Original Paper

Phonetica 2018;75:151–181
DOI: 10.1159/000481099

Received: March 28, 2016
Accepted: August 28, 2017
Published online: February 8, 2018

Explaining Coronal Reduction: Prosodic Structure and Articulatory Posture

Benjamin Parrell[a, b] Shrikanth Narayanan[a]

[a]University of Southern California, Los Angeles, CA, and [b]University of Delaware, Newark, DE, USA

Abstract

Consonant reduction is often treated as an allophonic process at the phonological planning level, with one production target (allophone) being substituted for another. We propose that, alternatively, reduction can be the result of an online process driven by prosodically conditioned durational variability and an invariant production target. We show that this approach can account for patterns of coronal stop (/t/, /d/, and /n/) production in both American English and Spanish. Contrary to effort-driven theories of reduction, we show that reduction does not depend on changes to gestural stiffness. Moreover, we demonstrate how differences between and within a language in the particular articulatory postures used to produce different coronal stops automatically lead to reduction to what have normally been considered distinct allophones – coronal approximants ([ð̞]) and flaps ([ɾ]). In this way, our approach allows us to understand different outcomes of coronal stop reduction as the dynamic interaction of a single process (durationally driven undershoot) and variable spatial targets. We show that these patterns are reflected across a wide variety of languages, and show how alternative outcomes of reduction may fit within the same general framework.

1 Introduction

Consonant reduction processes, sometimes referred to as lenition, have been the focus of a large body of phonetic and phonological research. Generally, reduction refers to the production of a consonant with reduced spatial extent or articulatory constriction degree (Kirchner, 2004). A large body of phonological research has studied consonant reduction as an alternation between 2 (or more) distinct allophones of a single underlying phoneme. A number of phonological explanations have been put

Some of the data presented here have appeared in the Proceedings of the 10th International Seminar on Speech Production.

KARGER

E-Mail karger@karger.com
www.karger.com/pho

Benjamin Parrell
Department of Communication Sciences and Disorders
University of Wisconsin – Madison, 1975 Willow Drive
Madison, WI 53706 (USA)
E-Mail bparrell@wisc.edu

forward to explain this allophonic alternation: reduction has been suggested to be caused by changes in phonological features (such as [±cont]), spreading of features from adjacent segments, feature underspecification, movement along the sonority scale, and decreases in consonant strength (see Kirchner (1998) and Lavoie (2001) for thorough reviews of these proposals). A variant on these phonological theories is that reduction is planned in order "to increase intensity and thereby reduce the extent to which the affected consonant interrupts the stream of speech" (Kingston, 2008, p. 1). In this view, reduction is a way for a speaker to signal the absence of a prosodic break to a listener, providing a possible functional motivation for an allophonic alternation hypothesis.

Another proposal is that reduction results from a planned decrease in articulatory effort, captured in an optimality-theoretic framework using a family of constraints against expending effort in speech (Kirchner, 1998, 2004). The problem with viewing spatial reduction as a change in effort is primarily one of circularity: since there is no clear way to measure metabolic effort in speech movements, it is simply assumed that reduced segments are less effortful (Pouplier, 2012). It is also not clear whether a simple measurement, such as displacement relative to time, is able to capture articulatory effort given the complex structure of the speech motor system (Kingston, 2008; Pouplier, 2012). Most importantly, effort-based constraints still predict a categorical alternation among various allophones, although the number of categories may be larger than that considered in other allophonic accounts of consonant reduction (and the larger the number of categories, the larger the number of needed constraints).

One problem for any analysis of reduction as an allophonic substitution process is accounting for the wide variability of productions seen in speech, which is often relegated to a separate phonetic level or ignored entirely. If reduction is the result of an allophonic substitution rule, there should be some qualitative differences between full and reduced productions. In fact, there is often no clear dividing line between "full" and "reduced" segments, but rather a continuum of possible productions (e.g., Stone and Hamlet, 1982). Moreover, if reduction is an allophonic alternation, it is unclear why it should be heavily conditioned by prosodic structure across languages (Fougeron and Keating, 1997; Lavoie, 2001). Nor is it straightforward to explain why reduction of the same or similar consonants can lead to different outcomes. For example, /d/ is frequently reduced in many languages, but the outcome of this process can be either an approximant or a flap.

Here, we suggest a cause of reduction that can account for (1) the variability seen in production, (2) the common link between prosodic position and consonant reduction, and (3) divergent outcomes of coronal stop reduction between languages. We propose that consonant reduction results from the undershoot of an articulatory target due to prosodically conditioned shortening of movement durations. These articulatory targets are taken to be invariant at the level of phonological control, following the theory of articulatory phonology (Browman and Goldstein, 1992b, 1995a). Note that the idea that consonant reduction is equivalent to a reduction in articulatory movement, on its own, is not new (e.g., Bauer, 2008; Browman and Goldstein, 1992b). Critically, the current proposal suggests why that variation in the extent of articulatory movements may occur.

Our proposal linking articulatory undershoot, movement duration, and prosodic position comes from a large literature showing that prosodic structure has substantial effects on the spatiotemporal realization of speech movements. Gestures produced in the vicinity of a large prosodic boundary, such as an intonational phrase, are both longer in duration and larger in their spatial extent than those produced in the absence of a boundary or near a small boundary, such as a syllable boundary (Byrd et al., 2000; Cho and Keating, 2001; Fougeron and Keating, 1997). This has been shown across a number of languages, though the precise patterns related to each level in the prosodic hierarchy may be language specific (Keating et al., 2003) and may even vary between speakers (Byrd and Saltzman, 1998; Parrell et al., 2013).

When the duration of speech gestures is reduced, their spatial extent is typically also reduced, although the details of this relationship may be modified by factors such as speaking style, rate, or stress (Kelso et al., 1985; Lindblom, 1983, 1990; Moon and Lindblom, 1994; Vatikiotis-Bateson and Kelso, 1993). This reduction can be seen as the simple outcome of a movement that is unable to reach its target (undershoot) when its duration is restricted (Parrell, 2011). Given that productions near small prosodic boundaries show shorter durations than those near larger boundaries, it follows that their spatial extent will also be reduced because of this reduced duration. This has been confirmed by a number of articulatory studies examining the prosodic effects on the spatial magnitude of speech gestures (Byrd et al., 2000; Cho et al., 2011; Fougeron and Keating, 1997). Stress may also modify this relationship, with gestures in unstressed syllables generally showing less movement than those in stressed syllables for the same movement durations (Mooshammer and Fuchs, 2002). Overall, the patterns seen in previous work suggest that productions in "weak" prosodic positions – that is, those near small or nonexistent prosodic boundaries and in unstressed syllables – will be particularly prone to reduction in both duration and magnitude (Lavoie, 2001).

Importantly, our proposal also provides a principled prediction of the outcome of coronal stop reduction: the outcome will be determined by the precise nature of the articulatory target for the stop. Grossly speaking, reduction in the spatial extent of a gesture with a target at the alveolar ridge may still lead to achievement of contact with the palate, though possibly with a smaller amount of linguopalatal contact. This has been shown for alveolar flaps in English (Byrd, 1994; Lavoie, 2001; Saw, 1993). The short duration of this palatal contact would give rise to the perception of a flap, with no time for the build-up of intraoral pressure required to produce a stop burst. On the other hand, reduction in the duration of gestures with dental targets may still lead to contact with the teeth, but not the palate. Data from electropalatographic studies in Spanish where electrodes were placed at the top of the front teeth suggest that even reduced productions of dental /d/ typically contact the front teeth (Hualde et al., 2011; Lavoie, 2001). However, limited contact with only the teeth and not the palate would still allow for air to escape the oral cavity, giving rise to the perception of an approximant, rather than a flap. Note that while this prediction generalizes across two broad categories of alveolar and dental stops, small differences in the exact targets and degree of reduction may alter these expected patterns – for example, increased temporal reduction or a reduced spatial target for alveolar stops may lead to incomplete linguopalatal contact and the production of an approximant, rather than a flap. Such differences are discussed in section 4.1.

In the current study, we test the relationship between spatial reduction, duration, and prosodic structure using American English and Spanish as test cases. Both languages show reduction of intervocalic coronal stops in prosodically weak positions, but the outcomes of this reduction are different in the 2 languages. Spanish reduces

Phonetica 2018;75:151–181
DOI: 10.1159/000481099

intervocalic dental /d/ (and, variably, /t/) to an approximant ([ð̞][1]) in phrase-medial position. American English, on the other hand, reduces alveolar /d/ and /t/ to a voiced flap ([ɾ]) before an unstressed vowel. While reduction in both languages has traditionally been described as a symbolic, allophonic alternation rule, recent experimental work in both languages has questioned this analysis. We briefly review past phonological and phonetic research on stop reduction in both languages before outlining the specific hypotheses tested in the current study.

1.1 English Flapping

For American English, rule-based accounts of coronal flapping state that /t/ and /d/ are replaced by [ɾ] when they occur intervocalically, following a stressed vowel, and before an unstressed vowel. There is a substantial variation in the precise formulation of this rule among different authors, however. Kahn (1976), for example, stresses the importance of the following vowel being unstressed, allowing for flapping to occur between unstressed vowels. In some formulations, the conditioning context for flapping can span a small prosodic boundary (such as a word boundary) but not a larger phrase boundary (Hayes, 1995).

An alternative account of American English flapping is that flapping is a gradient process with productions of [t] and [d] on one end of a continuum and [ɾ] on the other, rather than a categorical rule (though see Warner and Tucker (2011), who show that this continuum extends past flaps to approximants and full acoustic deletion, an issue we discuss in section 4). One of the first studies to propose a gradient rather than categorical explanation of flapping examined the duration of many American English consonants (Umeda, 1977). This study found a range of durations for flaps and variable voicing of /t/ in environments that condition the flapping rule. Zue and Laferrierre (1979), examining the acoustic speech signal, found similar variability in the duration of flaps, as well as identifying 2 different types of productions that fell somewhere between the short flaps and full stops: a short flap-like duration with a burst, and a long consonant duration with no burst. A number of articulatory studies, following on the findings of these 2 papers, have explicitly investigated flapping in American English, looking at the patterns of variability and possible dimensions along which flaps and full stops differ (Fukaya and Byrd, 2005; de Jong, 1998; Stone and Hamlet, 1982; Turk, 1992).

Stone and Hamlet (1982) measured the movements of the jaw and tongue, as well as speech acoustics, during a task requiring reiterant productions of the syllable "da" in an alternating stressed-unstressed pattern (e.g. [da də da də da də]). They found a range of acoustic productions, which they grouped into categories: a voiceless or partially devoiced /d/, a fully voiced /d/, a short /d/ that was either voiced or voiceless, and a flap of variable duration. For the articulatory measures, they found that the more canonical-/d/-like tokens had a higher jaw position and higher acceleration of jaw opening movements out of the consonant as well as more palatal contact (as measured from an electropalatograph). For some participants, 2 syllables were produced with a single jaw movement when the medial /d/ was produced as a flap. This study also found a correlation between jaw height and the amount of linguopalatal contact, with /d/-like tokens

[1] We have chosen to use [ð̞] to represent a nonrhotic coronal approximant as (1) most of the work on Spanish reduction uses [ð] to represent this production and (2) using [ɹ] may create confusion between nonrhotic coronal approximants and true rhotics.

DOI: 10.1159/000481099

produced with both higher jaw position and more palatal contact. These patterns were all gradient, rather than categorical, suggesting a continuum of productions rather than an allophonic alternation.

Turk (1992), based on the result from Zue and Laferriere (1979) that flaps are much shorter than full stops, examined the durational variability of all stops in American English, finding that both labial and velar stops (except for /g/) show durational reduction in the environments where /t/ and /d/ are generally produced as flaps. Turk interprets this result to mean that the shortening of /t/ and /d/ in these environments is not caused by a specific flapping rule but by more general prosodic requirements (though /t/ and /d/ do shorten more than noncoronal stops).

The prosodic conditioning of flapping was investigated in more detail in de Jong (1998). This study used an X-ray microbeam system to measure the movements of the tongue and jaw during production of the words "tote" and "toad" in the phrase *I said, "Put the (target word) on the table,"* where participants placed a nuclear accent on either "put," "on," or the target word. This had the effect of placing the final coronal consonant of the target word in an unstressed, poststress, or prestress position. The results of this study indicate that, while the acoustics differentiate between stops and flaps at least quasi-categorically, there is little articulatory evidence for a categorical distinction. The one exception is a slightly fronted tongue body position during production of full stops compared to flaps. The lack of a clear articulatory distinction leads de Jong to posit that this may be a case where a gradient articulatory change gives rise, through quantal articulatory-to-acoustic relationships, to a more categorical acoustic percept. He goes on to say that, if this is the case, "from the speaker's perspective there is no reason to posit a rule which specifically demands the production of a flap before unstressed vowels. Rather, what is necessary is to understand the language's segmental and prosodic convention sufficiently to know when a salient consonant release is necessary, and when not" (de Jong, 1998, p. 309). That is, the articulatory control system is set up so that prosodically driven variation – which affects all segments (Turk, 1992) – causes variability in the realization of the coronal stops that leads, in some cases, to flapping. The author examines 2 possible causes for the flapping: that they arise from reduction in jaw movement or that they are caused by an increase in overlap between the tongue tip gesture for the stop and the following vowel (de Jong et al., 1993). Because there were no clear differences in jaw position, the first explanation was rejected. There is some support for the second account (lower and more retracted tongue body positions), but the authors point out several problems with this account, suggesting it may be an oversimplification.

The most recent articulatory study to examine flapping in American English looked at word-final /t/ in both phrase-medial and phrase-final position, looking at only the position of the tongue tip as measured via electromagnetic articulometry (Fukaya and Byrd, 2005). As found in previous studies, there was generally an acoustic distinction between flaps and stops, corresponding to stops being relatively long and voiceless and flaps being relatively short and voiced. No such difference was found in the articulation however, where participants produced tokens with gradient spatial and temporal characteristics (although there was a high degree of variability between participants). The authors go on to suggest that the falling-stress environment (stressed vowel-consonant-unstressed vowel) may be particularly conducive to temporal reduction in gesture duration, perhaps leading to phonologization of flapping in this context when it occurs word-internally.

In sum, these studies suggest that, at an articulatory level, flapping is the result of a continuous reduction process rather than a clear alternation between flaps and full stops. Results from acoustic studies suggest a clearer distinction, though some results indicate a continuum even at the acoustic level (Stone and Hamlet, 1982). In any case, it seems likely that possible categorical differences at the acoustic level result from a nonlinear articulatory-to-acoustic mapping rather than reflecting a true difference in production (de Jong, 1998). This is consistent with our hypothesis that spatial reduction of coronal stops in English is the result of prosodically conditioned durational variability rather than an allophonic alternation. The fact that this reduction process often leads to productions of approximants or a minimal acoustic trace (Warner and Tucker, 2011) is also consistent with this hypothesis, but it is difficult to see how these productions would arise from a process where flaps are categorically substituted for full stops.

1.2 Spanish Spirantization

The traditional description of reduction in Spanish states that the voiced stops /b, d, g/ are produced as full stops only in phrase-initial position, when immediately following a nasal, or (for /d/ only) when following /l/. In all other cases, they are produced as voiced approximants, a process known as *spirantization* (e.g., Harris, 1969). While some analyses state that this process results in voiced fricatives, experimental results show clearly that there is no evidence of frication in these productions, and they are best characterized as approximants (Martínez Celdrán, 2008). Many subsequent studies have adopted this allophonic split, with various proposals for the driving phonological cause (e.g., Hualde, 1988; Mascaró, 1984; Piñeros, 2002).

A growing body of experimental work has shown that these allophonic alternation accounts do not accurately reflect the variability of productions found in real speech. By using acoustic methods to analyze the produced constriction degree of voiced stops, a single unimodal distribution is found ranging from full stops to wide approximants (Simonet et al., 2012), though the precise nature of the distribution can differ between dialects (Carrasco et al., 2012). Importantly, a broad range of produced constriction degrees is found in all contexts, in contrast to the predictions of an allophonic alternation account (Hualde et al., 2011). Although there are tendencies towards wider constriction degrees where the allophonic accounts have proposed approximants, there is no evidence for a categorical alternation between full and reduced productions. Although these studies are almost entirely based on acoustics, there is a good relationship between the acoustic measures they have used and the articulatory constriction degree (Parrell, 2010). Moreover, an articulatory study of stop lenition in Spanish similarly found a continuum of constriction degrees rather than clear stop and approximant allophones (Parrell, 2011).

In sum, a simple allophonic alternation process cannot describe the actual constriction degree produced for voiced stops in Spanish. In fact, the actual production of voiced stops has been found to be heavily influenced by a number of suprasegmental factors, including speech rate (Soler and Romero, 1999), stress (Cole et al., 1999; Eddington, 2011; Ortega-Llebaria, 2004), and movement duration (Parrell, 2011). Overall, the lack of evidence for discrete allophonic categories and the influence of suprasegmental factors suggest that reduction of voiced stops in Spanish may be best explained as a dynamic process conditioned by a large number of factors (including both segmental and suprasegmental influences) rather than an allophonic alternation between stops and approximants. This is in agreement with our hypothesis in the current study.

1.3 Comparing American English and Spanish Reduction

In both American English and Spanish, reduction processes that were traditionally analyzed as allophonic alternations have been shown to be much more variable processes. In both cases, prosodically conditioned durational variability has been hypothesized to underlie spatial reduction. The outcomes of this process are quite different in the 2 languages, however. In American English, though spatial and temporal reduction seems to apply to all places of articulation in a similar way, the magnitude and frequency of reduction is larger for coronals than for labials or velars (Browman and Goldstein, 1995b; Turk, 1992), though why there is a difference between these places of articulation remains unclear. Moreover, coronal stop reduction in American English affects both the voiced /d/ and voiceless /t/ equally, and results in an articulation that generally maintains at least midsagittal contact between the tongue and palate. In Spanish, on the other hand, reduction affects all places of articulation with similar articulatory and acoustic outcomes, and occurs much more frequently in the voiced compared to the voiceless stops. Perhaps the most obvious difference between the two languages is that while coronal stops in American English reduce to a flap ([ɾ]), in Spanish they reduce to coronal approximants ([ð̞]).

Reduction in nasal coronal stops (/n/) has received relatively little attention in either language. While not traditionally included in allophonic accounts of reduction, phonetic studies suggest that, at least for Spanish, nasal stops also reduce in the same contexts as oral stops (Honorof, 2003). A few accounts of flapping in English include /n/ (e.g., Trager and Smith, 1951), but the vast majority of both theoretical and experimental work does not. Experimentally, the evidence for reduction of /n/ in English is mixed, with one study finding both temporal and spatial differences as prosodic position varies (Fougeron and Keating, 1997) and one finding no difference (Lavoie, 2001). However, both studies used electropalatography to measure reduction, and this method can only indirectly assess spatial reduction of speech movements. Although electropalatography can measure changes in linguopalatal contact patterns, it provides no information about tongue movements that fail to reach the palate.

Reduction of /n/, if found, follows directly from the proposal here, which views reduction as the consequence of durational reduction, which will affect gestures in the same prosodic context in a similar manner; however, the perceived outcome of this reduction may be affected by other adjacent gestures, such as nasal opening for /n/ (Browman and Goldstein, 1992b). Interestingly, while Spanish and English oral stops differ in their place of articulation (dental in Spanish and alveolar in English), nasal stops in both languages are produced at the alveolar ridge (Byrd, 1994; Kochetov and Colantoni, 2011; Martínez Celdrán et al., 2003). Our hypothesis predicts that reduction in nasal stops should therefore result in a flap in both languages.

We recognize that other differences exist between the 2 languages. For example, English shows substantial vowel reduction in unstressed syllables, while this is typically not found for Spanish. Additionally, English coronal reduction is sensitive to stress, while Spanish reduction is only weakly affected by stress or word boundaries (Cole et al., 1999; Parrell, 2011). While this suggests that the domain of reduction differs between the 2 languages, our focus here is on the variability seen in the outcomes of this process and not the particular prosodic positions that condition this reduction. Additionally, Spanish has a separate phonological rhotic flap, which is not present in English. We address possible influences of this contrast in the Discussion.

1.4 Proposal and Predictions

Our hypothesis is that consonant reduction is a dynamic process driven by reduced movement durations of gestures in prosodically weak positions and that the final outcome depends on the specific invariant articulatory target for that gesture. Our hypothesis makes the following predictions:

1) The magnitude of speech gestures should be conditioned by their duration, and this relationship should hold across both full and reduced productions. We should see similar relationships in both Spanish and English.

2) The relationship between duration and magnitude should hold for movements of all articulators used to create a particular constriction gesture. So, for coronal stops, we should see similar patterns of spatiotemporal variation in tongue tip, tongue body, and jaw movements.

3) Reduction should be seen even in the absence of changes in movement stiffness, which have been suggested to reflect differences in articulatory effort.

4) The final outcome of reduction should be determined by the location of the constriction target. Specifically, reduction of alveolar targets will lead to contact between the tongue and palate (flaps) while reduction of dental targets may lead to contact between the tongue and the teeth but not between the tongue and the palate (approximants). This predicts that not only will English and Spanish reduction outcomes differ, but that Spanish will show a different type of reduction for /n/ (alveolar) than for /d/ or /t/ (dental).

We test these predictions using real-time magnetic resonance imaging (MRI) to examine the spatial and temporal characteristics of coronal reduction in English and Spanish. Additionally, we present results from a large database of languages with reduction (Kirchner, 2004) to show how these patterns may apply more broadly (see section 4.1).

2 Methods

2.1 Participants and Stimuli

Four participants participated in the current study. Two were native speakers of General American English. Two were native speakers of Iberian Spanish (spoken in the central and northern part of Spain). Both Spanish speakers were current residents of the USA and had lived in Spain through completing college. No participant reported any history of speech or hearing impairment.

Stimuli were designed to elicit coronal oral and nasal stops (/t/, /d/, and /n/) in a symmetric or near-symmetric low vowel context. Low vowels were chosen to maximize the amount of movement between the vowels and intervocalic stops. The prosodic position of the consonant was varied to elicit a range of productions including both full and reduced forms.

For American English, prosodic conditions included the stop in noninitial, word-initial, and phrase-initial positions. Both flanking vowels for the word- and phrase-initial conditions were /ɔ/ (the 2 American English speakers in this study consistently differentiated between /ɔ/ and /ɑ/). For the noninitial condition, it was not possible to use the same vowels – a falling stress pattern (and often reduced second vowel) is the conditioning factor for word-internal flapping. For this condition, the vowel context was [ɑCɑ] or [ɑCə], which was chosen both to give a fairly close match to the vowels in the rest of the stimuli and to limit coarticulatory tongue movement between the full and reduced vowels. In order to induce more variability in production, stimuli in the noninitial condition used contrasting emphatic stress. In these stimuli, the target consonant appeared in the coda of a monosyllabic word with the vowel /ɑ/ (*pot, pod, swan*). The words appeared in the middle of the carrier phrase “Put the X on the table” and emphatic stress (shown in the stimuli with capital letters) was placed on either *put*, the target word, or *on*. This effectively places the target coronal consonant in unstressed, poststress, or

DOI: 10.1159/000481099

Table 1. American English stimuli used in the coronal reduction study

Category	/t/	/d/	/n/
Noninitial	I said put the po**t** ON the table.	I said put the po**d** ON the table.	I said put the swa**n** ON the table.
Noninitial	I said put the PO**T** on the table.	I said put the PO**D** on the table.	I said put the SWA**N** on the table.
Noninitial	I said PUT the po**t** on the table.	I said PUT the po**d** on the table.	I said PUT the swa**n** on the table.
Noninitial	He didn't say "bo**tt**om" any more.	He didn't say "ba**d**a bing" any more.	He didn't say "Tiajua**n**a" any more.
Word-initial	He didn't awe **T**awny anymore.	He didn't awe **D**awnie anymore.	He didn't awe **N**aughty anymore.
Phrase-initial	He didn't awe. **T**awny did.	He didn't awe. **D**awnie did.	He didn't awe. **N**aughty did.

The target consonants are all 3 coronal oral and nasal stops (/t/, /d/, and /n/). A variety of prosodic contexts are used to facilitate spatiotemporal production variability. In the first 3 carrier phrases, the location of the emphatic stress is changed to generate prestress (ON), poststress (target word stressed), or unstressed (PUT) conditions. The other 3 stimuli present the target consonant in noninitial, word-initial, and phrase-initial positions.

Table 2. Spanish stimuli used in the coronal reduction study

Category	/t/	/d/	/n/
Noninitial	Ella dice "má**t**a" también.	Ella dice "ná**d**a" también.	Ella dice "gá**n**a" también.
Noninitial	Ella dice "ma**t**ámos" también.	Ella dice "na**d**ámos" también.	Ella dice "ga**n**ámos" también.
Noninitial	Ella dice "MA**T**ÁMOS" también.	Ella dice "NA**D**ÁMOS" también.	Ella dice "GA**N**ÁMOS" también.
Word-initial	Ella dice "copa," "**t**ápa," y "mesa."	Ella dice "tipa," "**d**áma," y "mujer."	Ella dice "mapa," "**n**áve," y "faro."
Phrase-initial	Ella dice "mapa." **T**ambién lo digo.	Ella tiene una capa. **D**ámela.	Ella no dice "tapa." **N**ádie lo dice.

The target consonants are all 3 coronal oral and nasal stops (/t/, /d/, and /n/). A variety of prosodic contexts are used to facilitate spatiotemporal production variability. The target consonant appears in noninitial position in both lexically unstressed and lexically stressed syllables, as well as in a lexically stressed syllable that receives emphatic stress. The target consonant also appears word-initially in a list and at the beginning of a sentence to induce different prosodic boundary strengths. Lexical stress on the target words is marked with an accent (´) for illustrative purposes, even when not normally used in Spanish orthography.

prestress positions, and such stimuli have been shown to cause a relatively large amount of spatiotemporal variation in production (de Jong, 1998).

For Spanish, /a/ was used as the target vowel both preceding and following the target consonant for all stimuli. As for English, these stops occurred in noninitial, word-initial, and phrase-initial conditions. A slight modification of the word-initial condition was necessary as this position in Spanish, when it occurs within a prosodic phrase, does not differ from the noninitial condition in that language (e.g., Cole et al., 1999; Parrell, 2011). This was addressed by placing the word-initial target in a list, which was designed to induce a smaller prosodic boundary than that which would occur in sentence-initial position (cf. Byrd, 2000; Parrell et al., 2013). For clearer comparisons with the English data, we will continue to use word-initial to indicate this category. In the noninitial condition, stress variation was included both at the lexical level (with the target consonant in onset position of stressed and unstressed vowels) as well as at the emphatic level, with utterances without emphatic stress contrasted with emphatic stress on the word containing the target consonant in a lexically stressed syllable, again cued visually by capital letters. As for English, this variation in stress was included to increase the spatial and temporal variability in the data set. Data for the Spanish tapped rhotic /ɾ/ was collected concurrently with the other coronal segments for Spanish but is not included in the current analysis.

A full list of the stimuli used for both languages appears in Table 1 (English) and Table 2 (Spanish). For each language there were a total of 18 stimuli, which were randomized into 2 blocks of 9. Blocks were presented in an alternating fashion for a total of 6–8 repetitions per target phrase, giving 108–144 tokens per speaker (6–8 repetitions × 18 stimuli).

Phonetica 2018;75:151–181
DOI: 10.1159/000481099

2.2 Real-Time MRI Data Collection

Data were acquired using an MRI protocol developed especially for research on speech production, detailed in Narayanan et al. (2004). Participants were supine during the scan with the head restrained in a fixed position to facilitate comparisons across acquisitions. For the English data, a 13-interleaf spiral gradient echo pulse sequence was used (TR = 6.164 ms, field of view = 200 × 200 mm, flip angle = 15°). For the Spanish data, a 9-interleaf spiral sequence was used (TR = 6.028 ms, field of view = 200 × 200 mm, flip angle = 15°). For both sequences, a 5-mm slice located at the midsagittal plane of the vocal tract was scanned with a resolution of 68 × 68 pixels, giving a spatial resolution of approximately 2.9 mm per pixel. Videos were reconstructed with a 13-frame (for English) or 9-frame (for Spanish) sliding window, with 1 frame reconstructed at every repetition time pulse. This gives an effective frame rate of 162.2 frames/s (13-interleaf sequence) or 165.9 frames/s (9-interleaf sequence). Similar techniques for high frame rate reconstruction are described in Proctor et al. (2015). Synchronous noise-cancelled audio sequences were collected at 20 Hz during MRI acquisition (Bresch et al., 2006).

2.3 MRI Data Analysis

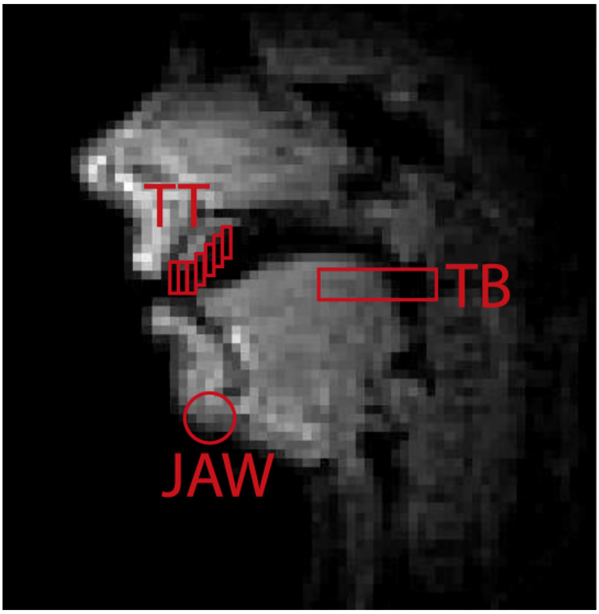
All measurements of speech articulator motion were extracted from the MRI images by means of pixel intensity values (Hagedorn et al., 2011; Lammert et al., 2010; Proctor et al., 2011). This method is based on the idea that the changes in pixel intensity of a particular pixel over time reflect changes in tissue density at that point in the vocal tract. Lower intensities correspond to the absence of tissue (air) while high values signify the presence of one of the speech articulators at that particular point. In any given arbitrary region of the vocal tract, then, the average pixel intensity in that region will reflect the proportion of the region occupied by the speech articulators. By placing these regions at relevant locations in the vocal tract and measuring the average intensity over time, we are able to estimate speech articulator motion in that region. Because the shape of the vocal tract will vary considerably between participants, these relevant regions (described below) were defined on a by-participant basis relative to each participant's anatomy. Each region was defined such that the relevant speech articulators (tongue tip, tongue body, jaw) were always present in the region, avoiding any floor effects that could be caused by the complete absence of the articulator from the region.

For the current study, we are interested particularly in the forward motion of the tongue body during the transition from vowel to coronal consonant, the motion of the tongue tip towards the palate, and the raising of the jaw. Tongue body movement was measured by defining a long, horizontal region in the pharyngeal area of the vocal tract. This region has a vertical span from the top of the epiglottis to the bottom edge of the velum at its lowest position. Defining the region in this way focuses the measurement on only movement of the tongue body, without interference from the presence or absence of these other structures. The tongue body region spanned horizontally from the rear pharyngeal wall to a point roughly in the middle of the hard palate, including for each participant 1 pixel of the pharyngeal wall and 2–3 pixels of the tongue during production of /i/ (the most forward position of the tongue in the data set). Because the pixel values in the pharyngeal region were found to vary substantially from sample to sample, the mean pixel intensity in the tongue body region was normalized by the mean pixel intensity in the entire image on a frame-by-frame basis.

Jaw movement was measured with a circle with a radius of 2 pixels that was placed at the base of the jaw between the jaw inflection point and the hyoid bone. The circle was placed such that when the jaw was closed, some part of the jaw was still in the circle, and that when the jaw was maximally open, the circle was not entirely filled by the jaw. This avoids possible saturation effects that might limit the accuracy of the measurement at extremes of jaw position. The precise location of the circle was manually determined for each participant. For examples of these region-of-interest locations, see Figure 1.

Tongue tip movement was measured in a slightly different way. The tongue tip can contact a large number of places along the width of the palate. It would be plausible, then, to use a large region covering the entire length of the palate. However, such a region would give very different average intensity values when the tongue tip contacts the palate at a particular point compared to when contact is made with the tip and blade along a wide portion of the palate. In order to measure the movement of the tongue tip and contact with the palate more accurately, a set of smaller regions was used. Each region had a horizontal width of only 1 pixel, with a vertical span of 4 pixels beginning at the palate. These regions were arranged in a horizontal array beginning just posterior to the teeth, past the alveolar ridge, to the end of the hard palate. For each participant, the one of these

DOI: 10.1159/000481099



Color version available online

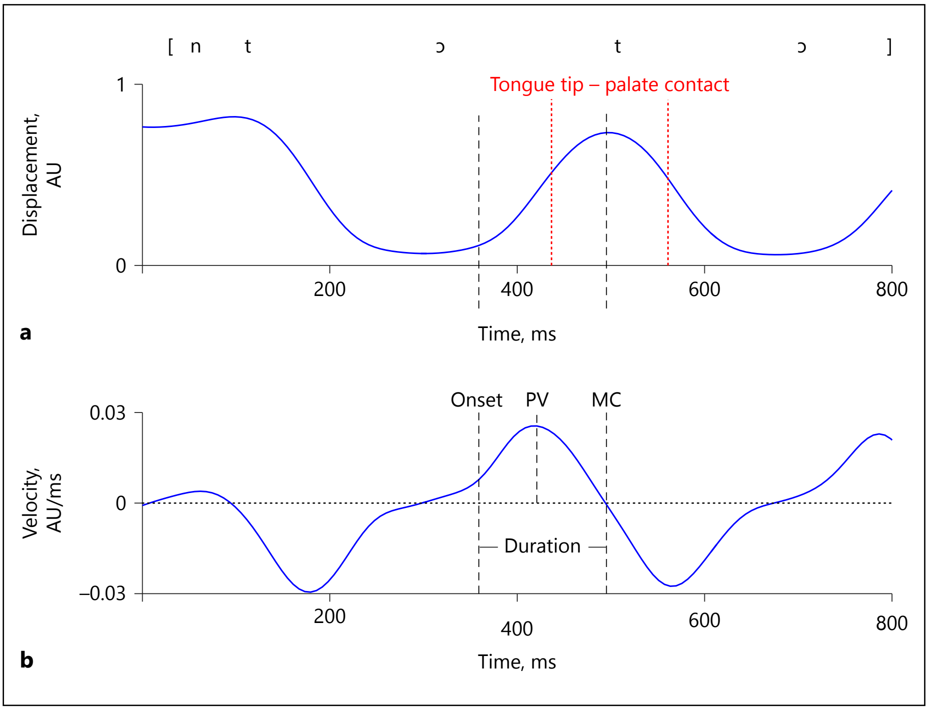
Fig. 1. Representative region of interest locations for 1 American English participant measuring jaw, tongue body (TB), and tongue tip (TT) movement.

regions with the highest maximum pixel intensity during production of each target consonant was chosen to index tongue tip movement. Each speaker was highly consistent in the location of the produced constrictions. For English, both speakers produced all consonants at the same point, near the inflection of the alveolar ridge. For Spanish, however, each consonant was produced at a slightly different location. Each participant was consistent within each segment, however, and the locations were similar between participants. /d/ was measured at the first point on the palate, immediately posterior to the upper teeth; /t/ was measured at a point slightly behind the teeth; /n/ was measured at the alveolar ridge.

In order to be able to compare the final outcome of reduction across languages and across segments, both the location and width of linguopalatal contact were measured. Tongue contact with the palate was measured by placing a set of 1 × 1 pixel regions along the length of the palate. This gives an increase in intensity only when the tongue is in that exact pixel (i.e., when the tongue is contacting the palate). A pixel intensity threshold was used to determine the presence of tongue-palate contact for each pixel. This was done by initially finding the maximum and minimum pixel intensity values across all tongue contact regions across all repetitions for each speaker. Subsequently the threshold for measuring contact was set at 25% of this range plus the minimum value found. This threshold was found to adequately measure contact when compared manually against the MR images. For each repetition, the width of palatal constriction was measured as the number of pixel regions above threshold during the point of maximum tongue tip movement (described above). Width here refers to the horizontal extent of contact in the mid-sagittal plane. A width of 0 indicates that there was no contact between the tongue and palate. The location of the constriction was identified as the pixel along the palate with the highest intensity value at the peak of tongue tip movement. Note that this technique only measures contact between the tongue and palate. Because the teeth do not appear on MR images, it is not possible to measure any contact that may occur between the tongue and upper teeth.

All resulting signals (tongue tip, tongue body, jaw, tongue-palate contact) were smoothed using locally weighted linear regression (Lammert et al., 2010; Proctor et al., 2011). The weighting function used was a Gaussian kernel with a standard deviation of h samples, where $h = 4$. As samples lying more than $3\ h$ from the center of the kernel in either direction receive weights near zero, this gives a smoothing window width of roughly 150 ms given the sampling period of 6.164 ms (13-interleaf sequence) or 6.028 ms (9-interleaf sequence).

Phonetica 2018;75:151–181
DOI: 10.1159/000481099



Color version available online

Fig. 2. Example tongue tip trajectory from a participant producing [ntɔtɔ] from the phrase "didn't awe Tawny": displacement (**a**) and velocity (**b**). Units are in arbitrary units (AU) derived from the pixel intensity signal. Time points for measure gesture onset (onset), time of peak velocity (PV) and time of maximum constriction (MC), as well as the movement duration, are shown. Note the smooth displacement profile both before and after linguopalatal contact (marked in red dashed vertical lines and measured by hand from the real-time MR images) during the [t] in "Tawny."

Gestural identification was conducted using a velocity-based algorithm. Critical time points are shown in Figure 2. The identification algorithm used takes as input a manually located estimate of the gestural midpoint of 1 derived variable (here pixel intensity contours). Using the velocity of that 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 velocity minimum (measurement point: time of peak velocity). Where the preceding velocity minimum incorrectly indexed the start of articulator motion, a manual estimate of motion onset was used as the input to the algorithm, and the nearest velocity minimum to that point was selected. When these corrections were necessary, motion onset was estimated as the inflection point in the signal when the value began to rise towards the coronal construction during or immediately after the preceding vowel. Manual estimation of movement onset was often necessary for phrase-initial repetitions as these productions often had multiple velocity peaks.

Once the velocity minimum close to gestural onset was identified, the algorithm then locates the onset of gestural motion by locating a point where the velocity signal from the preceding minimum to the time of peak velocity crosses some arbitrary threshold of the velocity difference between the 2 points. This threshold was set to 20%. Movement duration was calculated as the time between gesture onset and the time of maximum constriction.

After measurement, tongue tip, tongue body, and jaw movements were normalized to a range from 0 to 1 for each speaker by subtracting the minimum value found for each region during the vowels preceding and following the target consonants across all repetitions, then dividing by the total range of each measure across all repetitions. Maximum displacement was calculated as the difference in normalized intensity between gesture onset and the time of maximum constriction. Similar methods have been shown to be highly correlated with displacements hand-measured in pixels (Lammert et al., 2013).

While the tongue body and jaw measurements straightforwardly measure articulatory movements, the tongue tip measure is a bit different. When the tongue tip is not in contact with the palate, this measure indexes articulatory movements in the same way tongue body and jaw are measured. However, when the tongue tip is in contact with the palate, it is not clear whether the measurement indexes movement in the same way – it may measure movement of additional tissue into the region of interest rather than spatial movement of the tongue tip (which has contacted the palate and so cannot move any further). This may be due to the incompressible nature of the tongue, which leads to spatial deformation of the tongue as it moves against the hard structures of the mouth (Baker, 2008; Kier and Smith, 1985; Perrier et al., 2003; Stone and Lundberg, 1996). In a sense, however, both spatial movement and deformation are the result of the same action, with more tissue moving into a particular space in the vocal tract. In fact, the tongue tip measure appears to respond similarly to both unrestricted motion and deformation-based motion. When the tongue tip trajectories are examined there is no noticeable deflection at the time when the tongue contacts the palate, such as would be expected if the tongue tip measure were indexing only spatial movement and not deformation (Fig. 2).

2.4 Statistical Models

Our goal is to test whether variability between different productions can be explained by durational variation. To this end, we constructed a nonlinear mixed-effects model to predict maximum displacement from movement duration. Because the effects of duration on maximum displacement show a saturation effect (Fig. 3), the relationship between duration and maximum displacement was modeled as a logistic rather than linear relationship. The SSlogis function in R was used for this purpose, where

$$\text{maximum displacement} = \frac{asymptote}{1 + e^{\left(\frac{xmid - duration}{scale}\right)}},$$

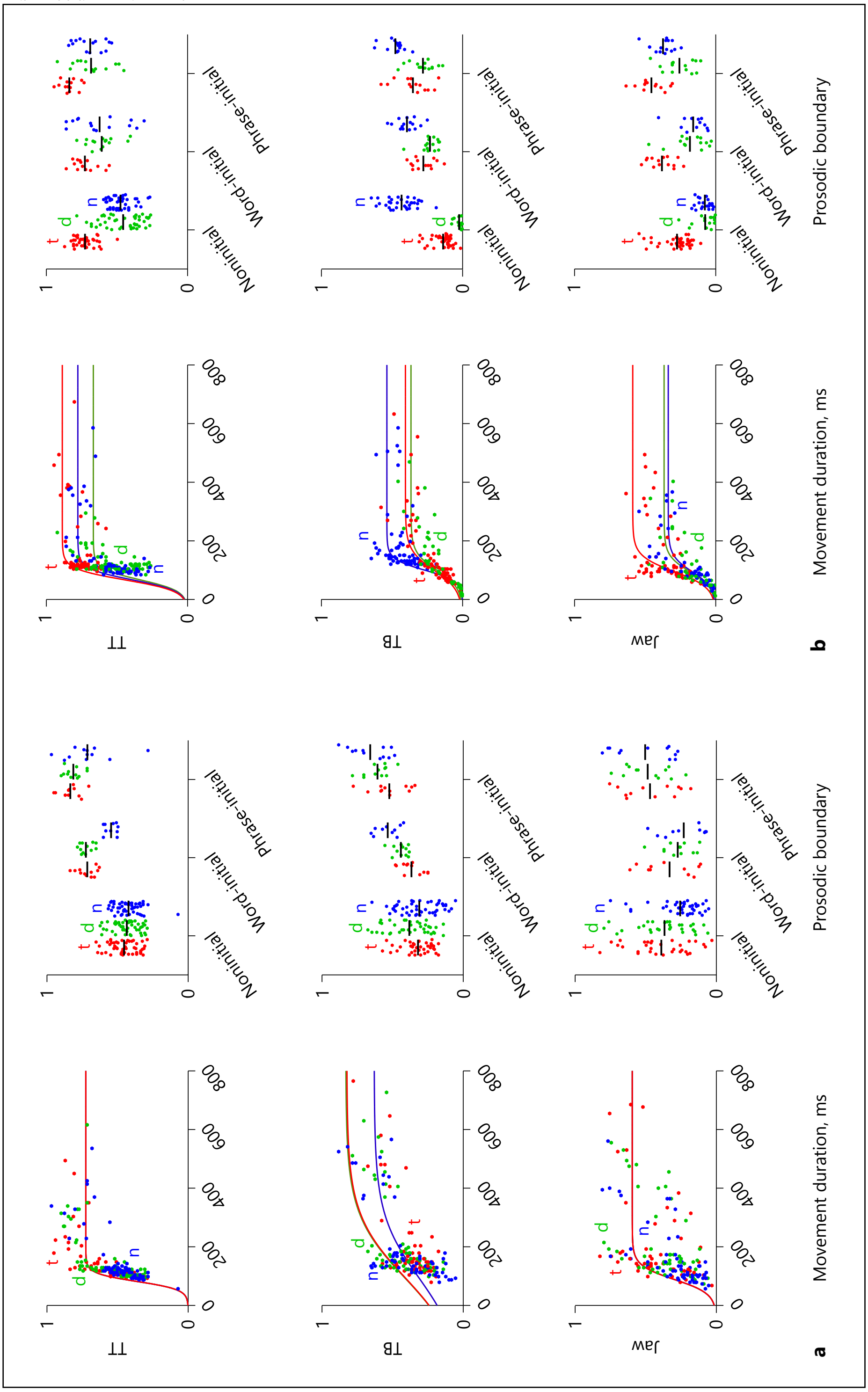
and *asymptote* is a numeric value representing the final asymptotic value at long duration, *xmid* represents the duration at the inflection point of the curve, and *scale* is a positive number reflecting the steepness of the curve. This model additionally included fixed effects of segment on *asymptote* and *scale* as well as random effects of both participant and token on *asymptote*. Models with both random effects on *asymptote* and *scale* in general did not converge. All models were built and assessed using the nlme package in R (Pinheiro et al., 2016). For each model, statistical significance of each predictor was assessed using the results of the p values given by the anova() function in R. Values for pairwise comparisons were generated from the summary() function in the nlme package. A full list of models used for all statistical analyses is shown in Table 3.

A separate model was constructed in order to test whether changes in maximum displacement are due to changes in gestural stiffness. Stiffness was determined by the slope of the regression line relating peak velocity to maximum displacement. Gestures with the same stiffness are expected to fall along the same peak velocity/maximum displacement line (though see Fuchs et al. (2011) for a counterargument against estimating stiffness as a dynamical control parameter from movement kinematics). If reduction were due to a change in stiffness, we would expect that reduced forms would

Fig. 3. Plot of maximum displacement results (normalized displacement) for English (**a**) and Spanish (**b**). For each language, the relationship between displacement and duration is shown in the left column, and that between displacement and prosodic position is shown in the right column. Movements of all 3 articulators are shown separately in each row. From top to bottom: tongue tip (TT), tongue body (TB), jaw. Tokens are shown separately by color: /t/ in red, /d/ in green, and /n/ in blue. Regression lines are shown for the displacement-duration plots and means for the displacement-prosodic position plots. *(For figure see next page.)*

3

Color version available online



DOI: 10.1159/000481099

Table 3. Nonlinear mixed-effects models used to evaluate data

Variable	Model
Displacement	Max. displacement ~ SSlogis(Duration, Asym, xmid, scal), fixed = list(Asym + scal ~ Segment, xmid ~ 1), random = Asym ~ 1 | subject/token
Contact width	Contact width ~ SSlogis(Duration, Asym, xmid, scal), fixed = list(Asym ~ Segment, xmid + scal ~ 1), random = Asym ~ 1 | subject
Contact location	Contact location ~ SSlogis(Duration, Asym, xmid, scal), fixed = list(Asym ~ Segment, xmid + scal ~ 1), random = Asym ~ 1 | subject
Peak velocity	Peak velocity ~ max. displacement + max. displacement:prosodic boundary + max. displacement:segment, random = ~1 | subject/token

Models fit movement duration as a logistic relationship to maximum displacement.

have a lower slope for the relationship between peak velocity and maximum displacement. The statistical model includes peak velocity as the dependent variable and maximum displacement, as well as the interaction between (1) maximum displacement and prosodic boundary and (2) maximum displacement and segment, as fixed factors. Models also included random intercepts for participants and tokens. The relevant comparison will be the interaction between displacement and prosodic boundary, which will reflect differences in the velocity/displacement slope by prosodic position. An interaction between displacement and segment would similarly reflect differences in stiffness by segment.

To compare the final outcome of reduction across languages and across segments, models were constructed to compare both the location and width of linguopalatal contact. These models predicted the location or width of contact based on movement duration (fit as a logistic function) with a fixed effect of segment on *asymptote*. The models additionally included a random effect of subject on *asymptote*. Models with random effects of token on *asymptote* and/or fixed effects of segment on *scale*, which were included in models on maximum displacement, did not converge. For this reason, these factors were left out of the models evaluating constriction location and width.

3 Results

Results will be presented based on the predictions listed in the Introduction. These are:

1) The magnitude of speech gestures should be conditioned by their duration.
2) All articulators should show similar reduction patterns.
3) Reduction should be seen even in the absence of changes in movement stiffness.
4) The outcome of reduction will be determined by the target place of articulation.

Our principal hypothesis is that the magnitude of speech gestures, as indicated by their maximum displacement, should be conditioned by the duration of those gestures. This relationship was found to be significant for both Spanish and English (see Table 4 for detailed results). This relationship can be seen in the left column in Figure 3a (English) and Figure 3b (Spanish). This figure also shows that this relationship is asymptotic, such that there is little to no change in maximum displacement at durations longer than approximately 200 ms. This relationship was captured in our models using a logistic, rather than linear, fit. Detailed results on the significant effects for each model, discussed below, can be seen in Table 4.

Table 4. Summary of statistical results for models for English and Spanish

	Articulator	Parameter	*F* value	*p* value
English				
	Tongue tip	asymptote	130.0	<0.0001
		asymptote ~ seg.	0.01	0.98
		scale	666.6	<0.0001
		scale ~ seg.	2.9	0.06
		xmid	330.1	<0.0001
	Tongue body	asymptote	29.0	<0.0001
		asymptote ~ seg.	7.8	<0.001
		scale	187.5	<0.0001
		scale ~ seg.	2.1	0.14
		xmid	358.2	<0.001
	Jaw	asymptote	37.7	<0.0001
		asymptote ~ seg.	1.7	0.19
		scale	11.0	<0.01
		scale ~ seg.	1.2	0.31
		xmid	209.7	<0.0001
Spanish				
	Tongue tip	asymptote	455.9	<0.0001
		asymptote ~ seg.	16.9	<0.0001
		scale	261.8	<0.0001
		scale ~ seg.	1.2	0.30
		xmid	52.6	<0.0001
	Tongue body	asymptote	72.8	<0.0001
		asymptote ~ seg.	11.2	<0.0001
		scale	110.3	<0.0001
		scale ~ seg.	110.2	<0.0001
		xmid	459.5	<0.0001
	Jaw	asymptote	1.5	0.21
		asymptote ~ seg.	27.2	<0.0001
		scale	8.4	<0.01
		scale ~ seg.	2.7	0.07
		xmid	169.3	<0.0001

All models showed significant effects of *asymptote* and *scale*, suggesting that changes in movement duration were related to changes in displacement. Segment (seg.) did not affect the *scale* parameter, suggesting a similar relationship between duration and displacement for all segments. For some models, the final asymptote did vary by segment. See text for details.

Importantly, the relationship between the duration and spatial extent of speech movements was found to be similar across all articulators for both languages. All segments generally patterned in the same way, with smaller movements associated with shorter durations, as indicated by the significant effect of the scaling factor representing the slope of the logistic function (a flat function, indicative of no relationship, would have an insignificant value for the slope of the fit). An effect of segment on the slope of the logistic curve was found only for the tongue body in Spanish, though post hoc tests showed no significant pairwise contrasts. No significant effects of segment on the slope were found for English. These results suggest that the relationship between duration and magnitude is generally similar across segments in both languages.

When there was an effect of segment, it was generally found on the final asymptote values only, indicating that different segments show differences in their maximal movement extent for certain articulators. The difference in final asymptote value was seen only for the tongue body in English, where the asymptote value for /n/ was less than for /t/ or /d/ (/n/–/d/: $t = 2.8$, $p < 0.01$; /n/–/t/: $t = 2.9$, $p < 0.01$). The overall lack of segmental effects in English indicates that the spatial extent of movements is generally similar for all 3 coronal stops.

For Spanish, there was consistently an effect of segment on the final asymptote values, though which segments showed larger movements differed between the articulators. Overall, this suggests that the coronal stops differ from one another in Spanish in their maximal movement extent, unlike for English. For the tongue tip, /t/ had a higher asymptote than /d/, and /n/ did not differ from either segment (/t/–/d/: $t = 3.6$, $p < 0.001$). For the tongue body, /n/ had a higher asymptote than either /t/ or /d/ (/n/–/d/: $t = 4.3$, $p < 0.0001$; /n/–/t/: $t = 3.6$, $p < 0.001$). For the jaw, /t/ showed a higher final value than either /n/ or /d/ (/t/–/d/: $t = 3.9$, $p < 0.001$; /t/–/n/: $t = 4.8$, $p < 0.0001$). For all articulators, /d/ showed the least movement, as indicated by the lowest final asymptote values.

Productions sorted by prosodic context, rather than by duration, are shown in the right columns of Figure 3. While we do not directly compare these 2 analyses, note that there is consistently a number of productions in the no-boundary context that fall within the range seen in the word boundary contexts. This is true for all segments in both English and Spanish, across all articulators. Overlap in distributions of this type would be unexpected if there were a categorical alternation between word boundary and no-boundary contexts.

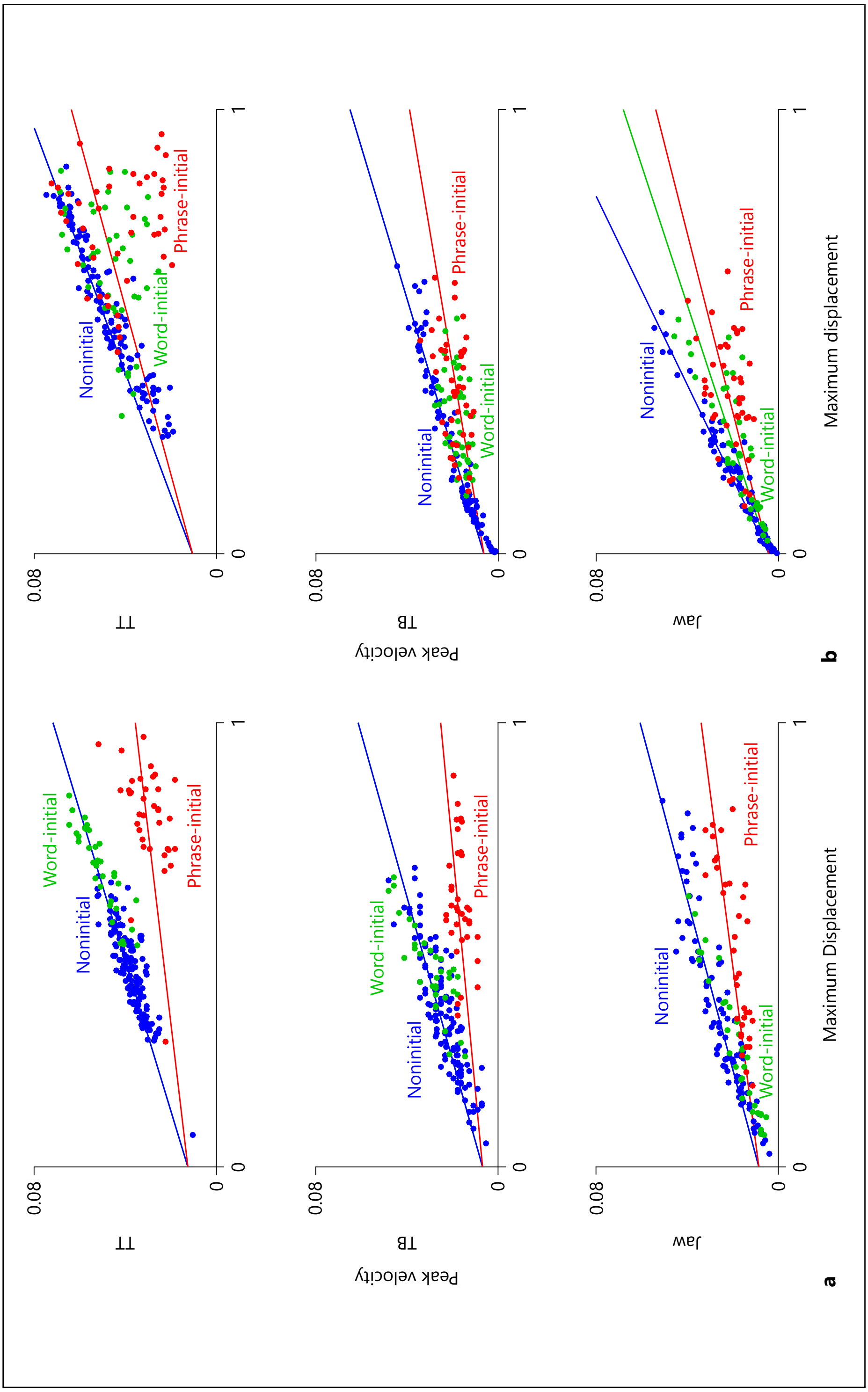
In terms of the relationship between maximum displacement and peak velocity, which we are using as an index of gestural stiffness, there exist clear differences in stiffness based on prosodic position for movements of all 3 articulators in both English and Spanish. This can be seen in the differences in the slope of the relationship between peak velocity and maximum displacement in Figure 4. Importantly, however, productions with small maximum displacements fall along the same regression line as those with larger displacements. If small displacements were due to a lowering of stiffness due to constraints on effort, we would expect productions with small displacement to have a shallower slope in their relationship to peak velocity. If anything, lower stiffness is associated with slightly *more displacement* as stiffness is lower for productions in phrase-initial position, consistent with results from many past studies (Beckman and Edwards, 1992; Byrd et al., 2000; Cho, 2006).

Fig. 4. Plots of peak velocity by maximum displacement for English (**a**) and Spanish (**b**). Peak velocity is measured in arbitrary units (i.e., displacement normalized from 0 to 1) per millisecond. Each articulator is shown separately. From top to bottom: tongue tip (TT), tongue body (TB), and jaw movements. Prosodic positions are separated by color: noninitial in blue, word-initial in green, and phrase-initial in red. All 3 articulators in both languages show a stiffness (slope of the regression line) difference between phrase-initial and noninitial productions, in line with previous work showing a lower stiffness in the former position. For English, word-initial productions have an identical slope to noninitial productions. For Spanish, word-initial productions are identical to phrase-initial productions except for jaw movements, where they have a slope in between noninitial and phrase-initial productions.

(*For figure see next page.*)

Phonetica 2018;75:151–181
DOI: 10.1159/000481099

Color version available online



All models showed a significant difference in the slope of the relationship between noninitial and phrase-initial conditions. In English, the word-initial condition consistently patterned with the noninitial condition, while in Spanish the word-initial condition generally patterned with the phrase-initial condition. The one exception to this is in the case of jaw movements for Spanish, where the 3 conditions were distinct from one another. Although there were significant differences between prosodic categories for Spanish, visual inspection of Figure 4 shows substantial overlap between all 3 prosodic conditions. This was generally not the case for English, where a clearer separation exists between phrase-initial and word-initial/noninitial conditions. In general there were no effects of segment on movement stiffness. The one exception to this pattern is for English tongue tip movements, where /t/ had a slightly higher slope (i.e. higher stiffness) than /d/ or /n/. Detailed statistical results are presented in Table 5.

Impressionistically, our results on the location of the coronal constriction agree with previous analysis of both English and Spanish (representative examples of all segments for each speaker are shown in Fig. 5). For English, all 3 coronal segments are produced at the alveolar ridge for both speakers (mean 3.2–3.3 pixels behind the teeth). For Spanish, /d/ is produced at the teeth (mean 1.3 pixels), /t/ is produced at a large area spanning the teeth and most anterior section of the palate (mean 1.6 pixels), and /n/ is produced at the alveolar ridge (mean 3.9 pixels). These patterns are consistent for both speakers. For all speakers, the place of articulation is consistent across prosodic conditions. These impressionistic results suggest that, as predicted, the outcome of coronal stop reduction is determined by the precise place of articulation: a flap for all segments produced at the alveolar ridge (all segments for English and Spanish /n/) and an approximant for those produced at the teeth (Spanish /d/).

The other notable impressionistic result shown in Figure 5 is that the width of linguopalatal contact in the sagittal plane generally increases from the no-boundary condition to the word-initial condition and again from the word-initial to the phrase-initial condition. This is found for both speakers across all segments in English, and for both speakers for Spanish /n/. Spanish /t/ shows a wide contact area regardless of the prosodic position for both speakers. Spanish /d/ does not show much contact with the palate for either speaker at any prosodic position in Figure 5 but does show a decreasing distance between the tongue and palate in word- and phrase-initial conditions, suggesting increased contact with the front teeth. In these examples, speaker SP1 achieves full linguopalatal contact for /d/ in phrase-initial position, while speaker SP2 shows clear contact with the teeth, as indicated by the small dark divot in the superior surface of the anterior portion of the tongue, which is indicative of tongue displacement caused by the front incisors. This is also seen, to a lesser extent, in the word-initial condition.

In fact, 50% of Spanish /d/ productions in noninitial position showed no contact with the palate, though Figure 5 suggests that at least some of these productions may still maintain contact with the teeth. This is in contrast to /t/ and /n/ in Spanish, which show no productions without linguopalatal contact. In English, roughly 10% of noninitial tokens fail to achieve palatal contact across all segments (/n/: 6%, /d/: 9%, /t/: 12%). In word-initial and phrase-initial conditions, all segments in both languages consistently achieve full closure. The one exception to this is Spanish /d/, where 7% of productions in both word-initial and phrase-initial position do not achieve palatal contact.

The quantitative results from the statistical models agree with the qualitative results seen in Figure 5. For constriction location, neither English nor Spanish showed

Phonetica 2018;75:151–181
DOI: 10.1159/000481099

Table 5. Summary of statistical results for models relating peak velocity to maximum displacement

	Articulator	Parameter	*F* value	*p* value	Post hoc comparisons
English					
	Tongue tip	max. displacement	103.3	<0.0001	–
		max. displacement: boundary	230.6	<0.0001	nb-pb: $t = 16.6$, $p < 0.0001$
					nb-wb: $t = 0.2$, $p = 0.86$
					wb-pb: $t = 19.4$, $p < 0.0001$
		max. displacement: segment	3.9	<0.05	t-d: $t = 2.2$, $p < 0.05$
					t-n: $t = 2.5$, $p < 0.05$
					d-n: $t = 0.5$, $p = 0.62$
	Tongue body	max. displacement	114.9	<0.0001	–
		max. displacement: boundary	130.7	<0.0001	nb-pb: $t = 14.8$, $p < 0.0001$
					nb-wb: $t = 0.13$, $p = 0.90$
					wb-pb: $t = 12.8$, $p < 0.0001$
		max. displacement: segment	0.3	0.73	–
	Jaw	max. displacement	437.5	<0.0001	–
		max. displacement: boundary	62.7	<0.0001	nb-pb: $t = 10.9$, $p < 0.0001$
					nb-wb: $t = 0.5$, $p = 0.59$
					wb-pb: $t = 6.8$, $p < 0.001$
		max. displacement: segment	0.9	0.41	–
Spanish					
	Tongue tip	max. displacement	173.0	<0.0001	–
		max. displacement: boundary	14.1	<0.0001	nb-pb: $t = 4.8$, $p < 0.0001$
					nb-wb: $t = 3.7$, $p = < 0.001$
					wb-pb: $t = 0.8$, $p = 0.43$
		max. displacement: segment	0.4	0.66	–
	Tongue body	max. displacement	298.7	<0.0001	–
		max. displacement: boundary	29.6	<0.0001	nb-pb: $t = 6.8$, $p < 0.0001$
					nb-wb: $t = 5.1$, $\mathrm{p} \leq 0.0001$
					wb-pb: $t = 0.8$, $p = 0.43$
		max. displacement: segment	0.1	0.86	–
	Jaw	max. displacement	435.9	<0.0001	–
		max. displacement: boundary	38.9	<0.0001	nb-pb: $t = 8.2$, $p < 0.0001$
					nb-wb: $t = 5.3$, $p \leq 0.0001$
					wb-pb: $t = 2.9$, $p < 0.01$
		max. displacement: segment	1.5	0.23	–

There is generally no effect of segment on this relationship, though a very small effect was found for tongue tip movements in English. Major differences existed by prosodic boundary such that phrase-initial productions had a lower stiffness than noninitial productions. Word-initial productions patterned with noninitial productions in English and phrase-initial productions in Spanish.

a significant effect of the scaling factor representing the slope of the logistic function, suggesting that there was no difference in place of articulation based on movement duration. Additionally, the model for English showed no significant effect of segment, in agreement with the observation that /t/, /d/, and /n/ are all produced at the alveolar ridge. For Spanish, however, there was a significant effect of segment on the asymptote

DOI: 10.1159/000481099

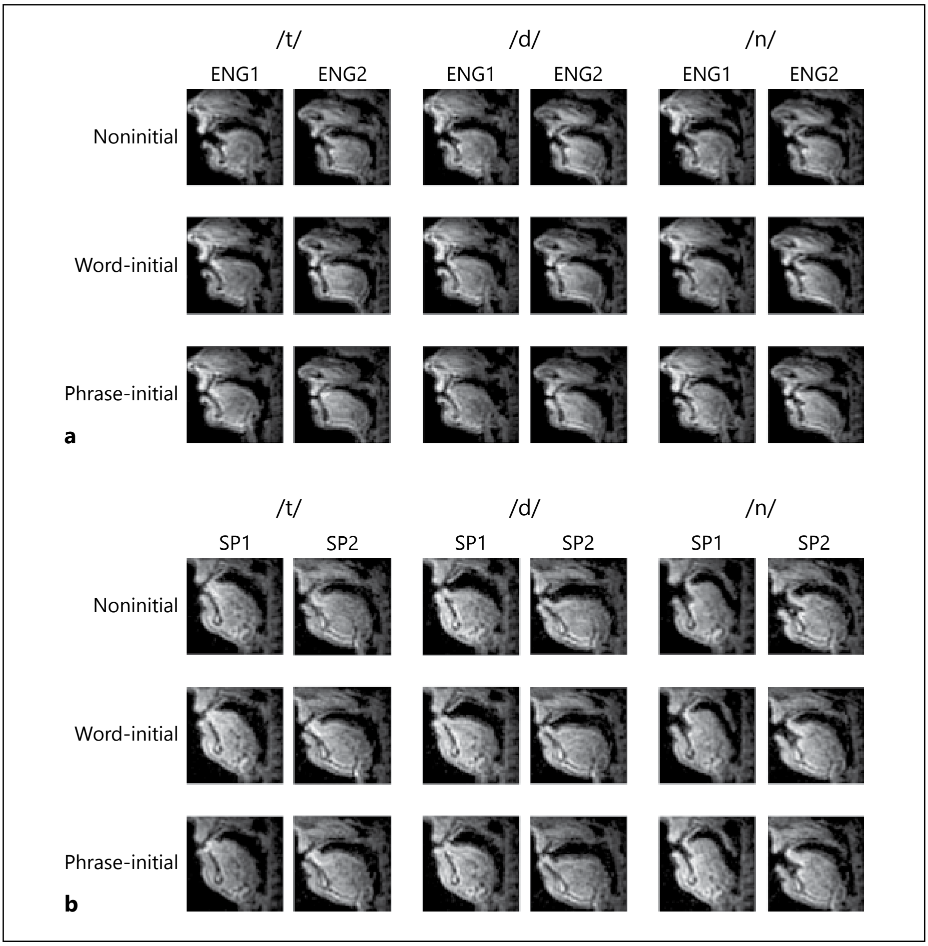


Fig. 5. Qualitative comparison of representative examples of productions in English (**a**) and Spanish (**b**). Representative productions for each participant are shown for /t/, /d/, and /n/ in all prosodic positions. Participants within each language showed very similar results. For the noninitial productions, /n/ in both languages is produced as flaps, with a small amount of linguopalatal contact and a slightly curled tongue tip. English /t/ and /d/ are also flapped, while Spanish /d/ is produced as an approximant at the teeth and /t/ is produced as a full stop at the front of the palate.

value, such that /d/ was more anterior than either /t/ or /n/ and /t/ was more anterior than /n/. This difference between /d/ and /t/ is unexpected, as they are generally grouped together as dental stops. More detailed results, including post hoc tests, are shown in Table 6.

In terms of constriction width, both English and Spanish showed a significant effect of the scaling factor of the logistic model, suggesting that contact width increased with movement duration. Again, this agrees with the impressionistic results examining Figure 5. There was no significant effect of segment for English, with a final asymptote value (i.e. maximum constriction width) of 4.8 pixels. There was a significant effect

Table 6. Results of statistical models for constriction location and width for English and Spanish

	Language	Parameter	*F* value	*p* value	Post hoc comparisons
Constriction width					
	English	asymptote	308.6	<0.0001	–
		asymptote ~ seg.	0.4	0.64	–
		scale	2.8	0.09	–
		xmid	11.6	<0.001	–
	Spanish	asymptote	142.0	<0.0001	–
		asymptote ~ seg.	366.0	<0.0001	t-d: $t = 2.2$, $p < 0.05$ t-n: $t = 21.1$, $p < 0.0001$ d-n: $t = 17.2$, $p < 0.0001$
		scale	0.3	0.56	–
		xmid	80.0	<0.001	–
Constriction location					
	English	asymptote	256.8	<0.0001	–
		asymptote ~ seg.	0.2	0.81	–
		scale	66.5	<0.0001	–
		xmid	3,179.9	<0.0001	–
	Spanish	asymptote	533.0	<0.0001	–
		asymptote ~ seg.	127.0	<0.0001	t-d: $t = 22.0$, $p < 0.0001$ t-n: $t = 5.6$, $p < 0.0001$ d-n: $t = 18.0$, $p < 0.0001$
		scale	599.5	<0.0001	–
		xmid	8.2	<0.01	–

Neither language showed a significant effect of *scale* for constriction location, suggesting that location did not vary with movement duration. The width of the constriction did vary with duration, however, with wider constrictions and longer durations. English did not show an effect of segment on either location or width, while Spanish showed differences by segment for both measurements.

of segment for Spanish, however. For Spanish, /d/ had less contact between the tongue and palate than either /n/ or /t/, and /t/ had more contact than /n/ (/d/: 0.9 pixels, /n/: 5.1 pixels, /t/: 6.2 pixels).

4 Discussion

This paper tested the hypothesis that spatial reduction of coronal stops is a continuous, dynamic process driven by prosodically conditioned durational variability rather than a strictly allophonic alternation. Based on this primary hypothesis, we made 4 concrete predictions:

1) There should be a clear relationship between the duration and spatial extent of gestures for coronal stops regardless of the language or segment.

2) Similar relationships should be seen not only for the tongue tip, but also for the other articulators involved in coronal stop production, the tongue body and jaw.

3) Since reduction is not caused by changes in articulatory effort, differences in spatial extent of gestures should be independent of changes in gestural stiffness.

4) The final outcome of reduction should be determined by the specific place of articulation of coronal stops.

The current results confirm all 4 of these predictions. We showed that there exists a clear, though nonlinear, relationship between movement duration and the extent of that movement. The precise nature of this relationship, including the slope and maximum movement extent vary between segments, articulators, and languages, but the relationship is significant in all cases. Interestingly, there is overlap in the range of movement between productions in the no-boundary and word boundary conditions for all segments in both languages. This is consistent with previous reports (e.g., Stone and Hamlet, 1982) and suggests substantially more variability than would be expected under a prosodically conditioned allophonic alternation account.

Neither are the current data well explained by changes in the stiffness of speech movements. Results from both languages indicate that while stiffness is modulated near larger prosodic boundaries, it is relatively constant across segments and across a wide range of displacements. Because the same stiffness was found spanning both small and large movements, this suggests that stiffness is not the driving factor in reduction, as might be suggested by a model where reduction results from effort minimization (Kirchner, 1998, 2004). These results are also consistent with past work that has shown temporally conditioned spatial reduction in labial stops with a constant stiffness, even when this does not lead to a loss of bilabial contact (Vatikiotis-Bateson and Kelso, 1990).

We note here that our measure of stiffness is an indirect one, as it relies on measure changes in pixel intensity in the MRI rather than directly on articulator movement per se. However, these changes in pixel intensity have been shown to correlate very closely with hand-measured changes in articulator position (Lammert et al., 2013). Moreover, our results closely match previous work comparing stiffness in prosodic position in English, showing that speech gestures have a lower stiffness in phrase-initial than non-phrase-initial position (Beckman et al., 1992; Byrd and Saltzman, 1998, 2003; Byrd et al., 2000; Cho, 2006; Edwards et al., 1991). Together, this suggests that our indirect measure of articulator stiffness is reliable.

As predicted, reduction occurs for the nasal /n/ as well as the oral stops in both Spanish and English. For both languages, /n/ shows a similar pattern of temporally conditioned spatial reduction as /t/ and /d/. This would be unexpected if reduction were an allophonic alternation process as the traditional accounts of reduction generally do not propose that /n/ also undergoes reduction, though a few accounts of English flapping do (e.g., Trager and Smith, 1951). It is, however, the expected outcome if reduction is a dynamic process as the tongue and jaw movements required to produce a nasal stop are more or less equivalent to those required for an oral stop. This is particularly true for English, where /n/, /d/, and /t/ are all produced with similar amounts of linguopalatal contact and at the same location along the palate.

Since Spanish /n/ is produced slightly differently than the Spanish oral stops (with contact at the alveolar ridge rather than the teeth/anterior palate), we predicted that /n/ would be similar to /d/ and /t/ in terms of the relationship between movement duration and maximum displacement, but that the final production would be a flap, rather than an approximant. This prediction was confirmed both qualitatively and quantitatively. In fact, Spanish /n/ is virtually identical to English /n/, with a small amount of contact at the alveolar ridge suggesting an apicoalveolar flap (Fig. 5). This finding contradicts a previous study, which suggested that /n/ may be produced as an approximant in Spanish (Honorof, 2003). However, that study used electromagnetic articulometry to track a point on the tongue roughly 1 cm posterior to the tongue tip. This method may miss the distinction between approximant and flapped productions as the tracked point on the tongue tip is posterior to the apex (meaning the actual constriction is approximated in electromagnetic articulometry but not measured), and electromagnetic articulometry does not allow for visualization of tongue shaping. Using real-time MRI, with its complete image of the midsagittal vocal tract, allows us to clearly see that coronal reduction outcomes are conditioned by the precise location of linguopalatal contact.

We have so far ignored the full productions of /t/ versus spirantization of /d/ in noninitial position in Spanish. This cannot be explained as just the consequence of durationally conditioned spatial reduction as both /t/ and /d/ show similarly attenuated displacement at short durations. One clue to what may be happening here comes from the word-initial and phrase-initial productions. Here, the productions of /t/ have much wider linguopalatal contact than /d/ (weak-phrase-initial /d/: 1.7 pixels, phrase-initial /d/: 2.4 pixels, noninitial /t/: 5.6 pixels, phrase-initial /t/: 5.6 pixels). A possible explanation for the wide sagittal contact in /t/ is that it comes from a virtual target beyond the surface of the palate. When the tongue hits the palate, it is impeded, and spreads out across the palatal surface. Based on this idea, it would follow that perhaps /d/ has a movement target that is closer to the palate, resulting in less spreading of the tongue, as has been proposed for voiced bilabial Spanish stops (Parrell, 2011). The data here seem to agree with this hypothesis. If this analysis is correct, the tongue tip for /d/ would just barely touch the hard palate when it has time to fully reach its target (as might occur at the long durations associated with occurring in phrase-initial position). This also suggests that further decreases in duration of /t/ could lead to incomplete oral closure, resulting in an approximant production. Although we did not observe this in the current study, productions of /t/ with incomplete closure have frequently been found in acoustic studies (e.g., Machuca, 1997).

Loss of contact for alveolar stops at extremely short durations may also explain results indicating that English /t/ and /d/ are often produced as approximants or are seemingly deleted entirely, at least in the acoustic signal (Warner and Tucker, 2011). In fact, in our data set we find that approximately 10% of noninitial productions of all English coronal stops fail to achieve palatal contact in the midsagittal plane. From an allophonic perspective, this would require another (optional) allophonic alternation. A different explanation arises naturally in our approach, where approximants are simply the far end of a continuum of productions linking full stops, flaps, and approximants.

Interestingly, past work has suggested that there is substantial variation in the precise location and orientation of the tongue tip for different speakers of English (Dart, 1998). In that data set, up to 20% of productions of coronal stops were produced laminally, with no apical contact. Laminally produced stops in our framework would be hypothesized to reduce to approximants, instead of the expected flaps. Although this is not typically accounted for in allophonic accounts, approximant productions are attested in English (Warner and Tucker, 2011). Perhaps these productions come from laminal, rather than apical, productions (possibly in addition to the shorter duration account suggested above). While the data in Dart (1998) show substantial variability, the total number of stops produced examined is relatively small (2 tokens each from 20 speakers, 40 total) and combines initial and final stops, all in front vowel contexts. In contrast, our speakers do not show the same amount of variation (see Results), it will be important to explore how such possible differences may influence reduction in future work, as well as testing whether approximant productions of coronal stops may result from extremely short durations and/or tongue shaping differences.

Overall, the spatial reduction seen in the current study is not well explained as the consequence of a categorical or allophonic alternation, as has been suggested in many phonological accounts of these reduction processes (e.g., Harris, 1969; Kahn, 1976). Rather, the amount of tongue tip, tongue body, and jaw movement in both languages varies dynamically with prosodically influenced changes in duration. This undershoot of a spatial target at shorter durations is not unique to this particular case of coronal reduction but is pervasive in speech production (e.g., Browman and Goldstein, 1990, 1992a; Lindblom, 1963). This suggests that the extreme reduction in the magnitude of these movements in noninitial position is due to the very short articulatory durations in these positions. This analysis also agrees with many previous findings on reduction in the 2 languages. All previous studies that have looked for evidence of clear allophonic alternation between full and reduced coronals in these languages have concluded that these sounds form a continuum of productions rather than distinct allophones. This is true both for American English (Fukaya and Byrd, 2005; de Jong, 1998; Stone and Hamlet, 1982) and Spanish (e.g., Cole et al., 1999; Hualde et al., 2011; Parrell, 2011; Simonet et al., 2012).

4.1 Cross-Linguistic Patterns

Moving beyond the 2 languages considered in the current study, our hypothesis predicts that when languages show coronal reduction, dental coronals should always reduce to approximants ([ð̞]), while alveolar coronals should always reduce to a flap/tap ([ɾ]). Although we have not, to this point, made an explicit prediction about retroflex stops, the logic driving the proposed distinction between dentals and coronals suggests that retroflex stops should also reduce to (retroflex) flaps: flaps are suggested to result when the gestural target is on the palate, while reduction to an approximant results from an articulatory target at the teeth (leading to limited linguopalatal contact as the magnitude of the gesture reduces and a loss of a complete oral seal). Retroflex stops, like alveolar stops, should maintain oral closure even with limited contact because the target is on the palate.

In the comprehensive survey of consonant reduction conducted by Kirchner (1998, 2004), there are 55 languages which show reduction of coronal stops. Of these, 5 show unexpected reductions to either fricatives (Turkana: /t/ > [s]; Pengo: /ʈ/,/ ɽ/ > [z]) or liquids (Karao: /t/ > [l], /d/ > [r]; Hausa: /d/ > [r]; Proto-Bantu: /d/ > [l]). These exceptions will be taken up below. In the remaining 54 languages, many are left unspecified in this survey for the precise characteristics of the coronal articulation. In the languages where the detailed manner and place of articulation are unknown, 20 show reduction to [ð] and 12 to [ɾ]. For the cases where the place is known, the results are presented in Table 7.

Note that all the dental stops reduce to [ð] and all the alveolar/retroflex stops reduce to [ɾ]. Interestingly, there are 2 languages in the survey that show reduction of both dental and alveolar stops, Purki and Yindjibarndi. In both cases, the dentals reduce to [ð] and the alveolars to [ɾ] (Yindjibarndi also shows reduction of retroflex stops to [ɾ]). These general results agree with our hypothesis that the precise location of the target constriction is the determining factor in the outcome of coronal reduction is borne out.

Phonetica 2018;75:151–181
DOI: 10.1159/000481099

Table 7. Results of coronal lenition at 3 places of articulation from the survey in Kirchner (1998, 2004)

	Dental	Alveolar	Retroflex
ð̞	5	0	0
ɾ	0	3	14

The numbers shown in each box reflect the number of occurrences of that pattern in the survey. Some languages that show reduction at more than 1 place of articulation may be included more than once. As predicted, dental stops reduce to approximants, and alveolar/retroflex stops reduce to flaps.

Interestingly, there are 22 languages in the survey that show reduction of voiced stops at all places of articulation. Of these 22, 15 show the pattern /b, d, g/ > [β, ð̞, ɣ]. Three of these languages also reduce /ɖ/ to [ɾ]. The remaining show the pattern /b, d, g/ > [β, ɾ, ɣ]. Given these data, it seems that the mechanism driving these patterns of reduction across multiple places of articulation (hypothesized here to be prosodically conditioned duration changes) always results in approximant labial and velar productions, but can lead to either approximants or flaps, depending on the language. This is expected given the hypothesis proposed here that prosodic variation underlies reduction to both flaps and approximants, with the eventual outcome conditioned by the particular language-specific articulatory posture of the tongue during coronal production.

We now turn briefly to exceptions to the predicted patterns of reduction, starting with the unexpected reductions seen in the Kirchner survey, where coronals reduced to fricatives or liquids. It is possible that the reductions to liquids, at least, are actually the result of an approximant production of an alveolar tongue tip gesture. Such a production would be expected if duration were to be reduced to the point where full contact between the tongue and palate could no longer be produced. Contact with only the center portion of the tongue would lead to a lateral, while a total loss of contact would lead to an alveolar approximant. Alternatively, the same result may arise when a stop has a less extreme target. It has been suggested that stops in many languages have a "virtual target" so that the goal for coronal stops is to move the tongue tip to a target point beyond the teeth or front portion of the hard palate (Löfqvist, 2005; Löfqvist and Gracco, 1997, 2002; Perrier et al., 2003; Westbury and Hashi, 1997). Critically, the exact virtual target for a stop will affect the duration at which it ceases to make palatal contact: a given gestural duration may lead to incomplete closure for a stop with a virtual target closer to the palate surface, while a stop with a more extreme virtual target with the same duration would maintain contact (Parrell, 2011). Thus, the same amount of durational shortening might lead to productions of flaps in a language with a virtual target far from the palate and an approximant in a language with a less extreme target.

In addition to the reduction to liquids in the Kirchner database, we are aware of at least a handful of examples that do not clearly fit with our hypothesis that alveolar stops should reduce to flaps, while coronal stops should reduce to approximants. First, we have the case of Danish, where alveolar stops reduce to approximants in syllable-final position (Basbøll, 2005). Just as for the reduction to liquids discussed above, approximant realizations could result either from loss of linguopalatal contact due to extremely short movement durations in prosodic conditions which favor reduction or a virtual target for alveolar stops in Danish close to the palate. Alternatively, the differences may have to do with syllable position – reduction for English and Spanish generally occurs intervocalically while in Danish it occurs syllable-finally. While the current proposal does not address syllable position effects on reduction, it is clear from other types of production changes that syllable position can play a critical role – e.g., intervocalic position often leads to voicing of voiceless consonants, while syllable-final position often leads to devoicing of voiced consonants (Kirchner, 1998). Further experiments and, perhaps, articulatory modeling may resolve this question by testing whether either a less extreme spatial target or shorter gestural duration is found for Danish alveolar stops compared to, e.g., English.

On the other hand, dental stops may result in flaps if the target position is actually on the palate posterior to the teeth rather than at the teeth themselves. In this case, we expect productions to maintain palatal contact at relatively short durations, just as we expect and have found for alveolar stops. Such differences in the precise place of articulation of dental stops might underlie the reports of reduction to a flap, rather than the expected approximant, in Dominican Spanish (Nuñez Cedeño, 1987) and Lekeitio Basque (Hualde et al., 1994). Again, further articulatory studies on these languages may resolve this question, including verification of the reported patterns.

It is also possible that reduction of coronal stops to approximants in some languages may be driven by the presence of an existing phonemic tap or flap, as exists in Spanish. While we cannot rule this out given the data here, it adds an additional layer of complexity to coronal reduction processes that our current hypothesis avoids. Additionally, the demonstrated existence of reduction of coronal stops to flaps even in languages that have existing rhotic flaps, such as the Dominican Spanish example cited above, show that any such constraints of the phonological system on coronal reduction cannot be universally applicable.

The variety of coronal reduction seen across English dialects shows that this process may be more complex than our hypothesis initially predicts. Different varieties of English reduce voiceless coronal stops to glottal stops (Browman and Goldstein, 1992b) or fricatives (Liverpool English; Honeybone, 2012). In the case of glottal stops, it is possible that durationally driven spatial reduction may still underlie this change: final voiceless stops may be produced with an additional glottal closure gesture even when produced with full oral closure, which would remain present even if the oral stop gesture fails to make palatal contact (Browman and Goldstein, 1992b). In the case of reduction of voiceless stops to fricatives, this change would entail a loss of oral contact with the retention of glottal spreading. Interestingly, the same oral aperture can lead to resonance (i.e., approximants) when combined with voicing but frication (i.e., fricatives) when the vocal folds are separated, due to the slower airflow in the former case (Ohala, 1983). This means that the reduction of voiceless stops to fricatives may in fact result from the same durationally conditioned shortening that results in approximant productions of voiced stops. Of course, these suggestions are only possible reasons why reduction of voiceless stops to glottal stops or fricatives occurs and remain to be confirmed or disconfirmed experimentally.

5 Conclusions

The current study examined reduction of coronal stops in American English and Spanish. The results show that the process of reduction is similar in both languages: shorter durations of movements associated with noninitial position result in smaller

Phonetica 2018;75:151–181
DOI: 10.1159/000481099

movements of all 3 articulators involved in making the coronal constriction – the tongue tip, tongue body, and jaw. The eventual articulatory outcomes of this durational and spatial reduction are conditioned by the precise manner in which each language produces coronal stops. Differences in constriction location result in flapping of American English for /t/, /d/, and /n/ as well as of Spanish /n/, spirantization of Spanish /d/, and full stop productions of Spanish /t/. These patterns are also found in a wide survey of languages, where alveolar and retroflex stops typically reduce to flaps, and dental stops always reduce to dental approximants. Though there are exceptions to this pattern, many could result from similar processes as seen in the current data with differences in the particular constriction targets (differences in either target location or target distance from the palate). These cases warrant further study to resolve whether such possible explanations are indeed correct.

While our results must be taken with some caution as they come from only 2 speakers each of only 2 languages, they suggest that reduction, both in Spanish and American English as well as more generally, may be better explained by a dynamic process that is the result of both invariant (e.g. constriction target, location) and variable (e.g. duration) factors rather than a process of allophonic substitution. The very reduced productions typically found in noninitial position in American English and Spanish are simply the far end of a continuum of prosodically conditioned variation in duration and magnitude, in agreement with previous experimental work examining American English flapping (Fukaya and Byrd, 2005; de Jong, 1998; Stone and Hamlet, 1982) and Spanish spirantization (Parrell, 2011; Soler and Romero, 1999).

Acknowledgment

This work was supported by NIH grant DC007124.

References

Baker TA (2008): A Biomechanical Model of the Human Tongue for Understanding Speech Production and Other Lingual Behaviors;(unpubl doctoral dissertation, University of Arizona, Tucson.

Basbøll H (2005): The Phonology of Danish. Oxford, Oxford University Press.

Bauer L (2008): Lenition revisited. J Linguist 44:605–624.

Beckman M, Edwards J (1992): Intonational categories and the articulatory control of duration; in Tohkura Y, Vatikiotis-Bateson E, Sagisaka Y (eds): Speech Perception, Production, and Linguistic Structure. Tokyo, Ohmsha Ltd, pp 359–375.

Beckman M, Edwards J, Fletcher J (1992): Prosodic structure and tempo in a sonority model of articulatory dynamics; in Docherty G, Ladd RD (eds): Papers in Laboratory Phonology II: Gesture, Segment, Prosody. Cambridge, Cambridge University Press, pp 68–86.

Bresch E, Nielsen J, Nayak K, Narayanan S (2006): Synchronized and noise-robust audio recordings during real-time magnetic resonance imaging scans. J Acoust Soc Am 120:1791.

Browman CP, Goldstein L (1990): Tiers in articulatory phonology, with some implications for casual speech; in Kingston J, Beckman ME (eds): Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech. Cambridge, Cambridge University Press, pp 341–376.

Browman CP, Goldstein L (1992a): “Targetless” schwa: an articulatory analysis; in Docherty GJ, Ladd DR (eds): Papers in Laboratory Phonology II: Gesture, Segment, Prosody. Cambridge, Cambridge University Press, pp 26–56.

Browman CP, Goldstein L (1992b): Articulatory phonology: an overview. Phonetica 49:155–180.

Browman CP, Goldstein L (1995a): Dynamics and articulatory phonology; in Port RF, van Gelder T (eds): Mind as Motion: Dynamics, Behavior, and Cognition. Boston, MIT Press, pp 175–194.

Browman CP, Goldstein L (1995b): Gestural syllable position effects in American English; in Bell-Berti F, Raphael LJ (eds): Studies in Speech Production: A Festschrift for Katherine Safford Harris. Woodbury, American Institute of Physics, pp 19–34.

Byrd D (1994): Palatogram reading as a phonetic skill: a short tutorial. J Int Phon Assoc 24:21–34.
Byrd D (2000): Articulatory vowel lengthening and coordination at phrasal junctures. Phonetica 57:3–16.
Byrd D, Saltzman E (1998): Intragestural dynamics of multiple prosodic boundaries. J Phon 26:173–199.
Byrd D, Saltzman E (2003): The elastic phrase: modeling the dynamics of boundary-adjacent lengthening. J Phon 31:149–180.
Byrd D, Kaun A, Narayanan S, Saltzman E (2000): Phrasal signatures in articulation; in Broe MB, Pierrehumbert JB (eds): Papers in Laboratory Phonology V. Cambridge, Cambridge University Press, pp 70–87.
Carrasco P, Hualde JI, Simonet M (2012): Dialectal differences in Spanish voiced obstruent allophony: Costa Rican versus Iberian Spanish. Phonetica 69:149–179.
Cho T (2006): Manifestation of prosodic structure in articulatory variation: evidence from lip kinematics in English; in Goldstein L, Whalen D, Best C (eds): Papers in Laboratory Phonology 8: Varieties of Phonological Competence. Berlin, Mouton de Gruyter, pp 1–34.
Cho T, Keating PA (2001): Articulatory and acoustic studies on domain-initial strengthening in Korean. J Phon 29:155–190.
Cho T, Lee Y, Kim S (2011): Communicatively driven versus prosodically driven hyper-articulation in Korean. J Phon 39:344–361.
Cole J, Hualde JI, Iskarous K (1999): Effects of prosodic and segmental context on /g/-lenition in Spanish; in Fujimura O, Joseph BD, Palek B (eds): Proceedings of the Fourth International Linguistics and Phonetics Conference. Prague, Karolinium Press, pp 575–589.
Dart SN (1998): Comparing French and English coronal consonant articulation. J Phon 26:71–94.
De Jong K (1998): Stress-related variation in the articulation of coda alveolar stops: flapping revisited. J Phon 26:283–310.
De Jong K, Beckman M, Edwards J (1993): The interplay between prosodic structure and coarticulation. Lang Speech 36:197–212.
Eddington D (2011): What are the contextual phonetic variants of /β, ð, ɣ/ in colloquial Spanish? Probus 23:1–19.
Edwards J, Beckman M, Fletcher J (1991): The articulatory kinematics of final lengthening. J Acoust Soc Am 89:369–382.
Fougeron C, Keating PA (1997): Articulatory strengthening at edges of prosodic domains. J Acoust Soc Am 101:3728–3740.
Fuchs S, Perrier P, Hartinger M (2011): A critical evaluation of gestural stiffness estimations in speech production based on a linear second-order model. J Speech Lang Hear Res 54:1067–1076.
Fukaya T, Byrd D (2005): An articulatory examination of word-final flapping at phrase edges and interiors. J Int Phon Assoc 35:45–58.
Hagedorn C, Proctor M, Goldstein L (2011): Automatic analysis of singleton and geminate consonant articulation using real-time magnetic resonance imaging. Interspeech 2011, Florence, pp 409–412.
Harris JW (1969): Spanish Phonology. Cambridge, MIT Press.
Hayes B (1995): Metrical Stress Theory: Principles and Case Studies. Chicago, University of Chicago Press.
Honeybone P (2012): Lenition in English; in Nevalainen T, Traugott E (eds): The Oxford Handbook of the History of English. Oxford, Oxford University Press.
Honorof DN (2003): Articulatory evidence for nasal de-occlusivization in Castilian. Proceedings of the 15th International Congress of Phonetic Science, Barcelona, pp 1759–1763.
Hualde JI (1988): A Lexical Phonology of Basque; unpubl doctoral dissertation, University of Southern California, Los Angeles.
Hualde JI, Elordieta G, Elordieta A (1994): Supplements of ASJU: The Basque Dialect of Lekeitio. Lejona, Universidad del País Vasco.
Hualde JI, Shosted R, Scarpace D (2011): Acoustics and articulation of Spanish /d/ spirantization. Proceedings of the 17th International Congress of Phonetic Science, Hong Kong, pp 906–909.
Kahn D (1976): Syllable-Based Generalizations in English Phonology; unpubl doctoral dissertation, City University of New York.
Keating P, Cho T, Fougeron C, Hsu C-S (2003): Domain-initial articulatory strengthening in four languages. Pap Lab Phonol VI:145–163.
Kelso JA, Vatikiotis-Bateson E, Saltzman E (1985): A qualitative dynamic analysis of reiterant speech production: phase portraits, kinematics, and dynamic modeling. J Acoust Soc Am 77:266–280.
Kier WM, Smith KK (1985): Tongues, tentacles and trunks: the biomechanics of movement in muscular-hydrostats. Zool J Linn Soc 83:307–324.
Kingston J (2008): Lenition; in Colantoni L, Steele J (eds): Selected Proceedings of the 3rd Conference on Laboratory Approaches to Spanish Phonology. Somerville, Cascadilla Proceedings Project, pp 1–31.
Kirchner R (1998): An effort-based approach to consonant lenition; unpubl doctoral dissertation, University of California, Los Angeles.
Kirchner R (2004): Consonant lenition; in Hayes B, Kirchner R, Steriade D (eds): Phonetically Based Phonology. Cambridge, Cambridge University Press, pp 313–345.
Kochetov A, Colantoni L (2011): Coronal place contrasts in Argentine and Cuban Spanish: an electropalatographic study. J Int Phon Assoc 41:313–342.
Lammert A, Proctor MI, Narayanan SS (2010): Data-driven analysis of realtime vocal tract MRI using correlated image regions. Proceedings of Interspeech 2010, Makuhari.

Lammert A, Ramanarayanan V, Proctor M, Narayanan S (2013): Vocal tract cross-distance estimation from real-time MRI using region-of-interest analysis. Proceedings of Interspeech 2013, Lyon.
Lavoie LM (2001): Consonant Strength: Phonological Patterns and Phonetic Manifestations. New York, Routledge.
Lindblom B (1963): Spectrographic study of vowel reduction. J Acoust Soc Am 35:1773–1781.
Lindblom B (1983): Economy of speech gestures; in MacNeilage P (ed): The Production of Speech. New York, Springer, pp 217–245.
Lindblom B (1990): Explaining phonetic variation: a sketch of the H & H theory; in Hardcastle JW, Marchal A (eds): Speech Production and Modelling. Dordrecht, Kluwer Academic Publishers, pp 403–439.
Löfqvist A (2005): Lip kinematics in long and short stop and fricative consonants. J Acoust Soc Am 117:858–878.
Löfqvist A, Gracco V (2002): Control of oral closure in lingual stop consonant production. J Acoust Soc Am 111:2811–2827.
Löfqvist A, Gracco VL (1997): Lip and jaw kinematics in bilabial stop consonant production. J Speech Lang Hear Assoc 40:877–893.
Machuca MJ (1997): Las obstruyentes no continuas del español: relación entre las categorías fonéticas y fonológicas en habla espontánea; unpubl doctoral dissertation, Universitat Autònoma de Barcelona.
Martínez Celdrán E (2008): Some chimeras of traditional Spanish phonetics; in Colantoni L, Steele J (eds): Selected Proceedings of the 3rd Conference on Laboratory Approaches to Spanish Phonology. Somerville, Cascadilla Proceedings Project, pp 32–46.
Martínez Celdrán E, Fernández Planas AM, Carrera Sabaté J (2003): Castilian Spanish. J Int Phon Assoc 33:255–259.
Mascaró J (1984): Continuant spreading in Basque, Catalan and Spanish; in Aronoff M, Oehrle RT (eds): Language Sound Structure: Studies in Phonology Presented to Morris Halle by His Teacher and Students. Cambridge, MIT Press, pp 287–298.
Moon S-J, Lindblom B (1994): Interaction between duration, context, and speaking style in English stressed vowels. J Acoust Soc Am 96:40–55.
Mooshammer C, Fuchs S (2002): Stress distinction in German: simulating kinematic parameters of tongue-tip gestures. J Phon 30:337–355.
Narayanan S, Nayak K, Lee S, Sethy A, Byrd D (2004): An approach to real-time magnetic resonance imaging for speech production. J Acoust Soc Am 115:1771–1776.
Nuñez Cedeño RA (1987): Intervocalic /d/ rhotacism in Dominican Spanish: a non-linear analysis. Hispania 70:363–368.
Ohala JJ (1983): The origin of sound patterns in vocal tract constraints; in MacNeilage FP (ed): The Production of Speech. New York, Springer, pp 189–216.
Ortega-Llebaria M (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 Face TL (ed): Laboratory Approaches to Spanish Phonology. Berlin, Mouton de Gruyter, pp 237–254.
Parrell B (2010): Articulation from acoustics: estimating constriction degree from the acoustic signal. J Acoust Soc Am 128:2289.
Parrell B (2011): Dynamical account of how /b, d, g/ differ from /p, t, k/ in Spanish: evidence from labials. Lab Phonol 2:423–449.
Parrell B, Lee S, Byrd D (2013): Evaluation of prosodic juncture strength using functional data analysis. J Phon 41:442–452.
Perrier P, Payan Y, Zandipour M, Perkell J (2003): Influence of tongue biomechanics on speech movements during the production of velar stop consonants: a modeling study. J Acoust Soc Am 114:1582–1599.
Piñeros CE (2002): Markedness and laziness in Spanish obstruents. Lingua 112:379–413.
Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team (2016): nlme: linear and nonlinear mixed effects models. R package version 3.1–128. http://CRAN.R-project.org/package = Nlme.
Pouplier M (2012): The gaits of speech: re-examining the role of articulatory effort in spoken language; in Solé M, Recasens D (eds): The initiation of sound change: perception, production, and social factors. Amsterdam, Benjamins, pp 147–164.
Proctor M, Lammert A, Katsamanis A, Goldstein L, Hagedorn C, Narayanan S (2011): Direct estimation of articulatory kinematics from real-time magnetic resonance image sequences. Interspeech 2011, Florence, pp 281–284.
Proctor M, Lo CY, Narayanan S (2015): Articulation of English vowels in running speech: a real-time MRI study; in The Scottish Consortium for ICPhS 2015 (ed): Proceedings of the 18th International Congress of Phonetic Sciences. Glasgow, University of Glasgow.
Saw CC (1993): Customized 3-D electropalatography display. UCLA Work Pap Phon 85:71–96.
Simonet M, Hualde JI, Nadeu M (2012): Lenition of /d/ in spontaneous Spanish and Catalan. Interspeech 2012, Portland.
Soler A, Romero J (1999): The role of duration in stop lenition in Spanish. Proceedings of the 16th International Congress of Phonetic Science, Saarbrücken, pp 483–486.
Stone M, Hamlet S (1982): Variations in jaw and tongue gestures observed during the production of unstressed /d/s and flaps. J Phon 10:401–415.
Stone M, Lundberg A (1996): Three-dimensional tongue surface shapes of English consonants and vowels. J Acoust Soc Am 99:3728–3737.

Trager GL, Smith HL (1951): An Outline of English Structure. Washington, American Council of Learned Societies.
Turk A (1992): The American English flapping rule and the effect of stress on stop consonant durations. Work Pap Cornell Phon Lab 7:103–133.
Umeda N (1977): Consonant duration in American English. J Acoust Soc Am 61:846–858.
Vatikiotis-Bateson E, Kelso J (1990): Linguistic structure and articulatory dynamics: a cross language study. Haskins Lab Stat Rep Speech Res SR-103/104:67–94.
Vatikiotis-Bateson E, Kelso JAS (1993): Rhythm type and articulatory dynamics in English, French and Japanese. J Phon 21:231–265.
Warner N, Tucker BV (2011): Phonetic variability of stops and flaps in spontaneous and careful speech. J Acoust Soc Am 130:1606–1617.
Westbury J, Hashi M (1997): Lip-pellet positions during vowels and labial consonants. J Phon 25:405–419.
Zue VW, Laferriere M (1979): Acoustic study of medial /t, d/ in American English. J Acoust Soc Am 66:1039.

Phonetica 2018;75:151–181
DOI: 10.1159/000481099