at 80 ms, with no visible release burst. At 200 ms, /b/_{-0.5} achieves full closure, with no visible friction and a clear release burst. On the other hand, for /b/₀, achieving the target constriction degree results in a brief period of acoustic silence followed by a period with high-frequency aperiodic noise and no release burst. It may be noted that there is devoicing of both /b/₀ and /b/_{-0.5} when their durations are set to

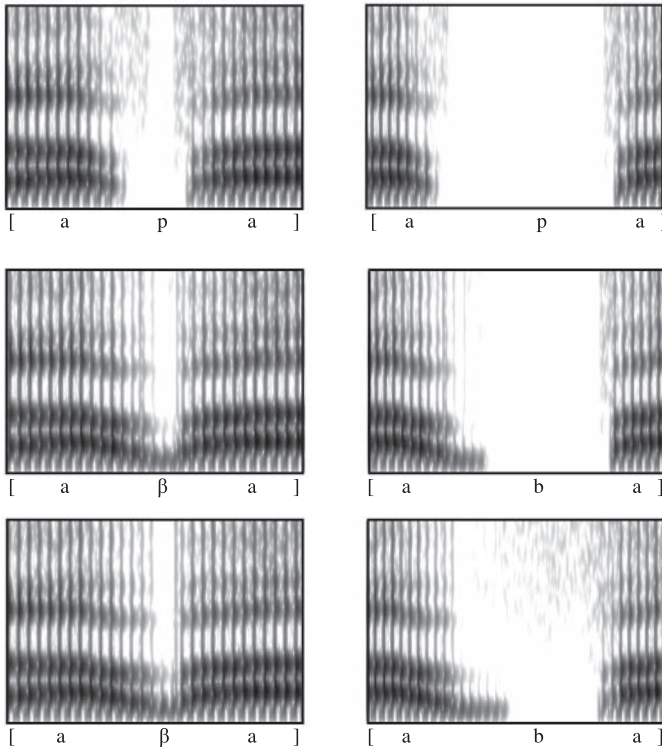


Figure 8. *Spectrograms of acoustic output from TaDA. In the top row, left-to-right: /p/ @ 80 ms, /p/ @ 200 ms; in the middle row, left-to-right: /b_{-0.5}/ @ 80 ms, /b_{-0.5}/ @ 200 ms. In the bottom row, left-to-right: /b₀/ @ 80 ms, /b₀/ @ 200 ms. Both /b_{-0.5}/ and /b₀/ show incomplete closure at 80 ms, but only /b_{-0.5}/ shows the attested variation with full closure at 200 ms.*

200 ms. This is due simply to passive aerodynamic effects as the walls of the oral cavity in the TaDA model are fixed, as is the position of the vocal folds; it does not model the expansion of the oral cavity that may happen either actively or passively in speech to maintain voicing during a long closure (Ohala and Riordan 1979). In any case, this passive devoicing is not particularly relevant to the argument being made here about the relationship between constriction degree and duration.

3.4. Discussion

The evidence from this modeling study supports the hypothesis that voiced and voiceless stops differ in target constriction degree, and that both must have constriction targets beyond the point of closure. Setting the target for /b/ to the point

of articulator contact (0 mm constriction degree), we see no period of occlusion, but rather light frication. Since we most often see full occlusion with release bursts for voiced stops with such long durations, we are forced to conclude the voiced stops must have a target beyond the point of closure. Using a small negative target for /b/ and a large negative target for /p/, on the other hand, does generate the appropriate patterns: /b/ is produced as a spirant at a short duration and as a fully occluded stop at a long duration; /p/ is produced always as a stop. These differences arise even though the amount of undershoot is equivalent for both stops.

4. Discussion

Taken together, these two studies provide evidence that Spanish /b/ should rightly be viewed as a stop, rather than an approximant. While the number of speakers is relatively small, the consistent patterns between speakers, as well as compatible evidence from articulatory modeling and the fact that the findings here are consistent with previous work, suggest these findings may indeed be robust. Importantly, the current study shows that there is a correlation between duration and constriction degree for /b/. If these segments were truly approximants that are produced as stops in certain contexts due to prosodic strengthening, articulatory overshoot or gestural incompatibility (Lavoie 2001), we would not expect to see a relationship between these two articulatory parameters in the absence of variation in stress, phrasal position, or segmental context. Neither is this relationship predicted in an allophonic account of the variable production of these segments (e.g., Carrasco and Hualde 2009).

Combining a hypothesized target for /b/ beyond the point of articulator contact and the established constriction-duration relationship, we can straightforwardly explain what has until now been seen as allophonic variation between stop and approximant productions as the dynamic consequences of an invariant gestural specification for constriction degree and fluctuations in the duration of the gesture. Phrase medially, the short duration of /b/ results in undershoot of the constriction target and, therefore, spirantization, while the increased duration of the closure gesture at a phrase boundary results in full closure. This dynamic process additionally, and uniquely, accounts for at least some of the variation in constriction degree of these sounds phrase medially. Although the initial hypothesis that these phonemes did not differ from /p/ in their target constriction degree was not supported by the data, the results do support that they do have a target constriction degree slightly beyond the point of articulator contact, though less than the hypothesized target for /p/.

While the current study only examined the labial stops, it is reasonable to believe that the theory proposed here also extends to the coronal and velar stops /d, g/ and /t, k/. Both /d/ and /g/ show the same quasi-allophonic alternation by prosodic

position as /b/, and all three are similarly affected by stress and segmental context. Although some of the particular details may be different (see, for example, the surprising finding of greater occlusion of /g/ in /aga/ vs /ugu/ contexts in Cole et al. [1999]), the overall patterns are very similar.

Importantly, this theory of the phonological specification of /b, d, g/ can also account for the variation we see in productions of these segments in other contexts. While the current study did not explicitly test these conditions, the predictions are nonetheless consistent with observed patterns. For these segments, full stops are produced in nasal + stop sequences. While other accounts must posit allophony (Carrasco and Hualde 2009) or articulatory incompatibility between fricatives and nasals (Lavoie 2001) to account for this pattern, it falls out directly in the current account as a dynamical consequence of an invariant gestural specification. In these nasal + stop sequences, the duration of active control of the shared constriction articulators is much longer than that for the intervocalic stops. This increased duration allows time for the articulators to better approximate or reach their target, resulting in full closure. Similarly, it explains why /d/ in /ld/ is realized with a stop while the voiced stops in /lb/ and /lg/ are produced as approximants. Since after /l/ the tongue tip is already at or close to its target for /d/, a stop is produced. On the other hand, because the tongue tip does not affect the duration or starting position of gestures involving the tongue lips or tongue dorsum, /b/ and /g/ are spirantized in this context. Our account also agrees with findings that /b, d, g/ are generally produced with greater constriction when they occur between high compared to low vowels (Ortega-Llebaria 2004; Carrasco and Hualde 2009; Hualde et al. 2010; though cf. Cole et al. 1999). As the articulators in the former case will start closer to their target than in the latter, the same duration will result in a closer approximation of the gestural target between high than low vowels. Lastly, the current account can also explain differences in produced constriction degree due to prosodic variation. The increased duration of consonants in a stressed syllable (Cole et al. 1999; Ortega-Llebaria 2004; Eddington 2010) leads directly, in this account, to more constricted productions; similarly, the reduced duration of these segments in fast speech (Soler and Romero 1999) leads to more undershoot and greater spirantization.

None of this prosodic or contextual variation (nor the relationship between duration and constriction degree in the absence of such variation) is predicted by an allophonic or approximant-target hypothesis. While it may certainly be possible to modify these theories to account for this variation, in the current proposal they can all be seen as the lawful dynamical consequences of a single invariant gestural target and variation in duration and initial position due to phonological specification, contextual effects, and prosodic position. This invariant gestural target, specifically, is one beyond the point of articulator contact.

It should be noted that the final theory proposed here is not radically different from the analysis of phrase-initial spatial strengthening and temporal lengthening in Cho and Keating (2001). That study found that phrase-initial coronal

stops in Korean were produced both with more lingual contact (measured by eletropalatography [EPG]) and greater duration than those same stops produced phrase-medially. This is particularly true of the lax stop /t/ and the nasal /n/, while the tense and aspirated stops (/t^{*}/ and /t^h/) show a much smaller difference. In fact, there is no difference in contact degree between the three stops and the nasal phrase initially, while there is a split between lax/nasal and tense/aspirated stops phrase medially. They say that the short durations found phrase medially cause undershoot of the gestural target, leading to less contact. This very closely parallels the findings of the current study: both /p/ and /b/ have increased duration and full contact phrase-initially, while the shorter durations found phrase-medially cause undershoot of /b/. While we did not find an effect of shorter durations on the constriction degree of /p/, that may reflect the limitations of the methodology employed. While EPG is sensitive to degree of palatolingual contact even beyond the point of closure of the oral tract, once the lips contact there is relatively little change in the EMA signal. It is possible that patterns similar to those found for Korean tense and aspirated stops may be found for Spanish /t/.

Given the categorical differences found in VOT between /b/ and /p/ in this study, it seems probable that they also differ in their specification for voicing. There is limited laryngoscopic evidence that /p, t, k/ are in fact produced with some, though relatively small, amount of spreading of the vocal folds (Martínez Celdrán and Fernández Planas 2007). Taking this into account along with the findings from this study, there is evidence for a three-fold distinction between the /b, d, g/ and /p, t, k/: they differ in the duration of the oral constriction gesture, in the constriction degree target of that gesture, and in the presence or absence of a glottal spreading gesture. While it may seem surprising that the two sets of stops differ in so many dimensions, it is important to see these differences in light of the relatively segmental nature of Spanish phonology, where many segments that share the same general feature or gestural category differ in their precise realization of that feature/category. For example, we find fine-grained place distinctions between /t/, /n/ and /s/ in addition to differences in stricture (for /s/) and velar opening gesture (for /n/) differences (Honorof 1999). These place distinctions, while not active in any phonological process, nonetheless exist and influence articulator movement, as can be seen in nasal place assimilation. We can draw a parallel between these and the voiced/voiceless stop distinction. Within Articulatory Phonology, “[g]estures of different organs ‘count’ as different and provide the basis for phonological contrast” (Studdert-Kennedy and Goldstein 2003); thus, the voicing distinction may rely principally on the presence or absence of the glottal spreading gesture (different organs), but the differences in the oral constriction gesture (same organ) are present and play a role in the phonetic production of these sounds.

Additionally, it may be possible to extend the finding that undershoot in voiced stops is the dynamic consequence of decreased duration to the voiceless series as well. As mentioned before, the glottal spreading gesture for the voiceless stops

seems to be rather small in magnitude. We might predict that decreasing the duration of the voiceless stops might then lead to undershoot of the open target for this gesture, resulting in continuous vibration of the vocal folds during the voiceless stop, which would explain the high rate of voicing found for these sounds in previous studies. Additionally, at extremely short durations, we might expect that even the oral closure gesture would undershoot its target, resulting in productions of voiced approximants more or less identical to the phrase medial voiced stops. This is exactly the attested pattern (Machuca 1997; Martínez Celdrán 2009). An experiment to test this hypothesis is currently underway.

The inclusion in this paper of duration at the level of phonological control, though not unprecedented, warrants discussion. In traditional phonology, duration is not part of the grammatical specification but is the result of the phonetic implementation of the phonology. Articulatory Phonology, on the other hand, does include temporal activation intervals in the phonological specification of gestures; in this manner, overlap in time of gestures can explain allophonic variation (Browman and Goldstein 1989; Browman and Goldstein 1995a; and others). While, in this sense, duration is phonologically specified, possible durations have usually been restricted to merely two possibilities: one (relatively shorter) for consonants and one (relatively longer) for vowels (Browman and Goldstein 1995a). The durational differences between voiced and voiceless stops in this paper indicate that this gross distinction may not be sufficient to capture the true role of duration in language. This is not, however, a novel claim. It is well known that voiced stops in many languages are significantly shorter than their voiceless counterparts (e.g., Lisker [1957] and Ladefoged and Maddieson [1996] for English; Braunschweiler [1997] for German). Additionally, increased duration seems necessary to distinguish geminates from singletons in at least some languages, as well as long from short vowels (Ladefoged and Maddieson 1996; Esposito and Di Benedetto 1999; Löfqvist 2006). For Spanish, there is articulatory evidence that fricatives are generally longer than stops, and that this durational difference is phonologically contrastive (Romero 1995). For any theory of phonology, it will be necessary to account for these differences, as in many languages they serve to distinguish one phoneme from another. This may be accomplished, in Articulatory Phonology, by specifying particular gestural durations as a percentage of cycle of the planning oscillator associated with each gesture, following and expanding on the proposal in Browman and Goldstein (1995b). A further discussion of what possible phonological durations might exist is beyond the scope of the current study, but such future work will provide important and necessary insight for phonological theory. It should be noted that the differences proposed here are on par with the durational differences reported for voiced and voiceless stops in other languages (Lisker 1957; Ladefoged and Maddieson 1996; Braunschweiler 1997). As such, whatever underlies the differences in those languages may well cause the durational differences in Spanish as well, and thus alleviate any need to posit Spanish as *uniquely* specifying duration phonologically.

5. Conclusion

In summary, evidence from this study supports the view that /b, d, g/ are voiced stops whose gestural target is one of complete closure, differing in the absence of a glottal spreading gesture from the voiceless stops /p, t, k/. The voiced stops, however, have a somewhat more open constriction target than the voiceless series, though this target is still beyond the point of articulator contact. This reduced constriction target, coupled with a shorter duration, explains the well-documented alternation between production of /b, d, g/ as approximants phrase medially and as full stops phrase initially.

Acknowledgment

This work was supported by NIH grant DC03172.

Correspondence email address: parrell@usc.edu

Notes

1. Phonological /b/ can be represented orthographically by either *b* or *v*.
2. A reviewer points out that this result may be due to the specific value set for the thresholds used in gestural identification. While that is a possibility, the thresholds used were chosen specifically to capture the kinematic patterns shown in the data and are consistent with previous work and so are likely to give a fairly accurate estimation of duration.

References

- Baayen, R. Harald, Douglas J. Davidson & Douglas M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59(4). 390–412.
- Baković, Eric. 1994. Strong onsets and Spanish fortition. *MIT Working Papers in Linguistics* 23. 21–39.
- Barlow, Jessica A. 2003. The stop/spirant alternation in Spanish: Converging evidence for a fortition account. *Southwest Journal of Linguistics* 22. 51–86.
- Bates, Douglas. 2005. Fitting linear mixed models in R. *R News* 5. 27–30.
- Beckman, Mary E. & Jan Edwards. 1992. Intonational categories and the articulatory control of duration. In Yoh'ichi Tohkura, Eric Vatikiotis-Bateson & Yoshinori Sagisaka (eds.), *Speech perception, production, and linguistic structure*, 359–375. Tokyo: Ohmsha, Ltd.
- Braunschweiler, Norbert. 1997. Integrated cues of voicing and vowel length in German: A production study. *Language and Speech* 40(4). 353–376.
- Browman, Catherine P. & Louis Goldstein. 1989. Articulatory gestures as phonological units. *Phonology* 6(2). 201–251.
- Browman, Catherine P. & Louis Goldstein. 1992. Articulatory phonology: An overview. *Phonetica* 49(3–4). 155–80.

- Browman, Catherine P. & Louis Goldstein. 1995a. Gestural syllable position effects in American English. In Fredericka Bell-Berti & Lawrence J. Raphael (eds.), *Studies in speech production: A festschrift for Katherine Safford Harris*, 19–34. Woodbury, NY: American Institute of Physics.
- Browman, Catherine P. & Louis Goldstein. 1995b. Dynamics and articulatory phonology. In Robert Port & Tim van Gelder (eds.), *Mind as motion: Dynamics, behavior, and cognition*, 175–194. Cambridge MA: MIT Press.
- Browman, Catherine P. & Louis Goldstein. 2000. Competing constraints on intergestural coordination and self-organization of phonological structures. *Bulletin de la Communication Parlée* 5. 25–34.
- Byrd, Dani & Elliot Saltzman. 1998. Intra-gestural dynamics of multiple prosodic boundaries. *Journal of Phonetics* 26. 173–199.
- Byrd, Dani & Elliot Saltzman. 2003. The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics* 31. 149–180.
- Byrd, Dani, Abigail Kaun, Shrikanth Narayanan & Elliot Saltzman. 2000. Phrasal signatures in articulation. In Michael B. Broe & Janet B. Pierrehumbert (eds.), *Papers in laboratory phonology V*, 70–87. Cambridge: Cambridge University Press.
- Byrd, Dani, Jelena Krivokapić & Sungbok Lee. 2006. How far, how long: On the temporal scope of prosodic boundary effects. *Journal of the Acoustical Society of America* 120(3). 1589–1599.
- Byrd, Dani, Sungbok Lee & Rebeka Campos-Astorkiza. 2008. Phrase boundary effects on the temporal kinematics of sequential tongue tip consonants. *Journal of the Acoustical Society of America* 123(6). 4456–4465.
- Carrasco, Patricio & José I. Hualde. 2009. Spanish voiced allophony reconsidered. Paper presented at the Phonetics and Phonology in Iberia Conference, Las Palmas de Gran Canaria, Spain.
- Cho, Taehong. 2006. Manifestation of prosodic structure in articulatory variation: Evidence from lip kinematics in English. In Louis Goldstein, Doug Whalen & Catherine Best (eds.), *Laboratory phonology 8: Varieties of phonological competence*, 1–34. Berlin: Mouton de Gruyter.
- Cho, Taehong & Patricia A. Keating. 2001. Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics* 29(2). 155–190.
- Cole, Jennifer, Jose I. Hualde & Khalil Iskarous. 1999. Effects of prosodic and segmental context on /g/-lenition in Spanish. In Osamu Fujimura, Brian D. Joseph & Bohumil Palek (eds.), *Proceedings of the 4th international linguistics and phonetics conference*. 575–589. Prague: Karolinum Press.
- Eddington, David. 2010. What are the contextual phonetic variants of /β, ð, ɣ/ in colloquial Spanish? Paper presented at Laboratory Approaches to Romance Phonology, Provo, Utah, 23–25 September.
- Edwards, Jan, Mary Beckman & Janet Fletcher. 1991. The articulatory kinematics of final lengthening. *Journal of the Acoustical Society of America* 89(1). 369–382.
- Espósito, Anna & Maria G. Di Benedetto. 1999. Acoustical and perceptual study of gemination in Italian stops. *The Journal of the Acoustical Society of America* 106(4). 2051–2062.
- Goldsmith, John. 1981. Subsegmentals in Spanish phonology: An autosegmental approach. In William Cressey & Donna Jo Napoli (eds.), *Linguistic symposium on romance languages 9*, 1–16. Washington DC: Georgetown University Press.
- Goldstein, Louis, Dani Byrd & Elliot Saltzman. 2006. The role of vocal tract gestural action units in understanding the evolution of phonology. In Michael Arbib (ed.), *Action to language via the mirror neuron system*, 215–248. Cambridge: Cambridge University Press.
- Goldstein, Louis, Hosang Nam, Elliot Saltzman & Ioana Chitoran. 2008. Coupled oscillator planning model of speech timing and syllable structure. In *Proceedings of the 8th phonetics conference of China and the international symposium on phonetic frontiers*. Beijing, 18–20 April.
- González, Carolina. 2003. *The effect of stress and foot structure on consonantal processes*. Los Angeles, CA: University of Southern California dissertation.
- Hanson, M. Helen & Kenneth N. Stevens. 2002. A quasiarticulatory approach to controlling acoustic source parameters in a Klatt-type formant synthesizer using Hlsyn. *Journal of the Acoustical Society of America* 112(3). 1158–1182.
- Harris, James. 1969. *Spanish phonology*. Cambridge: MIT Press.

- Herrera, Juana. 1997. Estudio acústico de /p, t, c, k/ y /b, d, y, g/ en gran canaria. In Manuel Almeida & Josefa Dorta (eds.), *Contribuciones al estudio de la lingüística hispánica: Homenaje al profesor Ramón Trujillo*, 73–86. Barcelona: Montesinos.
- Honorof, Douglas N. 1999. *Articulatory gestures and Spanish nasal assimilation*. New Haven, CT: Yale University dissertation.
- Hualde, José I. 1988. *A lexical phonology of Basque*. Los Angeles, CA: University of Southern California dissertation.
- Hualde, José I. 2005. *The sounds of Spanish*. Cambridge: Cambridge University Press.
- Hualde, José I., Miquel Simonet & Marianna Nadeu. (2011). Consonant lenition and phonological re-categorization. *Laboratory Phonology 2* (2).
- Hualde, José I., Miquel Simonet, Ryan Shosted & Marianna Nadeu. 2010. Quantifying Iberian spirantization: Acoustics and articulation. Paper presented at 40th Linguistic Symposium on Romance Languages, Seattle, WA. 26–28 March.
- Kirchner, Robert. 1998. *An effort-based approach to consonant lenition*. Los Angeles, CA: University of California, Los Angeles dissertation.
- Ladefoged, Peter & Ian Maddieson. 1996. *The sounds of the world's languages*. Malden, MA: Blackwell Publishing.
- Lavoie, Lisa M. 2001. *Consonant strength: Phonological patterns and phonetic manifestations*. New York: Routledge.
- Lewis, Anthony M. 2001. *Weakening of intervocalic [p, t, k] in two Spanish dialects: Toward the quantification of lenition processes*. Urbana-Champaign, IL: University of Illinois at Urbana-Champaign dissertation.
- Lisker, Leigh. 1957. Closure duration and the intervocalic voiced-voiceless distinction in English. *Language* 33(1). 42–49.
- Lozano, Maria del Carmen. 1979. *Stop and spirant alternations: Fortition and spirantization processes in Spanish phonology*. Bloomington, IN: Indiana University Linguistics Club.
- Löfqvist, Anders. 2005. Lip kinematics in long and short stop and fricative consonants. *Journal of the Acoustical Society of America* 117(2). 858–878.
- Löfqvist, Anders. 2006. Interarticulator programming: Effects of closure duration on lip and tongue coordination in Japanese. *Journal of the Acoustical Society of America* 120(5). 2872–2883.
- Löfqvist, Anders & Vincent L. Gracco. 1997. Lip and jaw kinematics in bilabial stop consonant production. *Journal of Speech Language-Hearing Association* 40. 877–893.
- Machuca, María Jesús. 1997. *Las obstruyentes no continuas del español: Relación entre las categorías fonéticas y fonológicas en habla espontánea*. Barcelona: Universitat Autònoma de Barcelona dissertation.
- Martínez Celdrán, Eugenio. 1991a. Duración y tensión en las oclusivas no iniciales del español: Un estudio perceptivo. *Revista Argentina de Lingüística* 7(1). 51–71.
- Martínez Celdrán, Eugenio. 1991b. Sobre la naturaleza fonética de los alófonos de /b, d, g/ en español y sus distintas denominaciones. *Verba* 18. 235–253.
- Martínez Celdrán, Eugenio. 1993. La percepción categorial de /b, p/ en español basada en las diferencias de duración. *Estudios De Fonética Experimental* 5. 223–239.
- Martínez Celdrán, Eugenio. 2008. Some chimeras of traditional Spanish phonetics. In Laura Colantoni & Jeffrey Steele (eds.), *3rd conference on laboratory approaches to Spanish phonology*, 32–46. Somerville, MA: Cascadilla Proceedings Project.
- Martínez Celdrán, Eugenio. 2009. Sonorización de las oclusivas sordas en una hablante murciana: Problemas que plantea. *Estudios de Fonética Experimental* 18. 253–271.
- Martínez Celdrán, Eugenio & Ana M. Fernández Planas. 2007. *Manual de fonética española*. Barcelona: Ariel.
- Mascaró, Joan. 1984. Continuant spreading in Basque, Catalan and Spanish. In Mark Aronoff & Richard T. Oehrle (eds.), *Language sound structure: Studies in phonology presented to Morris Halle by his teachers and students*, 287–298. Cambridge, MA: MIT Press.

- Nam, Hosung, Louis Goldstein, Elliot Saltzman & Dani Byrd. 2004. TADA: An enhanced, portable task dynamics model in MATLAB. *Journal of the Acoustical Society of America* 115(5). 2430.
- Ohala, John & Carol Riordan. 1979. Passive vocal tract enlargement during voiced stops. In Jared Wolf & Denis Klatt (eds.), *Speech communication papers presented at the 97th meeting of the Acoustical Society of America*. 89–92. Cambridge, MA: Massachusetts Institute of Technology.
- Ortega-Llebaria, Marta. 2004. Interplay between phonetic and inventory constraints in the degree of spirantization of voiced stops: Comparing intervocalic /b/ and intervocalic /g/ in Spanish and English. In Timothy Face (ed.), *Laboratory approaches to Spanish phonology*, 237–254. Berlin: Mouton de Gruyter.
- Parrell, Benjamin, Mike Proctor & Louis Goldstein. 2010. Towards a computational articulatory model of Spanish phonology. Paper presented at *Laboratory Approaches to Romance Phonology*, Provo, Utah, 23–25 September.
- Piñeros, Carlos-Eduardo. 2002. Markedness and laziness in Spanish obstruents. *Lingua* 112(5). 379–413.
- Pouplier, Marianne & Louis Goldstein. 2010. Intention in articulation: Articulatory timing in alternating consonant sequences and its implications for models of speech production. *Language and Cognitive Processes* (25). 616–649.
- Recasens, Daniel. 1986. *Estudis de fonètica experimental del català oriental central*. Barcelona: Publicacions de l'Abadia de Monserrat.
- Romero, Joaquín. 1995. *Gestural organization in Spanish: An experimental study of spirantization and aspiration*. Storrs, CT: University of Connecticut dissertation.
- Romero, Joaquín, Benjamin Parrell & María Riera. 2007. What distinguishes /p/, /t/, /k/ from /b/, /d/, /g/ in Spanish? Poster presented at Phonetics and Phonology in Iberia, Braga, Portugal.
- Roon, Kevin D., Adamantios I. Gafos, Phil Hoole & Chakir Zeroul. 2007. Influence of articulator and manner on stiffness. *16th International Congress of the Phonetic Sciences*. Saarbrücken, 6–10 August.
- Saltzman, Elliot. 1995. Dynamics and coordinate systems in skilled sensorimotor activity. In Robert F. Port & Tim van Gelder (eds.), *Mind as motion: Dynamics, behavior, and cognition*. Cambridge, MA: MIT Press.
- Saltzman, Elliot & Kevin G. Munhall. 1989. A dynamical approach to gestural patterning in speech production. *Ecological Psychology* 1(4). 333–382.
- Saltzman, Elliot, Hosang Nam, Jelena Krivokapić & Louis Goldstein. 2008. A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. *Speech Prosody 2008*. Campinas, Brazil. 175–184.
- Soler, Antonia & Joaquín Romero. 1999. The role of duration in stop lenition in Spanish. *14th International Congress of the Phonetic Sciences*. San Francisco, CA.
- Studdert-Kennedy, Michael & Louis Goldstein. 2003. Launching language: The gestural origin of discrete infinity. In Morton H. Christiansen & Simon Kirby (eds.), *Language evolution*, 235–254. New York, NY: Oxford University Press, USA.
- Torreblanca, Máximo. 1976. La sonorización de las oclusivas sordas en el habla toledana. *Boletín De La Real Academia Española* 56. 117–145.