Open Library of Humanities

Stop lenition in Canary Islands Spanish – a motion capture study

Karolina Broś*, Institute of Applied Linguistics, University of Warsaw, Warsaw, Poland, k.bros@uw.edu.pl

Peter A. Krause, Department of Psychology, California State University Channel Islands, CA, USA, peter.krause@csuci.edu

*Corresponding author.

In this paper, we investigate the phonological and prosodic effects of lenition in Canary Islands Spanish using new methodology. In a motion capture study conducted in the field, we show that using an internet camera for lip tracking can help disentangle several problems related to the contexts of stop lenition. First, we demonstrate that there is no post-consonantal lenition of stops in the studied variety of Spanish. We also show that the same kind of lenition blocking occurs after a preceding consonant is deleted, which supports containment-based approaches to phonological analysis and contributes to the debate on gradient activation and gestural masking effects in phonology. Furthermore, we show that lip aperture, as opposed to acoustic measurements, points to differences between stops produced in focus and non-focus positions, which constitutes an added value in studying lenition. All in all, we show that capturing lip movements is a viable method in investigating consonantal constriction and that studies on the subject would benefit from combining easy-to-use articulation tracking with acoustic analysis.

1. Introduction

Stop lenition is a well-known phenomenon across the varieties of Spanish spoken around the world. Otherwise referred to as weakening, it consists in the loss of consonantal features, e.g., by voicing or spirantization, eventually leading to complete segment elision. In the Canary Islands, lenition concerns both /b d g/ and /p t k/ (Alvar, 1972; Oftedal, 1985). While the phenomenon itself has been studied quite extensively, there are several outstanding theoretical questions involving 1) the role of intervocalic vs. post-consonantal environments, 2) prosodic-level effects of stress and focus and 3) how lenition is realized articulatorily.

While stops usually undergo lenition in intervocalic position in Canarian Spanish, it is not entirely clear whether /p t k/ and /b d g/ are also weakened after other sounds. While the literature on Spanish dialectology states that /p t k/ weakening is post-vocalic only and /b d g/ are also weakened after non-nasal consonants in most geographical varieties of the language, this assumption is not so easy to confirm for the dialect under study in this paper given the overlap with the widespread elision of syllable-final consonants. Previous studies on Canarian Spanish suggest that the dialect may be closer to some Latin American varieties in blocking rather than allowing approximantization post-consonantally. However, definitive conclusions cannot be drawn on the subject given that many of the underlying stops stand in a derived intervocalic position (i.e., become intervocalic after the deletion of a preceding consonant) and hence their post-consonantal behavior cannot be reliably reported (the preceding consonant is not pronounced). Nonetheless, it has been demonstrated that such stops behave differently from the same sounds occupying underlying intervocalic positions. More specifically, consonant weakening is at least partially blocked (Broś, 2016; Broś et al., 2021).

Taking the above into account, our current knowledge of the behavior of /p t k b d g/ in the Canarian dialect is limited. A study specifically designed to elicit both retained and deleted consonants preceding the stops is necessary to tease apart the contexts of stop weakening. This is the first goal of our experiment.

Furthermore, studies on lenition in Spanish have shown that stress is a key factor (e.g., Cole et al., 1999; Hualde et al., 2011). However, given that stop weakening in this language takes place in different sentence positions, very often at word junctures, it is worth looking into the possible prosodic effects beyond the prosodic word. Thus, the present study investigates weakening in larger utterances, in which phrase-level prosody may play a role. More specifically, this study will contrast consonants in focus and non-focus positions.

Finally, in our explorations of the key prosodic and phonological determinants of stop lenition, we use new methodology to disentangle articulation from acoustics and see their interplay. It is worth noting that while most studies on lenition in Spanish rely on acoustic measurements, especially the intensity of the target segment compared to the flanking vowels, there are only a

few articulatory investigations (among them Parrell, 2011). In this study, we use video recordings and extract information concerning lip movements to measure relative consonant aperture as a proxy of degree of lenition using a procedure elaborated by Krause et al. (2020). As a result, we provide an easily implemented and cost-effective alternative to electromagnetic articulography (EMA) that can be used in the lab and the field without major complications. It is our hope that the results of the study encourage researchers to take on similar tasks in the future as a means of expanding the methodological repertoire available to phonologists.

The remainder of the paper is organized as follows: We introduce Spanish stop weakening in more detail and the roles of particular prosodic and segmental features in sections 1.1. and 1.2., respectively. We also look at key acoustic and articulatory approaches to stop lenition in sections 1.3. and 1.4. We then present our study in Section 2 and the statistical results in Section 3. This is followed by a discussion of the results in Section 4.

1.1. Consonant lenition in Spanish and the Canary Islands dialect

In Spanish, the typical manifestation of intervocalic consonant lenition is the approximantization of /b d g/. In most Spanish dialects, the degree of openness of the resultant sound differs depending on such factors as stress, place of articulation, quality of the flanking vowels, speech rate, etc. (see Hualde, 2005 for an overview).[1] It is also possible to see /b d g/ delete completely in some cases.

Some dialects of Spanish have also developed a further change – weakening of the /p t k/ series. Canary Islands Spanish is perhaps the most notorious in this respect, with voiceless stops undergoing voicing and (variably) approximantization post-vocalically (Oftedal, 1985; Trujillo, 1980).

Furthermore, while voiceless stop lenition is restricted to post-vocalic environments, /b d g/ also weaken after other sounds. The common generalization in the literature is that /b d g/ are phonetically realized as stops after pauses, nasals and after the lateral /l/ (in the case of /d/), and as approximants elsewhere (Navarro Tomás, 1918; Lozano, 1979; Lipski, 1996; Hualde, 2005). Moreover, the degree of approximantization of the resultant sound depends on the preceding segment: the greater its constriction, the more constricted the target consonant (Machuca, 1997; Martínez-Celdrán & Regueira, 2008). Preceding fricatives, especially a preceding /s/, both when retained and when weakened to [h], typically trigger a less approximantized pronunciation of the underlying voiced stops (Eddington, 2011).

---

[1] The process has been often referred to as spirantization, although this is slightly misleading as the resultant sounds are not spirants in the general sense. These sounds have also been defined as 'spirant approximants' by Martínez-Celdrán (2008).

However, there have been reports on varieties of Spanish in which stop allophones of /b d g/ rather than approximants are preferred after consonants. Such a pattern can be seen in the Spanish spoken in Costa Rica, Nicaragua, El Salvador and Panama (Fernandez, 1982; Amastae, 1989; Lipski, 1996; Quesada, 1996; Carrasco et al., 2012, see Quesada, 2010 for a review), in Judeo-Spanish (Hualde, 2013) and in the Colombian highlands (Canfield, 1962).

It is not entirely clear which of the weakening trajectories is followed by Canary Islands Spanish. Previous literature suggests that the dialect may be closer to the abovementioned Latin American varieties. Focusing on the behavior of /b d g/ after /s/, several researchers observed that the sibilant is usually deleted in these contexts and a stop rather than an approximant is produced following this deletion in the Spanish of Gran Canaria (Trujillo, 1980; Almeida, 1982; Dorta & Herrera, 1993; Herrera, 1997; see also Hualde, 2005). Additionally, a compensatory durational effect takes place: The resultant stop is longer and even geminated according to some sources, marking the plural in the absence of the /s/ (Trujillo, 1981; Almeida, 1982). Phonologically, it has usually been interpreted as tense, with the notation [b: d: g:] used to refer to it.[2]

As for the comparison with other contexts, a study by Herrera (1997) looks at differences in VOT and duration of /b d g/ after deleted /s/ compared to /b d g/ after vowels and nasals. It is unclear, however, how /b d g/ behave after pauses and other consonants, especially aspirated /s/. [h] is reported in Almeida (1982) as a minority pronunciation of the /s/ that is followed by lax [b d g] unspecified for continuancy, as opposed to tense non-continuants [b: d: g:] observed after deleted /s/. These may be interpreted as approximants, which would mean that some approximantization takes place after fricatives in the dialect. There appears to be variation, however, as Almeida also lists [h] followed by tense [b: d: g:] as minority variants. Furthermore, a recent study of spontaneous speech from Gran Canaria shows some variability in voiced stop productions in the context of a preceding consonant (Broś et al., 2021). Here, preceding sound deletion diminishes the probability of weakening. Approximantization still occurs, but it is much less frequent (a drop from 95% to 30% on average). The study does not provide a comparison with pronounced preceding consonants, however, so we are again unable to say whether the same decrease in weakening rates would occur e.g., after an aspirated /s/.

To summarize, stop lenition in Canary Islands Spanish concerns both voiced and voiceless stops. Typically, /p t k/ are either voiced or voiced and approximantized, while /b d g/ are approximantized or deleted in post-vocalic position (Oftedal, 1985; Almeida & Díaz, 1988). Additionally, in post-vocalic contexts derived by preceding sound deletion, weakening is largely diminished or even blocked. There is also some indication that /b d g/ do not weaken after a preceding aspirated /s/, which would place Canarian Spanish next to Nicaraguan, Costa Rican and other varieties that diverge from the mainland Spain pattern (cf. Hualde, 2005). **Table 1** presents

---

[2] See Dorta and Herrera (1993) and Herrera (1997) for an overview of previous reports and notations.

a summary of these generalizations, listing the most frequent as well as minority realizations in each context.[3]

UR	Example	Majority realization	Other realizations
/p/	*la paciencia* 'the patience'	[la.**b̥**a.ˈsjen.sja]	[la.**p**a.ˈsjen.sja], [la.**b**a.ˈsjen.sja], [la.**β**a.ˈsjen.sja]
/p/	*Las Palmas*	[la.ˈ**p**almah]	[la.ˈ**b̥**almah], [la.ˈ**b**almah]
/b/	*la barrera* 'the wall'	[la.**β**a.ˈre.ɾa]	[la.**β**a.ˈre.ɾa], [laː.ˈre.ɾa]
/b/	*las vacas 'the cows'*	[la.ˈ**b**a.kah]	[la.ˈ**β**a.kah]

Table 1: Stop lenition in Canary Islands Spanish – majority and minority pronunciations by context. [b̥] is a partially voiced /p/, [β] is a bilabial approximant. The data are based on Broś et al. (2021). It is assumed that the /s/ is deleted as only such cases were analyzed in the cited study.

Given the data and the current state of knowledge concerning the occurrence of stop weakening in the Canarian dialect, it is worth delving into the role of the left-hand environment. More specifically, this study will investigate the effect of a preceding consonant, pronounced or not, on the constriction of the target stop.

1.2. The roles of lower and higher-level prosody in lenition

As mentioned in Section 1.1., stop lenition has been shown to be influenced by various factors other than the segmental environment. In this paper, we want to take a closer look at the prosodic effects, as not all their aspects have been explored to date. Previous research reported, for instance, that more lenition occurs in /b d g/ when the sound is preceded by a stressed syllable (Colantoni & Marinescu, 2010; Cole et al., 1999; Ortega, 2004). At the same time, other studies report sensitivity to the stress of the following vowel. According to Hualde et al. (2011), there is a significant increase in the voicing of /p t k/ in Iberian Spanish in unstressed syllables. Carrasco et al. (2012) found a similar effect of stress on /b d g/ in Costa Rican Spanish. Eddington (2011), on the other hand, showed that the following unstressed vowel is associated with more lenition while the following stress diminishes the rate of approximantization in /b g/ in several varieties of Spanish. Finally, studies on Canarian Spanish show that the weakening of both /p t k/ and /b d g/ is promoted by the following unstressed vowel (Broś & Lipowska, 2019; Broś et al., 2021).

---

[3] In this study, we do not focus on gradient effects nor on the probability of each surface option. Instead, similar to previous studies, we focus on majority pronunciations when presenting data and generalizations, and we only focus on relative measures i.e., relative changes in consonant constriction as marked by intensity and lip aperture depending on the context: stress, focus and preceding sound deletion; see Section 1.5.

Although previous studies focused on some word-level factors other than lexical stress, such as the presence of a word boundary (Cole et al., 1999; Eddington, 2011; Katz, 2016), to the best of our knowledge there has been no study so far that focuses on prosodic effects beyond the word domain.[4] Since lenition affects segments inside words and across word boundaries in Spanish and has been shown to be sensitive to stress, it is worth investigating higher-level prosody and phrasing effects.[5]

Such investigations should be informed by the accentual characteristics of Spanish. To provide some background, in Spanish, the main stress usually falls to the rightmost part of the sentence. Thus, the nuclear stress position is typically the last content word of the intonational phrase (Beckman et al., 2002; Hualde & Prieto, 2015). This is in line with the so-called Nuclear Stress Rule (NSR) by which nuclear stress is assigned to the rightmost lexical category of a sentence (Zubizarreta, 1998). Moreover, the Focus/Prominence Rule (FPR) requires that the focused constituent be the most prominent one. In a broad focus statement, where none of the words are under special emphasis, all lexical items are equally highlighted (Ladd, 1980) and stress is typically marked with an F0 rise in prenuclear position, with each peak being slightly lower than the previous one in the sentence (a pattern referred to as downstepping), until final lowering by which the pitch returns to the baseline level at the end of a neutral declarative sentence. This general rule has been confirmed for many varieties of Spanish (Face, 2000, 2001; Face & Prieto, 2007; Vanrell & Fernández-Soriano, 2018).[6]

Furthermore, it has been shown that Spanish should be classified as a stress accent language (Beckman, 1986), which uses duration in addition to pitch to signal accents. Thus, the nuclear position is metrically the strongest and its phrasally determined prominence is often associated with final lengthening (Ortega, 2006; Ortega & Prieto, 2007). Consequently, to explore the effects of phrasal/sentence stress on the lenition patterns of stops, we decided to design a study with sentences in which stops in stressed syllables would be placed either in prominent (focus) or non-prominent words. Syntactically, the right edge of an IP should fall on the object of an

---

[4] Ortega-Llebaria (2006) mentions that in her study focused on disentangling stress from accent in Spanish in terms of the phonetic cues involved, stressed syllables were produced with stop-like pronunciations of underlying voiced stops compared to unstressed syllables while no such effects could be seen in accented vs. unaccented syllables. Ortega-Llebaria argues that different cues are responsible for stress (here, intensity plays a major role) rather than accent in Spanish (in the latter case, pitch and duration are the determining factors). She reports, however, that intensity-related effects were not consistent across speakers.

[5] It is also worth mentioning that studies on lenition in different languages, such as Ennever et al. (2017) or Katz and Pitzanti (2019), have shown prosodic (or syntactic) effects beyond the word in that different degrees of lenition may apply depending on the level of a constituent in a phrase.

[6] There are certain differences in the pitch contours in declaratives and interrogatives in Caribbean Spanish and Canarian Spanish compared to the Iberian variants (see e.g., Dorta (ed.), 2013 and Hualde & Prieto, 2015 for a review). These are not relevant to the research questions pursued in this paper.

SVO sentence, and it is the object of the sentence that is the expected location of nuclear stress in Spanish. Thus, we determined the prominence of a given word based on the syntax of the sentence and compared sentence-final lexical items with those placed before the verb, toward the beginning of the utterance. The assumption of the study is that since stressed syllables inhibit lenition and promote more stop-like pronunciations, focus position should have an even greater effect on consonantal constriction.

1.3. Stop lenition – acoustic correlates

Stop lenition in Spanish and other Romance languages has been explored to date using a series of acoustic correlates. Some of them are qualitative or categorical, e.g., presence/absence of a burst in stops (Dalcher, 2008), presence/absence of formants on the spectrogram that helps identify approximants as opposed to stops or fricatives (Recasens, 2016; Katz & Pitzanti, 2019) or voicing (Herrera, 1997; Dalcher, 2008; Romero et al., 2007). Continuous measurements, on the other hand, typically include sound and/or constriction duration (Dalcher, 2008; Hualde et al., 2011; Lavoie, 2001; Parrell, 2011) and acoustic intensity-based calculations (Carrasco et al., 2012; Hualde et al., 2010; Hualde & Nadeu, 2012, Hualde et al., 2011). The latter is perhaps the most widely tested and accepted method of measuring the degree of constriction and hence weakening of consonants. Moreover, acoustic intensity has been shown to correlate quite well with lip aperture measured during articulation (Parrell, 2010). Thus, although intensity is not an articulatory measurement, it can, in principle, be used to determine consonant aperture. We therefore decided to use it as a baseline criterion for detecting different degrees of lenition in the present study.

While different ways of using intensity have been applied to date (intensity ratio, intensity difference and maximum velocity), intensity difference is by far the most popular as a marker of relative constriction degree. We will therefore focus on this parameter. In this case, two alternative measurements can be considered. Since stop lenition usually happens in intervocalic position, both a comparison with the preceding and the following vowel can be used. The intensity difference is obtained by subtracting the minimum intensity of the consonant from the maximum intensity of the vowel. In our study, we follow Martínez-Celdrán and Regueira (2008), Figueroa and Evans (2015) and Broś et al. (2021) in calculating the intensity difference with respect to the preceding vowel. We also consider the alternative measurement using the following vowel, as done by Hualde et al. (2011), Carrasco et al. (2012) and others. However, since we are interested in the effect of stress and accent (see Section 1.2.), we want to avoid confounding lenition with changes in the intensity of the vowel caused by its prosodic status. Thus, we believe that taking intensity from the preceding vowel is more suitable. We return to this issue briefly in the Discussion.

1.4. Ways of measuring aperture

While the obvious way of measuring aperture directly is electromagnetic articulography (EMA), its major disadvantage is that it is both time-consuming and resource intensive. For this reason, articulography studies are usually limited to just a few participants. Moreover, while all places of articulation can be studied via EMA, it has been shown that measuring constriction degree with this method can be quite complicated for velars and coronals. More specifically, Romero (1995: 39–40) reports that in the case of velars, no direct estimate of constriction degree can be determined for anatomical reasons, including the way soft palate is used during the production of velar sounds. Thus, constriction degree must be estimated indirectly by comparing the spatial positions of the articulators during the production of different velar sounds. In the case of coronals, using the very tip of the tongue instead of the blade may make it more difficult to obtain satisfactory measurements. Nevertheless, once these obstacles are overcome, consistent patterns of constriction degree and location can be found across subjects. To avoid problems with determining the constriction degree, Parrell (2011) decided to focus on labials only in his study on Spanish stop weakening.

To provide a measurement of constriction degree using EMA, the Euclidean distance between the sensors placed on the articulators, e.g., the lower and upper lip in the case of labials, is calculated. In his comparison of /p/ and /b/ productions in Spanish across different contexts, Parrell (2011) demonstrated that the articulatory data point to a difference in target constrictions between the voiceless and the voiced stop. The latter has a more open (i.e., less constricted) target, although it is still beyond the point of closure. The articulatory parameters show that combined with a decreased duration, this can lead to undershoot, resulting in approximantization in intervocalic contexts. At the same time, more constricted targets for /p/ do not lead to approximantization with undershoot (in the studied sample). Thus, articulation aligns with impressionistic generalizations and acoustic measurements of differences between /p/ and /b/, at least in Peninsular Spanish and similar varieties. By extension, any method that will allow for inferring the relative distance between the articulators (here, lips) should be equally suitable for investigating consonant lenition.[7] Motion capture is a promising technique in this respect.

Recently, there has been substantial interest in linking phonetic and phonological research with physical gestures. Since it has been shown that certain face and body movements serve as an enhancement mechanism for marking salient prosodic positions (McNeill, 2008; Wagner

---

[7] As for alternatives, ultrasonography could be potentially used to study aperture. However, due to its physical restrictions, it would enable analyzing perhaps only velars, at least in its 2D version. Nonetheless, it is not clear how precisely we would be able to measure the distance between the back of the tongue and the velum. Given small differences in constriction between voiced and voiceless stops, and between stops and approximants, high resolution data are necessary. A study by Shaw et al. (2020) is interesting in this respect as it looks at the lenition of velars in Iwaidja, showing a significant correlation between the intensity measurements and ultrasound-based articulatory data, which further confirms the strength of the former as a lenition parameter shown earlier by Parrell (2010).

et al., 2014), many researchers have taken on a multimodal approach to studying speech. For instance, Beskow et al. (2006) found that in Swedish, words with a focal accent are supplemented by greater facial movement than words in other positions in the sentence. Similarly, Sangari (2002) and Esteve Gibert and Prieto (2013) showed that the timing of head nods is aligned with prosodic boundaries and the metrical structure of the word in focus. Such studies used motion capture to track facial expressions, with sensors placed on the face and body, and optical tracking. Although they focused on stress and sentence focus, they did not look at segmental phenomena accompanying speech. The motion capture method, however, is in principle suitable for analyzing at least some aspects of consonantal phonology as it allows for tracking facial expressions such as vertical and horizontal lip movements. In this vein, Żygis et al. (2017) focused on orofacial gestures produced in questions and statements in German whispered speech using motion capture. According to their results, there is a significant difference in lip aperture between the two sentence types. The same applied to lip aperture measurements focused on particular vowels. Thus, reliable measurement of lip opening from video data should make it possible to study other more nuanced questions, such as the relative constriction of labial consonants. No study of this kind has been done to date, which makes the present paper the first to provide a motion capture account of the weakening process.

While the motion capture method using optical tracking with sensors is quite successful, it also requires a special lab and equipment. Fortunately, some fieldwork-friendly alternatives to such endeavors are now available. Importantly, Holbrook et al. (2019) demonstrated the viability of optically tracking lip movements by using inexpensive off-the-shelf video recording equipment and algorithms. Krause et al. (2020) extended this approach by basing their measurements on intelligent face-tracking. Recordings can be comfortably made in the lab or in the field using an internet camera. With a specially-designed algorithm based on the OpenFace face-tracking utility (Baltrušaitis et al., 2018), the method allows for extracting lip aperture and lip area measurements from video material, provided that the speaker was seated without substantial movement facing the camera. Most importantly, Krause et al.'s study shows that internet camera-based lip tracking is robust and reliable enough to distinguish speech productions at the segmental level. Furthermore, the method allows not only for detecting lip closure vs. lip opening, but also for distinguishing different types or degrees of lip closure by looking at relative lip compression, measuring on the outer and inner lip contour. This is important from the perspective of the present study as we are especially interested in various degrees of lip closure and slight lip aperture corresponding to voiceless and voiced stops, and approximants, respectively.

1.5. The present study – goals and hypotheses

The principal goal of this study is to investigate the prosodic and segmental phonological effects of stop weakening in Canary Islands Spanish using a novel method. A motion capture study of lip

movements gives us a link between the acoustics and articulation, and it allows us to determine whether lip movements tracked from video data can predict different degrees of consonant weakening (from stop voicing to strong approximantization) and whether such data can be used in future research alongside acoustic measurements. Thus, we hope that our study will constitute an important methodological contribution to laboratory phonology.

At the same time, we are interested in two important questions related to the weakening of underlying stops in Spanish. First, previous studies show that the deletion of a preceding consonant makes the stop behave differently than in underlyingly intervocalic position, which suggests that there is a phonological blocking effect that can be expressed as containment, i.e., non-deletion or incomplete deletion (non-pronunciation) of the apparently deleted consonant (Prince & Smolensky, 1993/2004; van Oostendorp, 2006). If the root node of the consonant is still there phonologically, the lack of weakening is explained. Such an interpretation, however, requires empirical support. An articulatory study can help us elucidate this question. A consonantal gesture can still be present even though the sound is not audible, e.g., due to gestural masking (or blending). As argued by Browman and Goldstein (1990), two gestures from two different tiers may sometimes mask or 'cover' each other, leading to apparent deletion. Some examples of this come from English. For instance, in sequences of words such as *must be*, the /t/ seems to be deleted in fast speech but palatography evidence shows that an alveolar closure gesture is there. In our case, it may be that a preceding consonant is masked by the following stop gesture and, as a result, apparently deleted. It is therefore worthwhile to look for comparative contexts. By looking at data elicited in a reading task, given that /s/ deletion is optional, we can compare contexts in which the stop is underlyingly intervocalic with two other types of outputs: a situation in which a preceding /s/ is retained (but e.g., weakened to [h]) and a situation in which it is deleted completely. In this way, we can look at the differences and similarities between the latter two contexts both in terms of acoustics and articulation to disentangle the effect of the preceding /s/ from the effect of a deleted segment. More specifically, we would like to know whether the /b/ in *la vaca* (the cow) /la#baka/, *las vacas* (the cows) /las#bakas/ with /s/ retention and *las vacas* /las#bakas/ with /s/ deletion are three separate surface categories (i.e., the post-deletion /b/ is somewhere in between the VbV and the VsbV) or whether approximantization is blocked in the latter two contexts in a similar way. If blocking takes place after a retained /s/, we should take it as evidence against post-consonantal weakening in the dialect. If blocking takes place after a deleted /s/ as well, there is evidence for containment, as mentioned above.

The second important question is the alignment of facial gestures with prominence in speech. As mentioned in Section 1.4., orofacial gestures produced during speech mark salient positions important for intonation and expressive language marking. This includes the position of focus. Knowing from previous studies that stop lenition in Spanish is sensitive to stress, we want to see whether this effect is magnified in a syllable that is stressed and in focus, i.e., accented, and

whether lip movement data will show such effects. At the same time, it will be interesting to see whether focus effects are marked both acoustically and articulatorily or only articulatorily. To the best of our knowledge, no such study has been conducted to date.

Given the three main goals of our study, our hypotheses are as follows.[8]

H1A. Target sounds will have a greater intensity difference (i.e., less lenition) in deletion contexts (VsCV and V(s)CV) compared to underlyingly intervocalic position (VCV). In deletion contexts, there will be a blocking effect when the consonant is retained or deleted.[9] In both cases, a greater intensity difference is expected. This can be summarized as an ascending lenition trend: V(s)CV, VsCV < VCV, i.e., V(s)CV, VsCV show *less lenition* than VCV.

H1B. Target sounds will have a smaller lip aperture (i.e., less lenition) in deletion contexts (VsCV and V(s)CV) compared to underlyingly intervocalic position (VCV). In deletion contexts, there will be a blocking effect when the consonant is retained or deleted. In both cases, a smaller lip aperture is expected. This can be summarized as an ascending lenition trend: V(s) CV, VsCV < VCV, as in H1A above.

H2A. Target sounds will have a greater intensity difference (i.e., less lenition) in stressed position (S) compared to unstressed (US), and an even greater intensity difference in stressed syllables in focus (SF). This can be summarized as an ascending lenition trend: SF < S < US in which SF shows *less lenition* than S, which in turn shows *less lenition* than US. The intensity difference gets smaller with more lenition.

H2B. Target sounds will have a smaller lip aperture (i.e., less lenition) in stressed position (S) compared to unstressed (US), and an even smaller lip aperture in stressed syllables in focus (SF). This can be summarized as an ascending lenition trend: SF < S < US, as in H2A above. Here, lip aperture increases with lenition.

H3. There is a correlation between the acoustic measurements (i.e., relative intensity difference) and articulatory measurements (lip aperture parameters) in predicting stop lenition. The greater the lip aperture, the smaller the intensity difference.

In all cases, we assume that there will be a difference between /p/ and /b/ in that there will be less lenition in the voiceless stop compared to the voiced one, as measured by both intensity difference and lip aperture.

---

[8] In all cases, we use '<' to refer to the amount of lenition and not to the measured acoustic or articulatory parameter, which may be in a reverse relationship with the tested conditions.

[9] Following the literature review above, the Spanish dialect under study most probably follows the Central American pattern in which /b d g/ weakening does not take place post-consonantally. We therefore hypothesize that blocking will take place both with the /s/ retained and deleted.

2. Methodology

Participants

A total of 21 participants were recorded (11 females, 10 males). However, 6 of the video files had missing frames and had to be excluded from the analysis, which gave us a total of 15 participant files subjected to motion capture and acoustic analyses. The subjects were all native speakers of the Gran Canarian Spanish variety, living in Las Palmas and in the northern part of the island (Gáldar, Firgas and Agaete), aged 24–55. They received a small remuneration for their participation in the study.

Equipment

The study was conducted in Las Palmas, Gran Canaria in the participants' homes, on the same laptop computer used to display the reading stimuli (Asus TUF Gaming F15 FX506LH-HN129, Intel Core i7, 16GB RAM). Video recordings were made with a Razer Kiyo webcam positioned above the laptop screen. Verbal audio was captured by the laptop's inbuilt microphone which gave better sound quality compared to the camera's mic, given the field conditions of the experiment (limited possibility of eliminating sound reverberation). Video and audio capture were controlled by a copy of OBS Studio, v28.1.2, running on the laptop. Video was captured at 30 frames per second in 1280x720 pixels of spatial resolution, with the auto-focus function disabled. Each experimental session was recorded in its entirety, resulting in raw videos of ~20–30 minutes in length. Additionally, backup audio inputs were recorded using Audacity.[10]

Experimental procedure

The participants sat in front of a computer, each at the same distance from the screen, and were instructed to read out a series of sentences as naturally as possible.[11] Participants controlled the flow of the experiment, using an arrow on the keyboard to pass from one sentence to another in a PDF file opened using Adobe Reader (with one sentence per page). They were instructed to sit as still as possible and not to move their heads during the study, especially not to tilt it sideways or lean forward. The whole procedure lasted around 25 minutes (+/– 3 minutes depending on reading speed and the time needed to understand the instructions). Prior to starting the experimental procedure, the participants gave their informed consent.

---

[10] http://audacityteam.org/.

[11] The distance from the screen was measured by the experimenter (first author). The setup was made such that the computer was fixed on the desk, with the screen placed at the same distance from the chair in which the participant was sitting. It must be noted, however, that the method is not sensitive to changes in distance from the camera as long as there is no tilting of the head or horizontal movements (see Krause et al., 2020 for details).

Materials

The materials used included 376 sentences containing a total of 560 target words starting with /p b/. The aim was to elicit /p b/ lenition in two contexts: intervocalically and after a weakened consonant which may or may not be deleted. The intervocalic context was always the same: both V1 and V2 were /a/. In the trials with deletion contexts, the preceding consonant was always /s/. However, since the deletion of a weakened word-final /s/ is optional in the dialect, we could not predict which speakers would delete the /s/ and in which sentences, although the prediction is that /s/ should be more prone to deletion before voiced obstruents. In any case, our aim was to compare both the deletion context with the intervocalic (non-deletion) context and the intervocalic context with different strategies in the deletion context, i.e., retention of a weakened consonant vs. its actual non-pronunciation.

The second factor that we wanted to explore was prosody, hence the words used were of different types: The initial /p b/ were either in a stressed or in an unstressed position. Additionally, to look at the prosodic effects beyond word stress, we placed the words either in the initial part of the sentence (the subject position in most cases) or toward the end of the sentence, where focus (and the main sentence stress) is expected to fall in Spanish. The different placement of target words in a sentence allowed for using up to two target words per utterance and decreasing the number of sentences needed to obtain a good signal-to-noise ratio for the motion capture analysis. Also, we did not want the reading task to be overly long or boring, hence, apart from limiting the number of sentences to read, we used unpredictable repetitions, i.e., the number of repetitions of each sentence, and hence words, differed, ranging from 4 to 10. Nevertheless, we made sure that the number of words belonging to each condition was the same, i.e., 80 repetitions of each stress/prosodic condition per sound and a total of 80 repetitions of the deletion context conditions. Three randomized lists were created and administered randomly to the participants. Some examples of sentences used in the study are presented in **Table 2** below.

Sentence	Translation	Condition	Sound
*La **barrera** estaba mal colocada y el portero no veía.*	'The wall was incorrectly placed and the goalkeeper could not see'	US	/b/
*La **paciencia** de esa mujer me tenía impresionado.*	'The patience of this woman had me impressed'	US	/p/
*La **banda** de música empezó el concierto con La **Bamba**.*	'The music band started the concert with La Bamba'	S, SF	/b/, /b/

(Contd.)

Sentence	Translation	Condition	Sound
*La **paga** mensual es **más baja** de lo que pensaba **Paco**.*	'The monthly pay is less than what Paco thought'	S, DEL, SF	/p/, /b/, /p/
*La **vaca** de Juan cuesta mucha **pasta**.*	'Juan's cow costs a lot of money'	S, SF	/b/, /p/
***Las Vacas** Locas es una banda de música de Tenerife.*	'The Mad Cows is a music band from Tenerife'	DEL	/b/

Table 2: Examples of sentences used in the study. Consonants placed in a deletion context were marked 'del', those in unstressed position were marked 'US', in stressed syllables – 'S', and stressed in focus – as 'SF'.

Data preprocessing

Audio data were extracted from the video (see 2.2 below) and then annotated in Praat (Boersma & Weenink 2022). The resultant TextGrids included a sound tier, a condition tier, a tier including information on deletion (whether it occurred, if applicable), as well as a word and a sentence tier. Only the sequences of interest were annotated, i.e., the VCV sequence or the VsCV sequence. Annotations were provided by one junior annotator and then corrected for time alignment and other details by the first author.[12] Note that the number of sounds produced in each condition was different for each speaker. First, speakers were instructed to repeat the whole sentence, should they make a mistake during reading. This produced additional target words whenever the speaker made a mistake elsewhere in the sentence than where the target word was placed (the participants did not know which words were important for the experimenter). On the other hand, some words had to be excluded because it was impossible to identify the approximant acoustically between vowels.[13] All such cases were /b/ sounds, usually produced in the sequences

---

[12] For VsCV sequences, it had to be determined whether a given sound was deleted or not. The sounds were first auditorily categorized and then, based on visual inspection of the spectrograms and waveforms, the first author of this paper determined whether frication noise was present in the middle and/or higher frequency ranges. Note that /s/ undergoes weakening in the dialect, which usually consists in debuccalization to [h] or [ɦ] of variable length. More information on the exact distribution of these output segments is available in Broś (2022). For the purposes of this study, we had to decide whether the /s/ was apparently deleted or retained. If there was no acoustic trace of the /s/ after the vowel (i.e., no frication nor presence of aperiodic noise immediately following the vowel on the waveform), we deemed the segment deleted (compare Figs. 16 and 17). We also annotated glottal fricatives as a category separate from full [s], as such sounds are usually shorter and less perceptible. We did not distinguish between voiced and voiceless glottal sounds, as voicing depends on the following segment and is not relevant for the measurements taken in this study.

[13] Overall, the expected number of observations per participant was 560, which should give us a total of 8,400 observations. The dataset contained 8,184 observations, i.e., 216 fewer than expected.

la baba (the saliva) or *la vacunación* (the vaccination). Given that both flanking vowels are the same, /b/ deletion is very likely in spontaneous speech in the dialect, and for some of our participants, it was quite prevalent even in read speech. Examples of TextGrid annotations are provided in the Appendix.

2.1. Acoustic measurements

The TextGrids prepared at the preprocessing stage were used to extract acoustic data with a custom Praat script prepared by the first author. We extracted information on the duration and intensity of each segment (both minimum and maximum) together with the corresponding condition, word and sentence. The database was later checked for annotation inconsistencies and manually corrected. The rest of the operations were performed in R (R Core Team, 2020). This included the calculation of intensity difference. Two alternative calculations were made – intensity difference A measured the difference in intensity between V1 and the consonant, whereas intensity difference B measured the difference between the consonant and V2.[14]

2.2. Motion capture

We collected motion capture data for this study using the following general strategy: We took digital video recordings of participants' faces while they read aloud the stimulus sentences. We then isolated the video of the target VCV and VsCV sequences and processed it using the OpenFace 2.0 face-tracking system (Baltrušaitis et al., 2018). We used the resulting facial coordinate data to compute the dependent variables of interest for each trial.

As mentioned above, video recordings were made in OBS Studio, using the Razer Kiyo webcam. The final video files were encoded to .mp4 format for later processing. In some cases, several experiment sessions were run back-to-back, causing participant audio and video to come slightly out of sync, due to memory leakage in OBS Studio. This de-synchronization was monitored for and corrected at a later processing step (see below).

A custom Python script segmented the session videos to isolate the key VCV sequences. This script obtained the beginning and ending time points of each VCV sequence from the Praat TextGrid files produced during acoustic coding (see previous section). The script was first used to generate a handful of video segments for a given participant. These segments were then visually inspected to ensure that the appropriate lip movements appeared in their entirety. (Because the cut points were based on acoustic information coded into the TextGrids, de-synchronization of audio and video could result in cutting off part of the critical movement.)

---

[14] The results will be presented for intensity difference A only. We briefly comment on the implications of the other parameter in the Discussion.

If movement was systematically cut off, the Python script was modified to begin cutting slightly (usually ~150ms) before or after the cut points indicated in the TextGrid. Once the initial video segments appeared satisfactory, the script was used to isolate the remaining VCV / VsCV segments. The file names of these segments preserved important trial information, including participant ID, the experimental condition of the sentence, the target word produced, whether the critical (phonological) consonant was /p/ or /b/, and whether the consonant was deleted during production.

The segmented videos were then processed using OpenFace 2.0. OpenFace is a neural network tool that automatically detects the most prominent face in each frame of a digital video and determines the coordinate positions of several key points. Although based on a two-dimensional image, these coordinates are expressed in an estimated three-dimensional space based on orthographic camera projection. The accuracy of this projection can be improved if OpenFace is provided with the intrinsic lens parameters of the camera used to take the image. Although that was not done in this case, absolute accuracy (i.e., correspondence to real-world units) is not essential to the case we are making here. Since all analyses are repeated measures, the most important requirement is the internal consistency of OpenFace's coordinate estimates within each participant. Therefore, all reports of distances in millimeters should be understood as reflecting OpenFace's *estimated* millimeters.

We used a second custom Python script to parse the OpenFace output files and generate a summary datasheet for each participant's motion tracking data. Key descriptive and predictor variables from the acoustic coding were first extracted from the video segment file names. Then the coordinate values at each frame of the segment were processed to generate several new dependent variables. Of central importance were the positions of the upper-middle and lower-middle points of the inner lips (OpenFace parameters 62 and 66, respectively, see **Figure 1**). At each video frame, the script computed vertical lip aperture by taking the Euclidean distance between these points in the (x, y) coordinate plane. These values were used as the basis of the following additional DVs.

2.2.1. Maximum lip aperture

The script identified the first maximum in lip aperture, which was presumed to correspond to the peak separation of the lips during the first vowel.

2.2.2. Minimum lip aperture

The script identified the minimum in lip aperture, which was presumed to correspond to the tightest closure of the lips during the bilabial consonant.
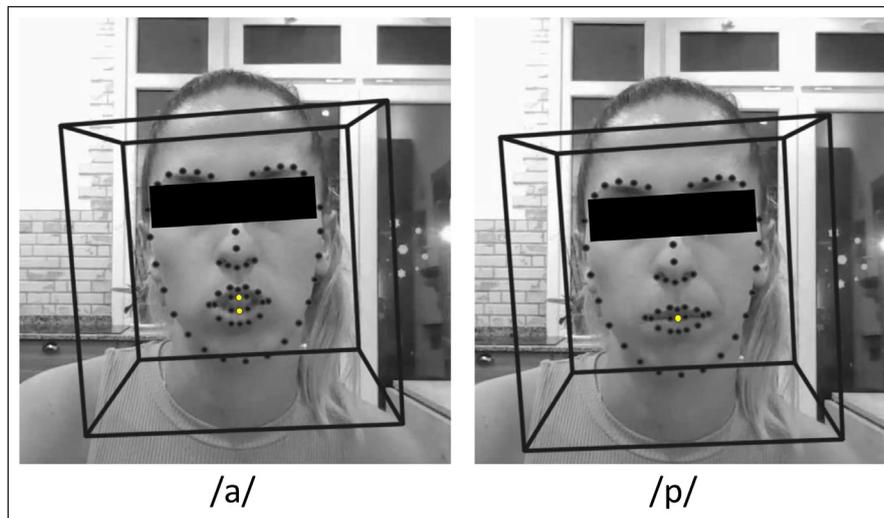
Figure 1: Two motion-tracked frames depicting the OpenFace key points. These frames were extracted from a real production of the phrase 'mucha pasta,' in which the obstruent was fully realized. The left-hand panel depicts the oral configuration during the /a/ of *mucha*, and the right-hand panel depicts the oral configuration during the /p/ of *pasta*. Parameters 62 and 66, which were used in computing vertical lip aperture, are highlighted in yellow.

2.2.3. Difference between maximum and minimum (i.e., relative aperture)

The script calculated the delta between the prior two values. Hereafter, we will refer to this variable as relative aperture, due to its surface similarity with the acoustic variable of relative intensity.

2.2.4. Maximum closure speed

Operating in the window of time between the maximum and minimum lip apertures, the script computed the change in aperture between each pair of adjacent frames, and then identified the largest *decrease*. The absolute value of this decrease was presumed to correspond to the maximum speed achieved during the bilabial closure.

2.2.5. Time-normalized lip aperture trajectory

To facilitate visualization of the lip-closure kinematics, the script also computed time-normalized trajectories of the vertical lip aperture over the course of each video segment. Time normalization aids in visually comparing these trajectories because each analyzed sequence (and hence each video segment) was not of a uniform duration. The script divided the time separating the first and last frames of a sequence into 10 equally sized intervals (each representing 10% of sequence duration). It then estimated the vertical lip aperture at the end of each interval via linear interpolation.

Graphs of these time-normalized trajectories, organized by independent variables of interest, appear in **Figures 2** and **3**.
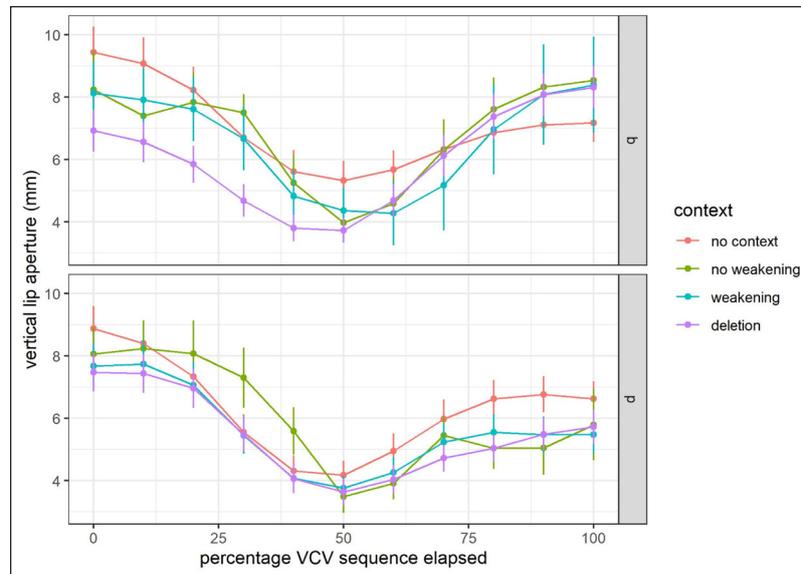
Figure 2: Graph of vertical lip aperture trajectories as a function of sequence type. Values are based on raw means, time-normalized to 11 time steps via linear interpolation. Error bars: +/− 1 SEM, calculated for the mean of participant means. 'No context' is equivalent to intervocalic position (VCV), 'no weakening' refers to /p b/ preceded by non-weakened /s/ (VsCV), 'weakening' refers to /p b/ preceded by /s/ weakened to [h] (VhCV), and 'deletion' refers to a V(s)CV sequence in which the /s/ is not pronounced.
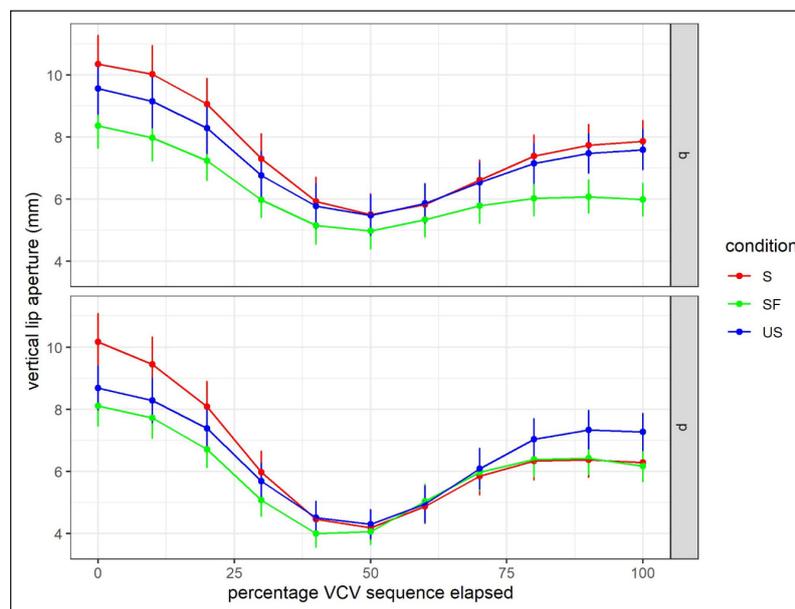
Figure 3: Graph of vertical lip aperture trajectories as a function of condition (after removing trials in the deletion context). Values are based on raw means, time-normalized to 11 time steps via linear interpolation. Error bars: +/− 1 SEM, calculated for the mean of participant means. S refers to stressed position, SF refers to stress in focus position, and US refers to unstressed position.

2.3. Statistical analyses

Statistical analyses were conducted using the packages *lme4* (Bates et al., 2018), for building models, and *emmeans* (Lenth, 2019), for the calculation of simple effects. Descriptive plots and effects plots were generated using the *ggplot2* (Wickham, 2016) and *ggeffects* (Lüdecke, 2018) packages, respectively.

3. Results

In this section we present the results of the acoustic analysis as these are the point of departure for subsequent statistical steps (3.1.). We then explore the correlation between the key acoustic parameter, i.e., relative intensity difference and the articulatory measurements to determine which measurements obtained in the course of lip movement analysis from video data are the most suitable when looking at the degree of stop weakening (3.2.). After this step, we present the statistical analysis of the articulatory data (3.3.).

3.1. Acoustic data analysis

To estimate the degree of lenition depending on the consonant and condition, we first looked at the deletion context, which was an independent variable with four levels: no deletion context (i.e., underlying VCV, annotated as 'no context'), deletion context (V(s)CV) in which deletion occurred ('deletion'), deletion context (VsCV) in which the /s/ was weakened to [h] or [ɦ] ('weakening'), and deletion context in which deletion failed to occur and no weakening happened instead, i.e., /s/ was produced as [s] ('no weakening'). We fit a linear mixed effect model with consonant (*p/b*), /s/-deletion context and their interaction as predictors, and intensity difference A as a dependent variable. The participant's ID and target word were clustering variables. After the adjustment of the initial maximal model structure aimed at obtaining a non-singular model that converged, the model fit was the following:

$$intensity\ difference\ A \sim consonant * context\ +\ (1 + consonant\ |participant)\ +\ (1|word)$$

The model was run on a total of 8,184 observations, of which 7,077 represented a no deletion context (i.e., underlying VCV); 406 had non-deleted preceding segments (VsCV), 127 of which were retained as /s/ and 279 weakened to [h]; and a further 701 had deleted preceding segments (V(s)CV). We performed *F*-testing of fixed effects, obtaining denominator degrees of freedom via Satterthwaite approximation. The results show main effects of consonant, $F(1, 37.97) = 99.5$, $p < 0.001$, and context, $F(3, 277.6) = 253.17$, $p < 0.001$, and a significant interaction between the two ($F(3, 277.49) = 3.719$, $p < 0.05$).

We then estimated marginal means from the model to assess the simple effects of context and to take a closer look at the interaction, testing via *t*-test, and adjusting for familywise error via Tukey's method. The simple effects are as follows: The intensity difference of the segment

is significantly smaller in no deletion contexts (underlying VCV) compared to both the deletion contexts in which the preceding sound was retained (VsCV, $t = -25.99$, df $= 197.1$, $p < 0.001$ when retained as [s], and $t = -21.34$, df $= 259.6$, $p < 0.001$ when weakened to [h]) and deleted (V(s)CV, $t = -21.99$, df $= 30.1$, $p < 0.001$). At the same time, the intensity difference calculated for the segment in deletion contexts with preceding consonant retention (VsCV) was significantly greater than when the preceding consonant was deleted (V(s)CV, $t = 11.40$, df $= 8151.4$, $p < 0.001$ in the case of the [s], and $t = 6.27$, df $= 8136.5$, $p < 0.001$ in the case of [h]). The difference between [s] and [h] outputs also resulted significant ($t = 3.17$, df $= 8118.6$, $p = 0.008$).

As for the interaction, all contrasts were significant except for the intensity difference of the /p/ in weakened /s/ vs consonants retained as [s] ($t = 0.379$, df $= 8117.8$, $p = 0.99$). **Table 3** presents the marginal means for the interaction term and the interaction effects are shown graphically on **Figure 4**.

Consonant	context	emmean	SE	Df	lower.CL	upper.CL
/b/	no context	5.36	0.548	28.9	4.24	6.48
/p/	no context	18.05	1.345	15.8	15.20	20.90
/b/	no weakening	24.02	0.829	127.9	22.38	25.66
/p/	no weakening	33.82	1.687	37.7	30.41	30.41
/b/	weakening	20.09	1.196	540.8	17.74	22.44
/p/	weakening	33.53	1.534	25.8	30.37	36.68
/b/	deletion	16.93	0.713	70.5	15.51	18.35
/p/	deletion	29.81	1.547	26.7	26.63	32.98

Table 3: Marginal means of intensity difference for the interaction between consonant and context in the first model. SE = standard error, df = degrees of freedom, CL = confidence levels.

To explore the effects of stress and focus on the relative intensity of the consonant, we excluded the deletion contexts from the database, which gave us a total of 7,077 observations, quite evenly distributed between the three levels of the condition variable (S = 2,170, US = 2,532, SF = 2,375). We then fitted a linear mixed effects model with consonant, condition and their interaction as predictors, and participant ID and target word as clustering variables. The resultant model that allowed for convergence was in the form

intensity difference A ~ consonant * condition + (1+consonant+condition|participant) + (1|word)
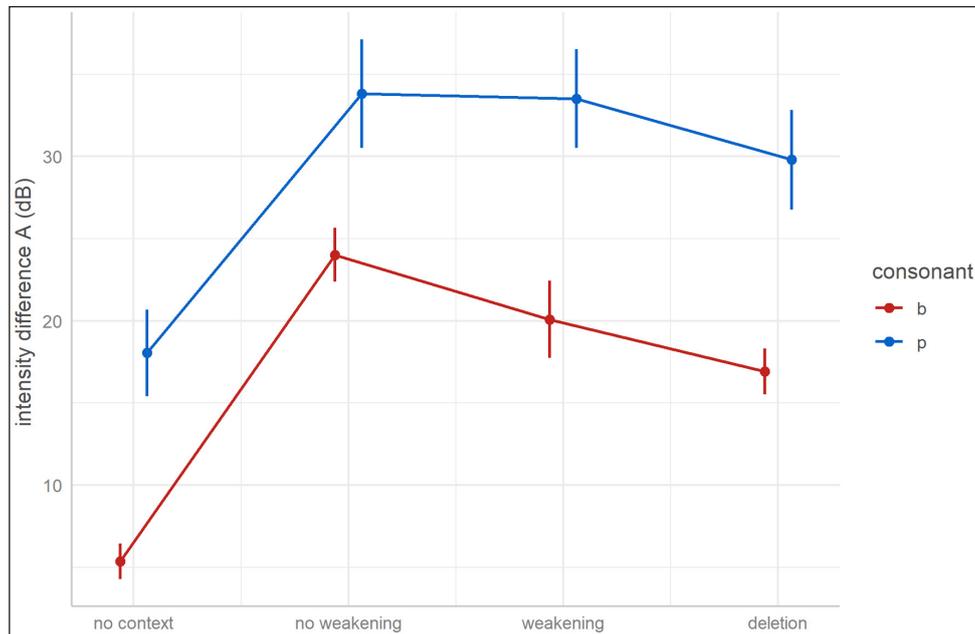
Figure 4: Effects plot, based on the estimated marginal means of the mixed-effects model, of the interaction between consonant and /s/-deletion context in predicting intensity difference A. (Error bars: 95% CI.)

The results show that the effect of consonant is significant ($F(1, 15.61) = 137.78, p < 0.001$). There was also a significant effect of condition ($F(2, 23.25) = 24.80, p < 0.001$) and a significant interaction between consonant and condition ($F(2, 134.31) = 10.708, p < 0.001$).

As in the deletion model, we estimated marginal means from the model. In this case, we wanted to explore the simple effects of condition and the interaction effect in more detail. The simple effects for condition show that target sounds in S position have a smaller intensity difference compared to the SF position, but the effect is not significant ($t = –1.95, df = 14.5, p = 0.16$). At the same time, they have a significantly greater intensity difference compared to US ($t = 5.18, df = 34.2, p < 0.001$). Also, consonants in SF position have a significantly greater intensity difference compared to US ($t = 6.66, df = 28.1, p < 0.001$). This means that there is less lenition in SF than in US contexts.

The pairwise comparisons for the interaction demonstrate that underlying /p/ has a greater intensity difference in SF position compared to both /p/ and /b/ in all other positions, and these effects are significant in all but one case: The comparison with /p/ in S position yielded insignificant results ($t = –2.739, df = 18.7, p = 0.113$). As for underlying /p/ in S position, it shows a greater intensity difference compared to /p/ in US position ($t = 6.040, df = 58.9, p < 0.001$), and compared to /b/ in S position ($t = –11.573, df = 17.5, p < 0.001$), SF position ($t = 9.587, df = 16.1, p < 0.001$), and US position ($t = 11.19, df = 16.2, p < 0.001$). Finally, /b/ has a significantly lower value of intensity difference in US position compared to /b/ in SF

position ($t = 3.171$, $df = 46.0$, $p < 0.05$), but not compared to /b/ in S position ($t = 2.561$, $df = 48.8$, $p = 0.126$). There is also no significant difference between /b/ in S and SF positions ($t = -0.945$, $df = 17.7$, $p = 0.929$). **Table 4** presents the *emmeans* for all consonants and condition levels, and **Figure 5** presents the interaction term graphically.

Consonant	condition	Emmean	SE	Df	lower.CL	upper.CL
/b/	S	5.91	0.56	21.8	4.75	7.08
/p/	S	19.02	1.47	14.8	15.88	22.15
/b/	SF	6.28	0.44	28.8	5.37	7.19
/p/	SF	20.09	1.29	14.9	17.33	22.85
/b/	US	4.77	0.43	26.8	3.87	5.67
/p/	US	16.58	1.32	14.8	13.76	19.39

Table 4: Marginal means of relative intensity for the interaction between consonant and condition in the second model; SE = standard error, df = degrees of freedom, CL = confidence levels. S – stressed position, US – unstressed position, SF – stressed in focus.
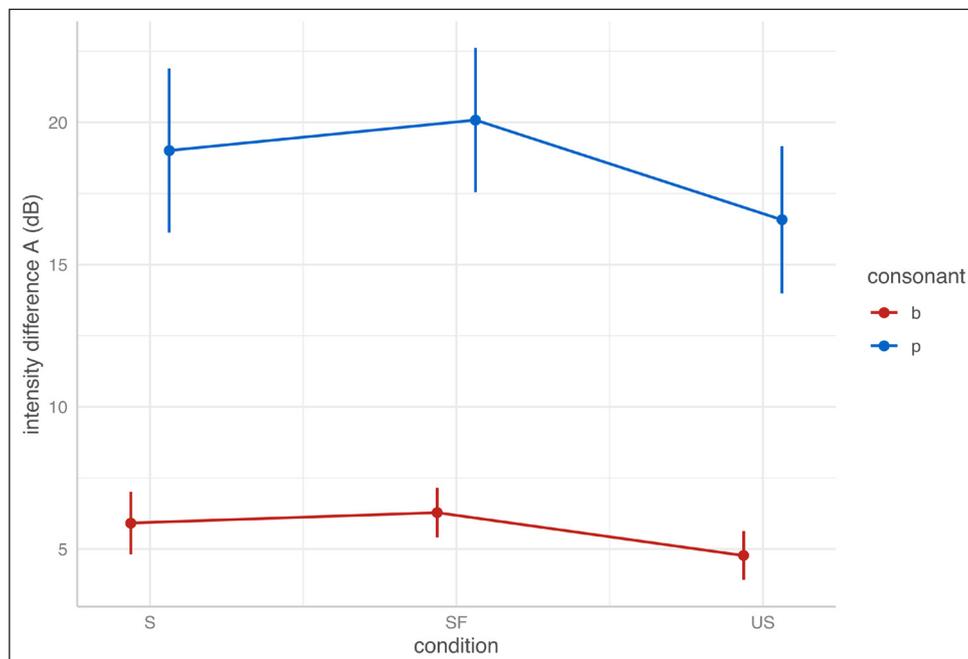
Figure 5: Effects plot, based on the estimated marginal means of the mixed-effects model, of the interaction between consonant and condition (S, SF, US) in predicting intensity difference A. (Error bars: 95% CI.)

3.2. Correlating the acoustics with the articulatory measurements

We next evaluated which of our measured articulatory variables most reliably covaried with acoustic intensity difference A. Intensity difference is an established means of assessing lenition. However, to our knowledge perhaps only two relevant articulatory studies have been performed to date.[15] Parrell (2011) used minimum lip aperture (dubbed "maximum constriction") to assess approximantization of Spanish labial consonants in different phonetic contexts. As for a direct comparison of the articulatory and acoustic metrics, Parrell (2010) presented a small-scale study on two subjects, showing a negative correlation between constriction degree and intensity difference. Similarly, Hualde, Shosted and Scarpace (2011) showed a good correlation between acoustic intensity and articulatory data based on electropalatography in intervocalic contexts for the coronal stop in Spanish.[16] The results of these studies suggest that a similar relationship might be revealed in our data.[17]

A priori, there appear theoretically well-motivated reasons to consider at least three of our measured variables: minimum lip aperture, relative aperture, and maximum closure speed. All three variables might be expected to predict intensity difference on the assumption that, for bilabial consonants, quieter/more consonant-like acoustic performance arises when the lips are brought closer together. On its face, relative aperture would seem most similar to (relative) intensity difference. However, while a canonical bilabial closure requires the lips to touch, most of the vowels in our inventory should be tolerant of a wide range of vertical apertures. This might reduce the explanatory power of relative aperture and/or maximum closure speed since both variables depend on the specific magnitude of maximum aperture.[18]

We made our determination empirically, using mixed-effects multiple regression. Because intensity difference, relative aperture, maximum closure speed, and minimum aperture all showed substantial positive skewness, we first transformed all four variables. Minimum aperture and relative aperture were converted to their binary logarithms. Relative intensity and maximum closure speed were converted to their square roots, as this yielded somewhat better distributions. Where necessary, small constants were added to the distributions to eliminate negative values

---

[15] But confer Romero (1995) who compared fricatives and approximants in Andalusian Spanish.

[16] Note, however, that they recommend caution when looking at stops preceded by /l/ or nasals, in which case acoustic data does not reflect articulation that well. This should be considered in future studies comparing the effects of different preceding sounds.

[17] Note, however, that while Parrell's aim was to measure the accuracy of different acoustic metrics in predicting constriction, our aim is to look at how well video-based lip movements correlate with the well-established lenition marker (i.e., intensity).

[18] A similar concern was raised in Section 1.3. in relation to the acoustic measurements. While intensity difference gives us a relative value, it also depends on the intensity of the vowel. For this reason, we controlled the vowels in the experiment, and we decided to use the preceding instead of the following vowel for calculations to assess the effect of stress and focus more reliably.

prior to transformation. For the remainder of this paper, references to these four variables should be taken as references to their transformed versions.

In the model, acoustic intensity difference was entered as the outcome variable, while minimum lip aperture, relative aperture, and maximum closure speed were entered as numeric predictors. Consonant (/b/ vs. /p/) was also entered as a categorical control variable. Participant ID and target word were entered as clustering variables. The random effects structure was determined via backwards selection from the maximal model. Backwards selection prioritized first de-correlating random intercepts and slopes and then eliminating random slopes – higher-order terms first, until a convergent, non-singular model resulted. The final model had the form

intensityDifference ~ consonant + minimumAperture + relativeAperture + maximumClosureSpeed + (consonant + minimumAperture + relativeAperture + maximumClosureSpeed || participant) + (minimumAperture + relativeAperture || word)

where the double-bar notation indicates that slope-intercept correlations have not been estimated for a given cluster. Despite correlations arising between multiple fixed effects, the model was not markedly collinear (all VIFs < 1.5). This model was run on a total of 8,184 observations.

T-tests of model slopes (using Satterthwaite-approximated degrees of freedom) revealed that of the three predictors of interest, only minimum lip aperture reliably differed from 0, estimate: –0.118, SE = 0.022, $t(28.62)$ = –5.36, p < .001. Unsurprisingly, the control variable of consonant was also reliable, estimate (using /p/ as the reference value): 1.248, SE = 0.159, $t(58.92)$ = 7.86, p < .001. An effects plot showing how minimum lip aperture predicts intensity difference appears in **Figure 6**.

3.3. Motion capture data analysis

We next examined the effects of the experimental manipulations on articulation. Following the findings described in the prior section, we used minimum lip aperture (binary log transformed) as the dependent variable of interest.

As we had for the acoustic results, we first tested the effects of the /s/-deletion context, treating it as a four-level predictor: trials with no deletion context, trials with deletion context but no recorded deletion and no weakening, trials with deletion context but no recorded deletion and weakening, and trials with deletion context and recorded deletion (i.e., the *no context, no weakening, weakening* and *deletion* levels, respectively). We fit a linear mixed-effects model with consonant, deletion context, and their interaction as predictor variables, and participant ID and target word as clustering variables. After model simplification, the final fitted model was of the form
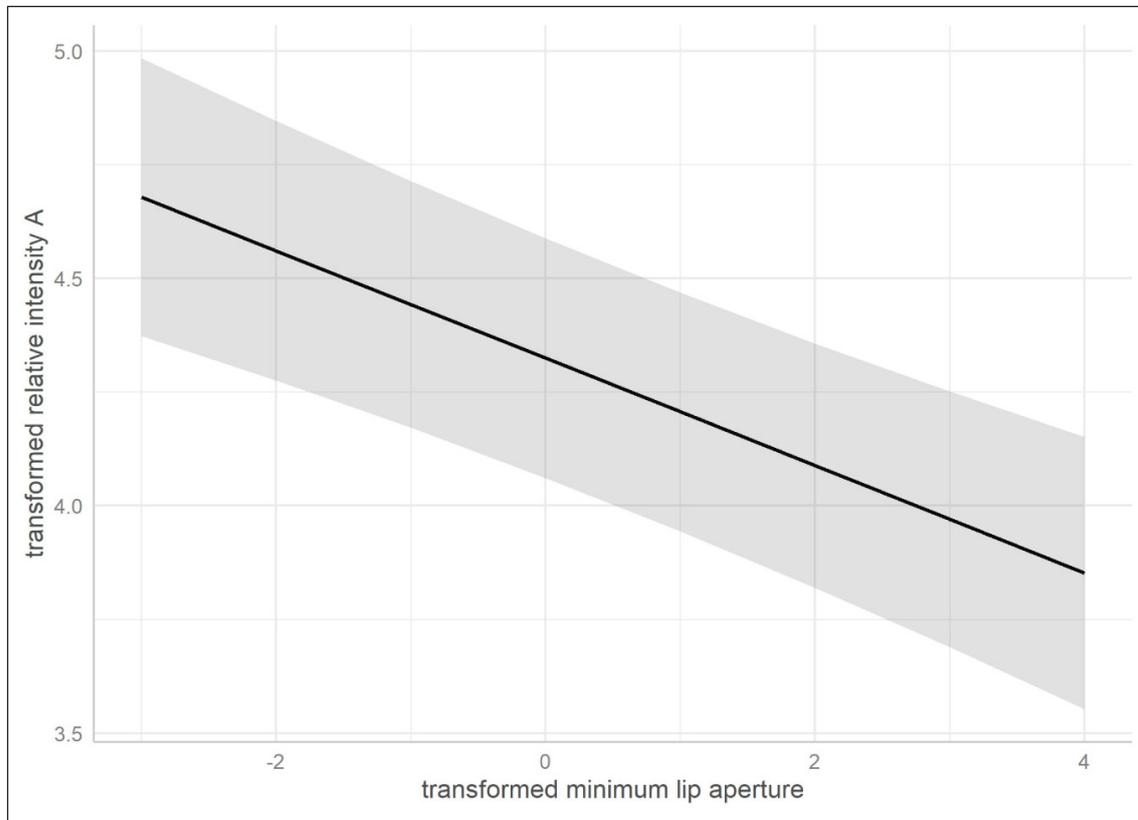
Figure 6: Effects plot, based on the estimated marginal means of the mixed-effects model, of how minimum lip aperture predicts intensity difference A. (Minimum lip aperture has been converted to its binary logarithm, and intensity difference A to its square root. Error band: 95% confidence.)

minimumLipAperture ~ consonant + deletion_context + consonant* deletion_context + (consonant + deletion_context + consonant* deletion_context|| participant) + (1 | word)

As was the case for the acoustic analysis, this model was run on 8,184 observations (7,077 VCV, 127 V[s]CV, 279 V[h]CV, and 701 V(s)CV). Because the deletion context predictor had four levels, we performed *F*-testing of fixed effects, obtaining denominator degrees of freedom via Satterthwaite approximation. The main effect of deletion context was reliable, $F(3, 35.96) = 14.09, p < .001$. The interaction was also reliable, suggesting the effect of deletion was moderated by consonant, $F(3, 230.10) = 11.30, p < .001$.

We estimated marginal means from the model to explore the interaction. We tested the resulting simple effects via t-test, using Satterthwaite-approximated degrees of freedom. Familywise error was contained using Tukey's method. For /p/-initial target words, none of the pairwise comparisons were reliable. For /b/-initial words, the effects were more nuanced. When /s/ was fully deleted, the resulting minimum apertures did not statistically differ from

those cases in which /s/ was fully retained or weakened to [h]. Trials in the no context level had larger minimum apertures than trials where /s/ was weakened, estimate: 0.883, SE = 0.113, $t(362.00) = 7.78$, $p < .001$, and trials where /s/ was fully deleted, estimate: 0.531, SE = 0.123, $t(21.9) = 4.32$, $p < .01$. Trials in the no context level may also have resulted in larger minimum lip apertures than trials where /s/ was fully retained, but the result was on the margin of reliability, estimate: 0.317, SE = 0.110, $t(12.90) = 2.87$, $p = 0.56$. Finally, trials in which /s/ was fully retained resulted in larger minimum lip apertures than those in which /s/ was weakened, estimate: 0.566, SE = 0.138, $t(30.20) = 4.10$, $p < .01$. **Table 5** presents the marginal means for the interaction term. An effects plot of the interaction appears in **Figure 7**. Note that the trends in this plot, as well as the fact that a reliable contrast arises between retained and weakened /s/ but *not* between retained and deleted /s/, suggesting that lip apertures might be quite small on weakened trials. We will revisit this point in the Discussion.

Consonant	Deletion	Emmean	SE	Df	lower.CL	upper.CL
/b/	no context	1.91	0.204	15.0	1.48	2.35
/p/	no context	1.55	0.229	22.3	1.08	2.03
/b/	no weakening	1.60	0.229	21.8	1.12	2.07
/p/	no weakening	1.37	0.267	37.6	0.83	1.91
/b/	weakening	1.03	0.231	24.6	0.56	1.51
/p/	weakening	1.40	0.246	29.5	0.90	1.91
/b/	deletion	1.38	0.235	24.5	0.90	1.87
/p/	deletion	1.45	0.269	38.5	0.91	2.00

Table 5: Marginal means of minimum lip aperture (binary-log-transformed) for the interaction between consonant and deletion. SE = standard error, df = degrees of freedom, CL = confidence levels.

Next, we removed trials in the deletion context from the dataset and examined the effects of the remaining conditions (7,077 total observations, S = 2,170, US = 2,532, SF = 2,375). We treated condition as a three-level independent variable: trials where target word was unstressed, trials where target word was stressed, and trials where target word was both stressed and focused (i.e., the *U, S,* and *SF* conditions, respectively). We fit a linear mixed-effects model with consonant, condition, and their interaction as independent variables, and participant ID and target word as clustering variables. After model simplification, the final fitted model was of the form
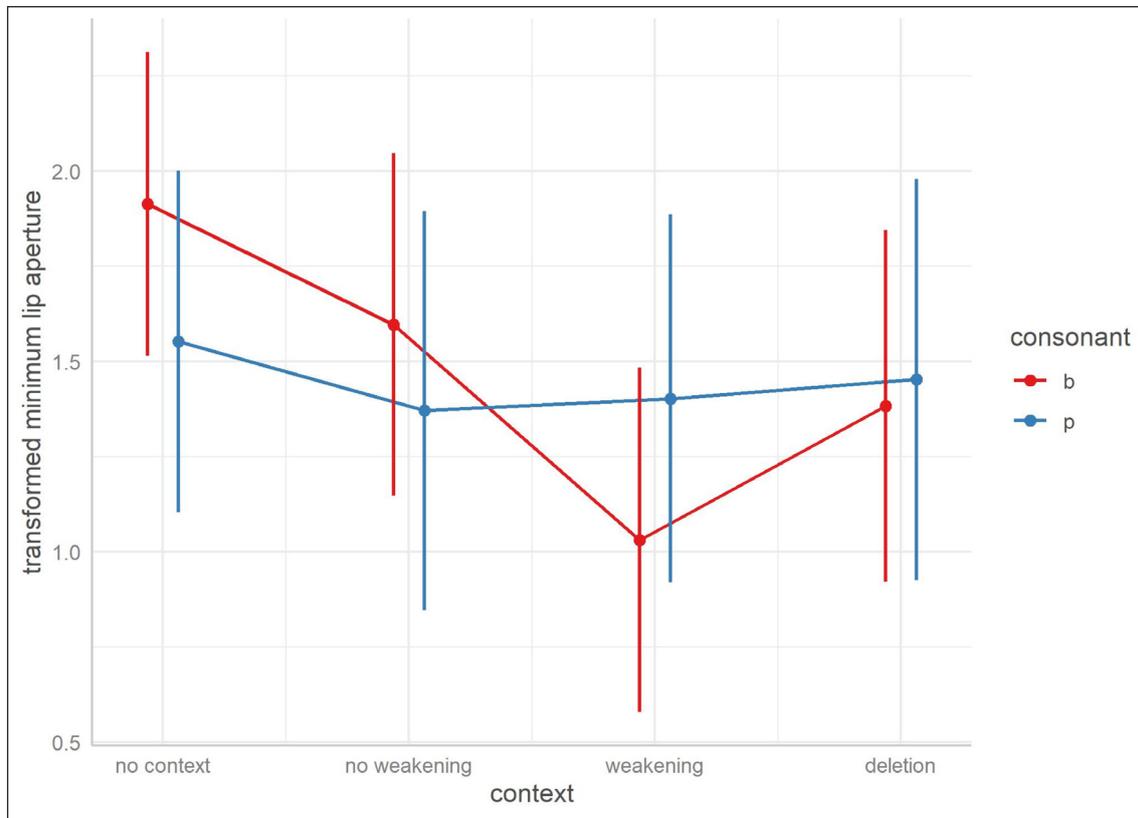
Figure 7: Effects plot, based on the estimated marginal means of the mixed-effects model, of the interaction between consonant and /s/-deletion context in predicting minimum lip aperture. (Minimum lip aperture has been converted to its binary logarithm. Error bars: 95% CI.)

minimumLipAperture ~ consonant + condition + consonant*condition + (consonant + condition | participant) + (condition | word)

Because the condition variable had three levels, we again performed *F*-testing of fixed effects. The interaction was not reliable. However, the main effect of condition was reliable, $F(2, 32.73) = 4.92, p = .014$. The main effect of consonant was also reliable, $F(1, 23.17) = 7.54, p = .011$.

We estimated marginal means from the model to assess the simple effects of condition, testing the simple effects via *t*-test, and adjusting for familywise error via Tukey's method. SF trials resulted in reliably smaller minimum lip apertures than S trials, estimate: –0.084, SE = 0.032, $t(21.7) = –2.60, p = 0.042$. SF trials may also have resulted in smaller minimum lip apertures than US trials, though the effect was on the margin of reliability, estimate: –0.134, SE = 0.056, $t(58.00) = –2.39, p = 0.052$. US and S trials did not differ from each other ($p = 0.661$). **Table 6** presents the marginal means for the main effect of condition. An effects plot of the main effect of condition appears in **Figure 8**.

Condition	Emmean	SE	Df	lower.CL	upper.CL
S	1.72	0.18	15.7	1.33	2.11
SF	1.63	0.18	15.4	1.25	2.01
US	1.77	0.18	15.4	1.38	2.15

Table 6: Marginal means of minimum lip aperture (binary-log-transformed) for the main effect of condition. SE = standard error, df = degrees of freedom, CL = confidence levels.
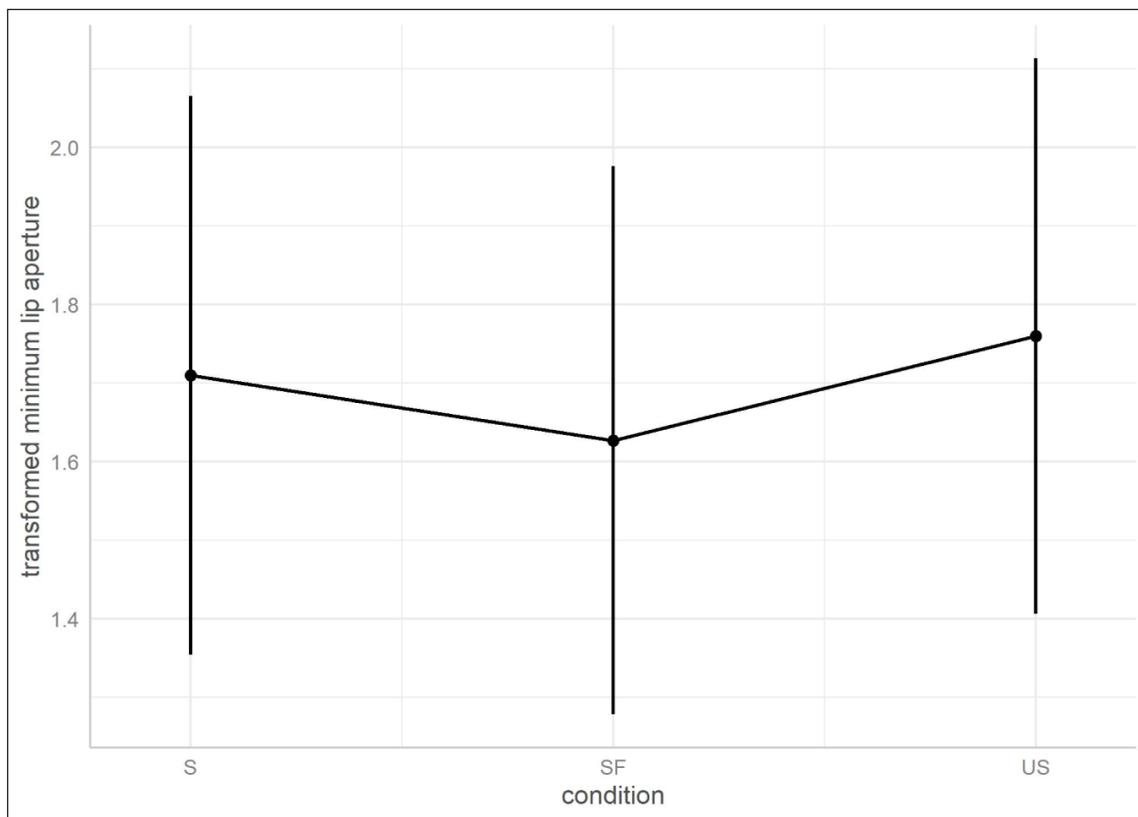
Figure 8: Effects plot, based on the estimated marginal means of the mixed-effects model, of the main effect of condition on minimum lip aperture. (Minimum lip aperture has been converted to its binary logarithm. Error bars: 95% CI.)

3.4. Bilabial closure timing in /s/ deletion contexts

Though not the primary focus of this study, the articulation of the bilabial closure may add relevant evidence to the question of whether the phonetic /s/ ([h]) is fully deleted, is truncated, or is masked by coproduction with /p b/. For example, 'early closure' might appear to be suggested

by **Figure 2**, in which the closure phase for full deletion trials appears to cover a much larger part of the early timeframe, compared to other contexts (at least for /b/ words). However, one should not be too hasty to draw conclusions from **Figure 2**, because it is plotted over proportional time. Judged in terms of the overall duration of the utterance, the bilabial closure will *always* be delayed for cases of /s/ retention, relative to cases of /s/ deletion, regardless of the underlying deletion mechanism. This is because in cases of /s/ retention, the implementation of the /s/ has been 'slotted' into the time between the first vowel and the implementation of /p b/ closure. Therefore, more specific follow-up was called for. Note that, since we have no direct measure of either tongue-tip or glottal constrictions, any evidence gleaned from this exercise is necessarily incomplete.

We generated a new Python script that detected kinematic landmarks of bilabial closure in the vertical lip aperture trajectory data. These landmarks were based on those estimated via the *findgest* algorithm in Mview (Tiede, 2005), although the specific implementation was new to our script. Briefly, the script searched for the moments of peak closure velocity and peak opening (release) velocity following the closure. During the closure phase, gestural onset was logged as the first moment at which the closure velocity rose above 20% of peak, and target achievement was logged as the first moment at which closure velocity fell beneath 20% of peak.

Gestural onset was frequently logged as 0 ms, creating a sizeable floor in the distribution, and making it an undesirable dependent variable. We therefore fit a linear mixed-effects model with target achievement time (binary-log-transformed) as the dependent variable, and *consonant, deletion context,* and their interaction as independent variables. The main effect of deletion context was reliable, $F(3, 16.61) = 5.39, p < .01$. After Tukey correction, two reliable pairwise contrasts were observed: between the *no context* and *weakening* levels, and between the *no context* and *deletion* levels (see **Figure 9**).

Since we modeled target achievement time, rather than gesture onset time, one might wonder if the delays in the *weakening* and *deletion* contexts (relative to *no context)* reflect additional time needed to form a tighter closure (as opposed to the approximantized closures observed for /b/ in the *no context* level). To address this concern, we estimated a follow-up model testing the linear association of minimum lip aperture with target achievement time (using consonant type as a control factor). However, none of the estimated effects were statistically reliable (all $ps > .2$).

We also fit a linear mixed-effects model testing how consonant and deletion context predicted acoustic duration of the overall phonetic sequence. There was a reliable main effect of deletion context, $F(3, 31.20) = 38.32, p < .001$. All pairwise comparisons were reliable after Tukey correction (see **Figure 10**).
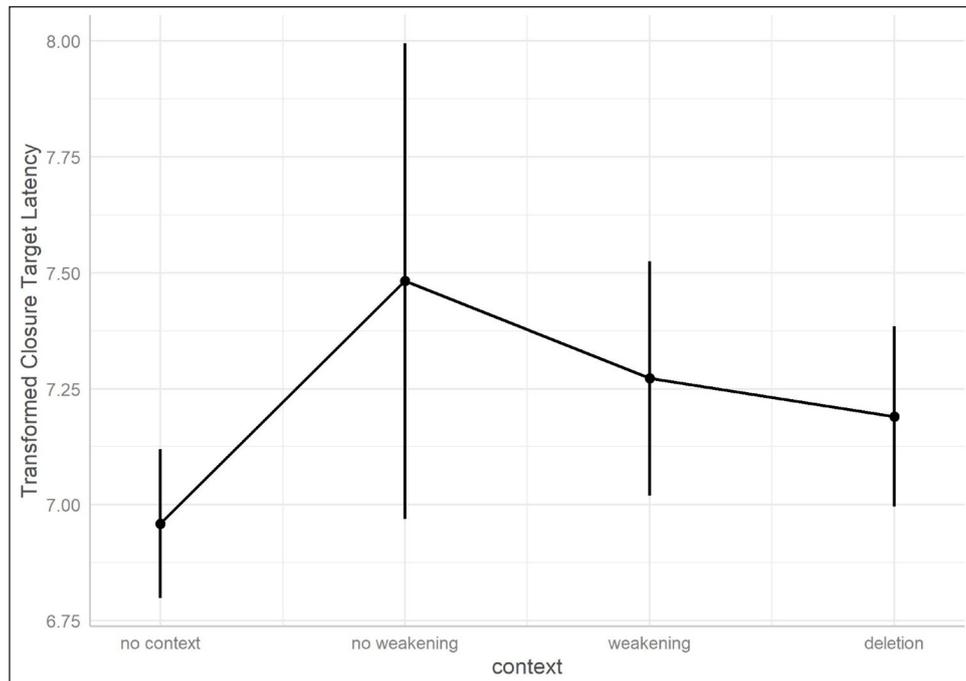
Figure 9: Effects plot, based on the estimated marginal means of the mixed-effects model, of the main effect of deletion context on target achievement timing for the bilabial closure. (Error bars: 95% CI.)
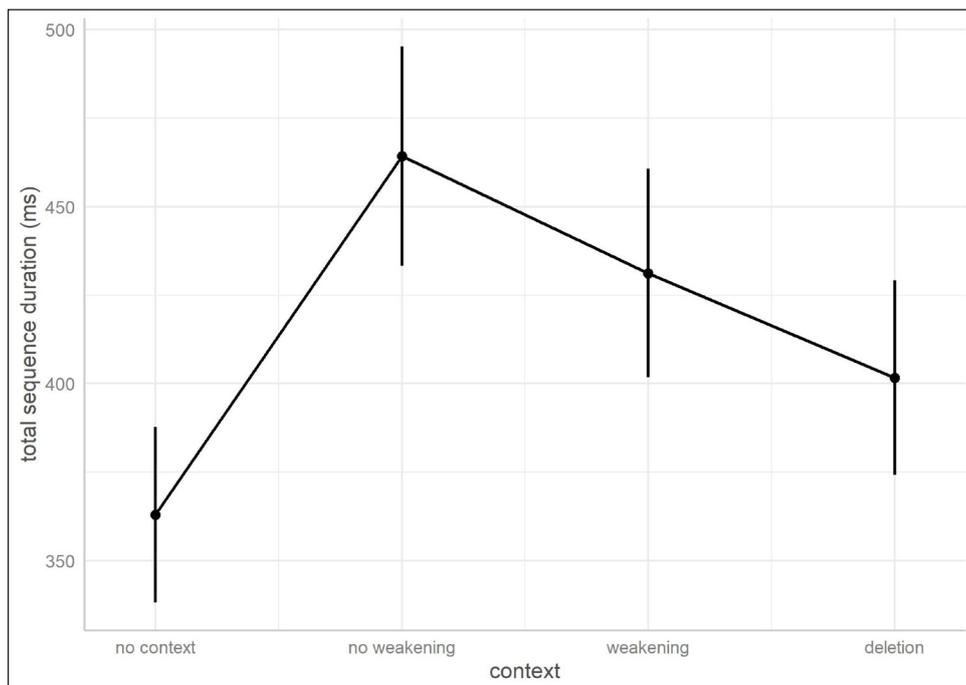
Figure 10: Effects plot, based on the estimated marginal means of the mixed-effects model, of the main effect of deletion context on phonetic sequence duration. (Error bars: 95% CI.)

We suggest that these differences in sequence duration likely arise due to one or more of the following issues. First, there may be differences in overall speech rate. Perhaps /s/ retention is more likely at lower rates of speech. Second, it makes sense for the acoustic duration of the /s/ to be "missing" from V(s)CV tokens. It is worth noting that the average 60-ms difference between the *no weakening* and *deletion* levels is roughly equivalent to the average acoustic duration of [s] in *no weakening* cases. Intriguingly, the durations of *deletion* and *no context* cases themselves differed, suggesting that the timing of V(s)CV and 'true' VCV tokens were not identical. This was further borne out by our finding that the bilabial closure target was achieved later for *deletion* cases than for *no context* cases, which might suggest that some inaudible trace of the /s/ remains phonetically present in *deletion* trials. If true, the present data cannot adjudicate whether the underlying gestures are truncated or 'masked' by coproduced /p b/ gestures (or both).

4. Discussion

To ensure optimal argumentation clarity, the results of the study are discussed in the order of the hypotheses presented in Section 1.5., i.e., in accordance with the three main goals set for the project.

4.1. Phonological effects – Hypotheses 1A and 1B

Hypotheses 1A and 1B stated that the target sound will have a greater intensity difference and a smaller lip aperture (i.e., less lenition) in deletion contexts compared to underlying VCV and that it will be the same in V(s)CV contexts, where the preceding consonant is deleted, and VsCV, where it is retained.

The results of the study show that there are significant differences in stop weakening depending on the phonological environment, as expected based on previous literature. Thus, stops in VCV sequences are significantly more weakened than stops in either V(s)CV or VsCV sequences. This was confirmed by acoustic and articulatory measurements.

As for the difference between the deletion contexts, there is a discrepancy between the acoustic and the motion capture analysis. Intensity difference shows slightly smaller values after the deletion of a preceding consonant (V(s)CV) compared to sequences with a retained /s/. This is against our expectations: We assumed that there should be no difference because weakening is not expected in either case. Importantly, the results tell us that stops do not weaken at all after the fricative [s] (according to previous studies, an intensity difference of 29 dB on average is equivalent to a full, usually voiceless stop pronunciation). Furthermore, it seems that the non-weakening environment is maintained both when the /s/ is weakened to a glottal fricative (around 27 dB on average) and after consonant deletion (23 dB).[19] These results point to a

---

[19] We base these numbers on the marginal means from the model (see **Table 3**).

'qualitative' gradient effect: The greatest lenition blocking[20] happens after a full-fledged alveolar fricative and dissipates in line with the degree of lenition of the underlying /s/. When there is only partial weakening by loss of oral constriction, there is less of a blocking effect and even less so with full consonant deletion. Nonetheless, the resultant sounds are nowhere near lenited consonants of the approximant type in the case of /b/ (around 5–10 dB) or voiced stop in the case of /p/ (around 15–18 dB). Post-consonantally, the outputs of /p/ do not fall below 29 dB of average intensity difference and the outputs of /b/ are not lower than 16 dB in this respect.

The discrepancy between the behaviors of /b/ and /p/ is also worth noting. The values of intensity difference are the same in voiceless stops regardless of whether the underlying /s/ was retained unchanged or realized as a glottal fricative. However, we can see a difference between there being some remnant of the /s/ and full deletion in both /p/ and /b/. The latter suggests some kind of a transition stage between underlying and derived VCV sequences. Further studies are needed to explore stop weakening after other consonants, e.g., sonorants, although it is worth noting that our data from this and other studies contain some instances of stops preceded by /l/ in which no lenition ensues, e.g., *el perro* (the dog), *salvajes* (wild). It is therefore quite plausible that both voiced and voiceless stop lenition in Canarian Spanish is post-vocalic only. All in all, we can say that Hypothesis 1A was partially confirmed.

As for the motion capture data, they show a difference in lip aperture between the VCV and deletion contexts, and no reliable difference between VsCV with a full [s] or glottal [h] and V(s)CV (with /s/ deletion). It is also worth noting that the discovered effects were only reliable in the case of /b/, suggesting a different behavior of voiced vs voiceless stops.[21] Moreover, there is also a surprising difference between sequences with /s/ weakening to [h] and /s/ deletion on the one hand, and sequences with a retained [s]. The trend suggests that /b/ preceded by a weakened /s/ may be the most constricted, which is contrary to the findings based on acoustic intensity. The stop preceded by an [s] is different from /b/ in underlying VCV sequences but seems to show an intermediate effect, which might be compatible with the hypothesis that there is some weakening after consonants but not as much as in post-vocalic contexts, in line with earlier acoustic studies of some dialects of Spanish. Unfortunately, we have no previous studies to compare our results to and hence we cannot say whether the minimum lip aperture values shown in the marginal means

---

[20] We use the term 'lenition blocking' here because the values of intensity difference suggest that there is no lenition in VsCV sequences regardless of the output of /s/. In the case of /p/, the values correspond to a voiceless stop while in the case of /b/ the values correspond to voiced stops rather than approximants, although in the latter case these stops may have a weaker constriction, at least after a deleted /s/.

[21] The lack of a reliable difference between deletion and no deletion contexts for /p/ may also be due to the limitation of the method. Perhaps lip closure cannot be measured as precisely in voiceless stops, whereas acoustic measurements are more sensitive to slight differences in muscle tension or voicing. Alternatively, the difference may lie in the type of lenition applied. In the case of /p/ it is mostly voicing while in /b/ it is full-on approximantization. In the former case, we may have a reduction in the degree of lip compression and the method used may be too coarse for capturing it.

table roughly correspond to a weakened (perhaps approximantized) variant of /b/ or simply a weaker stop pronunciation.

All in all, Hypothesis 1B was also partially confirmed. If we combine this result with the lip aperture trajectories from **Figure 2**, we can see that the level of aperture is indeed the same regardless of whether or not the preceding /s/ was deleted. A slight discrepancy between the trajectories can be observed in the part of the sequence before the target stop, where we see a slower drop in lip opening when /s/ is retained due to the presence of some form of a consonantal gesture (be it [s] or [h]). Interestingly, in /b/, the trajectories of V[h]CV and V[s]CV are aligned and differ from V(s)CV, whereas in /p/ there is alignment between V(s)CV and V[h]CV, with V[s]CV showing a different temporal path. These discrepancies should, however, be treated with caution given that time normalization results in distortions that may make visual comparisons unreliable. Another difference in the trajectories consists in the initial aperture of the lips depending on the context: The starting (vowel) level is lower when the vowel is flanked by a coda consonant, regardless of whether the consonant is pronounced or not.

All in all, these results lead to two important generalizations. First, there is no stop weakening after /s/ in the studied variety of Spanish, which is in line with some earlier experimental studies on different Spanish dialects (see Eddington 2011 for a review) and, indirectly, with some of the earlier reports on Gran Canarian Spanish (Almeida, 1982). This applies both to preceding alveolar fricatives and their weakened allophones. There is also a blocking effect of a preceding deleted /s/, which confirms previous findings, starting from reports of strengthened or tense voiced stops following /s/ deletion in Canarian Spanish (Trujillo, 1980, Dorta & Herrera, 1993) and followed by reports of unweakened voiced (and voiceless) stops following consonantal deletions in Gran Canarian (Broś et al., 2021) and some other Spanish dialects (e.g., Honduran, Amastae, 1989 or Costa Rican, Carrasco et al., 2012).

Second, the data are not entirely clear as to the phonological difference between contexts with /s/ deletion, /s/ weakening and /s/ retention. While acoustic data seem to point to a somewhat gradual weakening pattern and perhaps an intermediate category of stops produced in derived VCV contexts as opposed to underlying VCV on the one hand, and post-/s/ VsCV positions on the other, articulatory results are not fully in line with the acoustics, and /p/ does not necessarily follow the same pattern as /b/. While no differences have been detected for /p/ in motion capture data, perhaps due to the excessive coarseness of the method, the data for /b/ suggest a different pattern, i.e., a weakening/deletion blocking effect as opposed to (perhaps) some degree of weakening (an intermediate variant) after a full [s]. Nonetheless, taken together, the differences between the three deletion contexts are very small compared to the difference between VCV and deletion contexts in general and do not seem to reflect categorical changes, such as the change from a voiceless to a voiced stop or a change from a voiced stop to an approximant. Thus, we may look at them as gradient changes caused by articulatory overlap.

As for the phonological interpretation of the deletion context results, we can say that /p b/ weakening is blocked after consonants, and this blocking effect persists even after the consonant is deleted phonetically. There are several interpretations of this result. First, the apparently deleted segment is still there phonologically, i.e., it is not deleted completely, which makes the preceding vowel phonetically not adjacent to the stop consonant. Thus, the stop consonant behaves as if it was standing after a non-deleted segment and does not weaken. Another interpretation might involve gestural overlap in the phonetics component: the /s/ is not deleted completely and is instead masked by the adjacent stop gesture. As a result, no /s/ is phonetically present on the surface and the stop does not weaken because it stands after a consonant and not a vowel. Given the methods used in the experiment we cannot fully exclude the gestural overlap interpretation in cases of apparent deletion. On the one hand, we did find that the temporal trajectories in VsCV vs V(s)CV sequences differ by an average of 60ms in sequence duration, which roughly corresponds to the duration of the [s] in the dataset. This would suggest that the acoustic deletion of /s/ is complete, i.e., that no /s/ gesture occurs at all in these sequences. On the other hand, the durations of VCV and V(s)CV differ slightly and the timing of reaching the target for bilabial closure is delayed in the latter case. This might suggest that some remnant of the /s/ gesture is there nonetheless (see Section 3.4.). Thus, we leave the question of gestural overlap open for the time being.

Another question concerns gestural coordination in the sequence. In line with our data, the gestures in V(s)CV sequences are not coordinated in the same way as gestures in underlying VCV sequences. If full deletion ensues, we might expect gestural reorganization, which would lead to lenition in derived intervocalic contexts, yet this does not happen. Similar effects were reported by Shaw and Kawahara (2023), for instance, who analyzed gestural coordination in CVC sequences compared to CC sequences in which the intervening vowel was deleted in Japanese. They found a blocking effect in those consonantal contexts in which vowel weakening is less frequent, with no gestural reorganization and the preservation of the temporal structure of the sequence despite deletion. One of the interpretations put forward by Shaw and Kawahara (2023) is that less consistency in an optional process of vowel deletion leads to the preservation of a weakly activated gesture that inhibits the coordination between the now-adjacent consonants. Although our case looks at consonant deletion and vowel-to-consonant gestural coordination, there is a possible parallel. Since /s/ weakening and deletion is variable and optional across speakers, we might assume that the structural coordination between the vowel and /p b/ is maintained as with an intervening segment even in the phonetic absence of the /s/. Some support for this can be found in the delayed timing of reaching the bilabial closure in V(s)CV compared to VCV reported in Section 3.4.

The above interpretation might find its phonological equivalent in gradient symbolic representations (Smolensky & Goldrick, 2016; Zimmermann, 2019) which show gradient activation in phonology. Note that there seems to be a gradual change in the intensity difference

of /b/ depending on the featural representation of the /s/. While a full alveolar [s] triggers a fully voiced occlusion, weakening to [h] (weaker activation) changes the [b] in the direction of a weakened stop, and /s/ deletion (weaker still) leads to a pronunciation that is even closer to an approximant. However, these gradual changes do not lead to approximantization and hence do not involve a featural change in phonological terms. Besides, articulatory data do not confirm these effects; neither were they found for /p/. Thus, we are more inclined to determine that, in the absence of more convincing evidence, there seems to be no phonological difference between contexts with /s/ deletion and /s/ retention. The deleted consonant behaves as if it was still there, blocking the process of weakening in the same way as a retained consonant. As a result, a derived intervocalic environment leads to a different phonetic representation than an underlying one. The difference between the two environments is not a phonetic but a phonological (structural) one.

Phonological theory has dealt with the above problem on numerous occasions, based on the assumption that structural relations are not necessarily equivalent to surface phonological or phonetic relations. It is therefore possible to encounter some types of covert structure in the form of a root node that is left phonologically unparsed (Prince & Smolensky, 1993/2004) or that is projected but not necessarily pronounced (Goldrick, 1998). The usual situation, in which no deletion ensues, requires reciprocity between the projection and pronunciation lines connecting the underlying representation with the surface representation, i.e., full bidirectionality. Unpronounced projections are not visible nor perceptible on the surface because they cannot be interpreted phonetically. A schematic representation of our three types of sequences taking into account covert structure is presented in **Figure 11** below.
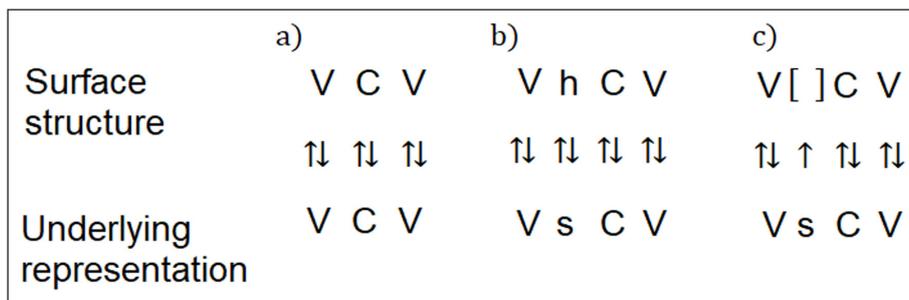
Figure 11: A schematic representation of the three types of sequences analyzed in the study. Projection lines are represented by upward arrows, pronunciation lines by downward arrows; segments correspond to lexical representations. Note that the pronunciation of the /s/ as [s] leads to the same structural representation as in b).

The above representations can be used as a basis for a formal analysis, which can be easily operationalized in one of the leading phonological frameworks. Looking at the opaque blocking effect in Gran Canarian Spanish, Broś (2016), for instance, uses a simplified version of Colored

Containment (van Oostendorp, 2006) and shows a successful analysis of the non-weakening of /p t k/ in deletion contexts under Optimality Theory. Although the assumptions made in Broś' paper were purely phonological, the present study confirms their correctness in capturing the blocking effect via acoustic and articulatory analysis. The preceding consonant may be unpronounced but still blocks stop lenition and hence must be present in the phonological output structure (the root node and its projection line are still there).

An interesting implication of a containment-based analysis is that it leaves the question of whether a given segment was deleted (no articulatory movement was made/planned) or masked by another gesture open. Note that the phonetic component interprets the output of phonological computation. If a given segment is there structurally (it contains a projection line) but is not pronounced, as in **Figure 11c**, this is treated differently than both an absent segment (**Figure 11a**) and a projected and pronounced segment (**Figure 11b**) and may be realized phonetically as gestural overlap. Otherwise, the lack of the pronunciation line might be interpreted as a lack of an /s/ gesture (target) and hence a phonetic VCV sequence. The unpronounced output of phonology might give rise to different, perhaps gradient phonetic interpretations, which is supported by the slight acoustic differences between sequences with retained and deleted /s/ in our data. Such an in-between option was also proposed by van Oostendorp (2008) in his analysis of incomplete neutralization in stop devoicing observed in many languages of the world. Thus, the containment approach does not allow for a formal distinction between the two types of phonetic interpretations of the data, which is an important issue raised by this journal's associate editor. Perhaps, ultimately, all deletion is due to gestural overlap, which is something that should be studied further in Spanish and other languages.[22]

4.2. Prosodic effects – Hypotheses 2A and 2B

Hypotheses 2A and 2B stated that target sounds should be characterized by a greater intensity difference and a smaller lip aperture (i.e., less lenition) in stressed vs. unstressed positions, and these tendencies should be even stronger in focus vs. no focus, resulting in a US > S > SF lenition trend.

In this context, the second part of our analysis revealed a key discrepancy between acoustic and articulatory measurements. The acoustic analysis showed that there is a difference between stops in stressed vs. unstressed positions in terms of degree of weakening. At the same time, there seems to be no difference between the two types of stressed syllables: S and SF. Thus, Hypothesis 2A is only partially supported. The motion capture data, on the other hand, demonstrate a

---

[22] Some evidence for this can be found e.g., in studies on English /t d/ deletion, according to which the lack of an articulatory target may happen but seems to be rare, and in apparent deletion identified as /t d/ being inaudible, the tongue movement corresponding to the coronal gesture does take place but has a smaller magnitude compared to an audible consonant (Purse, 2019).

significant difference between the focus position and the other two contexts, but no differences between S and US, contrary to our expectations. Hypothesis 2B was partially supported. All in all, the results suggest that lip movements show higher-level effects, such as sentence prosody (here: stress in focus position), which is compatible with previous studies on orofacial gestures produced during speech, while acoustic measurements are more sensitive to word-level processes and lexical as opposed to sentence stress.

Another explanation of the results might be sought in experiment design. While lexical stress was easy to elicit, focus was determined syntactically, on the expectation that speakers would put main emphasis on the rightmost word/phrase in the sentence, in accordance with the rules of nuclear stress placement and prominence in broad focus in Spanish (cf. Section 1.2.). Given the nature of the study, all the sentences were constructed as simple or complex statements, and we did not manipulate sentence structure or try to elicit narrow focus in a particular position e.g., by asking speakers to answer questions about particular parts of the sentence. We also did not control the number of words or follow the same syntactic structure in all sentences to provide a stimulus set, ensuring a more natural variability in /p b/ production. Consequently, it is possible that the speakers did not use neutral focus marking at least in some cases, which may have affected our results. The prominence of the consonant in S and SF conditions may have blended to some extent. Since we could confirm the effect of focus in the motion capture data, however, it seems that our expectations concerning intonation and prominence marking were borne out. If we look at the second vowel of the sequence depending on the condition, the data are in line with the expectation that the nuclear stress should be characterized by greater vowel duration in Spanish (see **Figure 12**).

At the same time, it must be noted that the intensity of the second vowel is the lowest in SF. Although this parameter is not a sentence prominence marker in Spanish and plays a greater role at the level of lexical stress distinctions instead (see e.g., Ortega-Llebaria & Prieto, 2007), the low intensity of the SF vowel may be an indication of an overlap between stress and accent with boundary effects. Domain finality, such as the end of a sentence or utterance, has been associated with various types of prosodic weakness and articulatory reduction, including accent retraction, pitch lowering, etc. (Cheng & Kisseberth, 1979; Hyman, 1977; Beckman & Pierrehumbert, 1986), which may have some impact on the intensity-based measurements centered on lenition.[23] All in

---

[23] Note that differences in vowel intensity directly affect the intensity difference, as signaled in Section 1.3. Since those differences are more marked in the second vowel, which may be stressed or unstressed, as well as bear nuclear stress, we decided to use the first vowel of the sequence to calculate the relative intensity difference. In this case, raw measurements also show that the maximum intensity of the vowel is the lowest in SF, yet the overall differences between positions are much smaller. As for the overall characteristics of the recorded sentences, our general observations are convergent with the expected pattern, i.e., a falling pitch contour (downstep) toward the end of the sentence as in typical Spanish declaratives and no pitch rise in nuclear position. We also see that maximum intensity levels fall at the end of the utterance (in the last phrase).
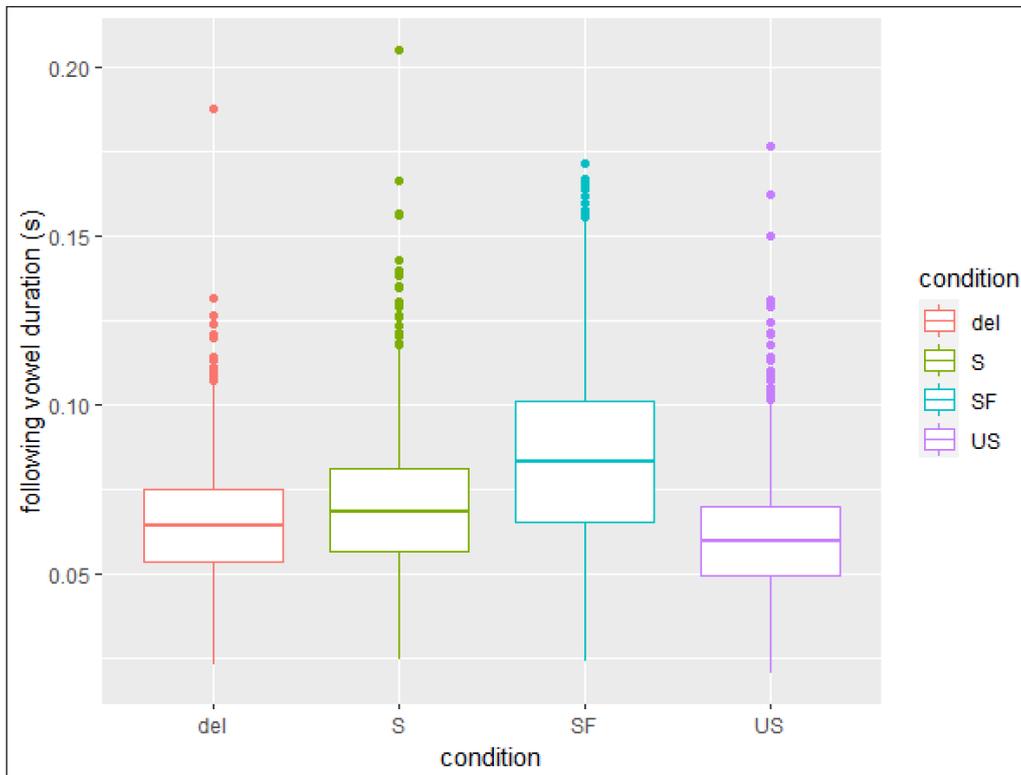
Figure 12: The duration of the second vowel depending on the condition. Note that this vowel is the longest in SF, and the shortest in US.

all, the acoustic data show sensitivity to intensity differences in stressed vs. unstressed syllables, in line with previous literature, and no discrimination of SF vs. S based on this parameter, also as expected (since accent is cued by duration and pitch rather than intensity in Spanish). Articulatory data, on the other hand, show sensitivity to the focus position. Thus, the cues relevant to stress and accent in Spanish are apparently reflected not only in the vowels, but also in onset consonants produced as a part of the studied syllables. Finally, it is worth mentioning that the discrepancy between the acoustics and articulation in capturing stress and accent effects in stop lenition in Canarian Spanish finds a parallel in at least one previous study. Shaw et al. (2020) found that acoustic and articulatory data are correlated in velar lenition in Iwaidja, as in this study, and that they seem to provide overlapping but distinct channels of information about lenition.

4.3. The reliability of lip measurements vis à vis acoustics – Hypothesis 3

Finally, Hypothesis 3 assumed that there is a correlation between the acoustic measurements and articulatory measurements in predicting stop lenition: the greater the lip aperture, the smaller the intensity difference. Our study showed that motion capture data obtained via lip tracking from outputs recorded with an internet camera can be successfully used to predict stop lenition. There

is a significant negative correlation between one of the three tested articulatory parameters, i.e., minimum lip aperture, and its acoustic equivalent, i.e., relative intensity difference. Thus, hypothesis 3 has found support in our data.

It follows from the above that lip measurements obtained from video recordings are fine-grained enough to accurately predict the degree of weakening in stops in intervocalic position. Furthermore, they not only constitute a viable, easy-to-implement alternative method of studying lenition, they also provide additional information on those aspects of speech which have not been captured by the acoustics, such as nuclear stress/focus effects (see Section 4.2.). As a further extension, they can be used to explore other prosodic characteristics of segmental phenomena and therefore be recommended to linguists interested in linking segmental phonology with intonation research.

Note that intensity, measured in dB, has been demonstrated to reliably show even the slightest differences between sounds undergoing lenition. For instance, Broś et al. (2021) showed that underlying voiced stops realized as [b d g] are significantly different from underlying voiceless stops that underwent voicing by lenition, the difference between them being in the order of a few dB of relative intensity. In our data, we did not label surface realizations as either lenited or not, voiced or voiceless, or approximants. We relied on the underlying specifications only. However, the acoustic data help us categorize these sounds based on the predictions made in previous studies. In **Figure 13**, we can see the intensity difference calculated for the target sounds depending on the condition. Judging by the mean values, we can conclude that /b/ is produced as an approximant in all VCV contexts (relative intensity below 10 dB) and as a stop in the deletion context (values close to 20 dB). At the same time /p/ is in the range of a voiced or a partially voiced stop in VCV contexts (15–22 dB), and in the range of a tense voiceless stop in deletion contexts.

Against this background, the reader may wonder how much precision and discrimination potential can be found in lip measurements made at a distance. A similar comparison with raw lip measurements as a discriminating factor shows that while differences in lip aperture are much smaller than the relative differences in intensity, the former reliably distinguish different degrees of weakening, which is exemplified by /b/ (see **Figure 14**). As for /p/, the differences between unweakened /p/ and its partially weakened or voiced counterparts, for example, are perhaps physically too small to capture by lip tracking. This was also pointed out in Section 3.

It is also worth mentioning that intensity measurements are a more general metric that 'sieves in' more than just differences in the constriction of the produced sounds. By contrast, lip aperture is by definition a consequence of fewer speech-relevant kinematic degrees of freedom than is acoustic intensity. Looking at this metric can only speak to relative consonant aperture and not e.g., voicing. Thus, any mismatch between the two metrics is possibly driven by the fact that intensity differences result from degrees of freedom other than lips and jaw.
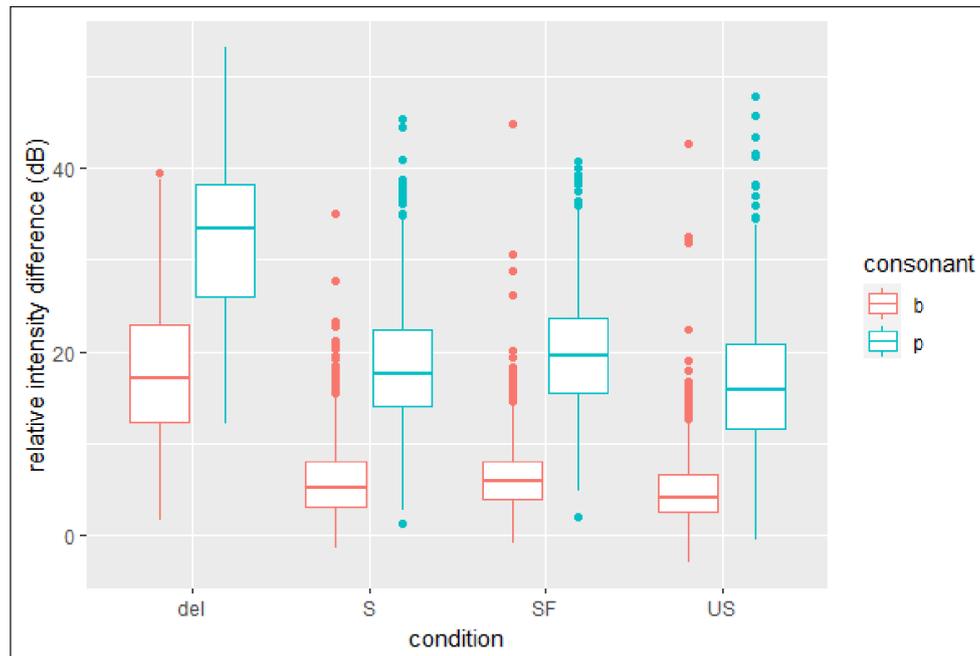
Figure 13: Intensity difference of the target sound by context and condition.
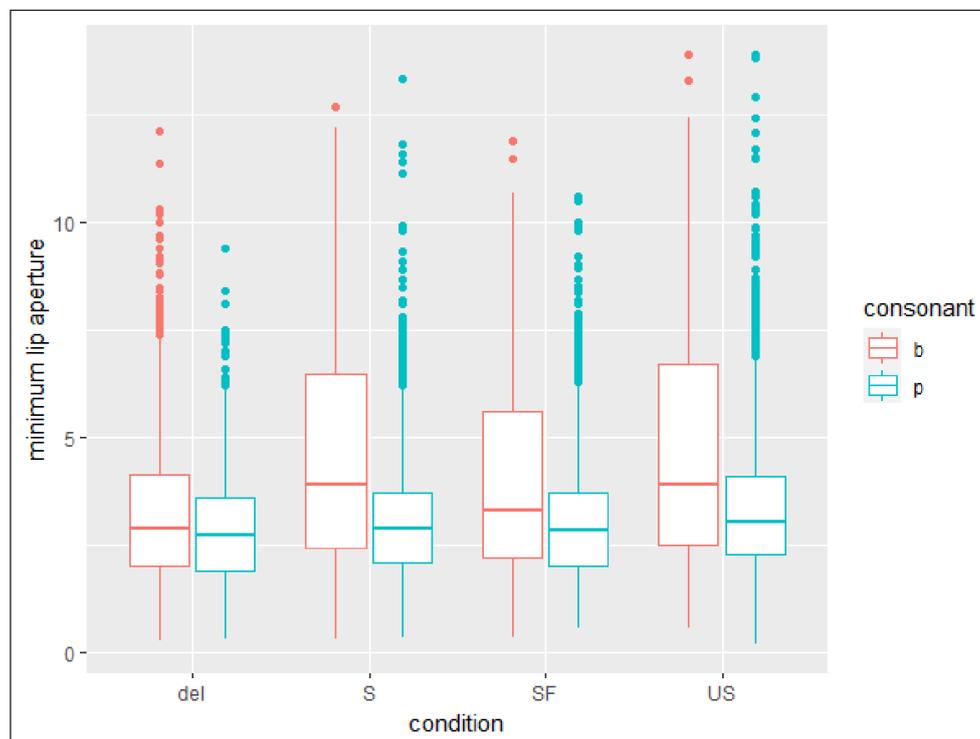
Figure 14: Minimum lip aperture for the target sound by context and condition.

To summarize the comparison between the two methods, given the intra- and inter-category discrepancies in the predictions allowed by the two types of measurements, we can conclude that they are complementary in nature and that each one of them constitutes an added value. For this reason, we encourage researchers to use both acoustic and articulatory data in future studies on consonant lenition.

5. Conclusion

In this paper, we have shown an analysis of voiced and voiceless stop lenition in Canarian Spanish focused on phonological and prosodic effects.

The data point to a reliable difference between post-consonantal and post-vocalic positions in the dialect, with no lenition in the former case. They also demonstrate that the deletion of a preceding consonant does not induce lenition, although the target sound becomes intervocalic in the process. Lenition blocking in such derived environments supports theoretical analyses assuming containment. The two findings confirm some of the previous reports on this and certain Latin American dialects of Spanish that diverge from the majority stop lenition pattern.

As for the prosodic effects, we have confirmed that stress inhibits stop lenition. Additionally, we have shown an effect of focus, with greater constriction of the target sound in accented syllables.

Finally, we have shown that motion capture data recorded with an easy-to-implement technique reliably predict consonant aperture and can be used alongside acoustic measurements to study lenition degree. Furthermore, lip measurements obtained from video recordings are not only correlated with the acoustics, but they also bring an added value by showing prosodic effects beyond what can be detected during acoustic analysis.

Appendix

1. List of sentences used as stimuli

La bamba es un baile latinoamericano muy conocido.

En toda América Latina se baila la bamba.

Mi padre pasó todo el fin de semana bailando la bamba.

La bajada de agua de nuestra casa se ha estancado otra vez.

A mi padre le duele la barriga.

La base científica del Covid es indudable.

La tienda de deporte más conocida de la isla se llama Base.

La banda de música empezó el concierto con la bamba.

La báscula demuestra que he bajado la barriga.

La bala le ha atravesado la barriga.

La baba de caracol se usa para producir cosméticos.

A la niña se le cae la baba.

El perro tiene hambre, se le cae la baba.

La vaca de Juan cuesta mucha pasta.

La vaca de Juan ha dado mucha leche este año.

A Juan se le escapó la vaca.

La papilla de avena le gustó mucho a la niña.

La paciencia de esa mujer me tenía bastante impresionado.

La paciencia de mi padre siempre ha sido objeto de admiración.

La patrulla ha estado buscando al ladrón desde hace días.

La paella valenciana es la más auténtica de todas las paellas.

La vacuna contra el Covid debería ser obligatoria para todos.

El jugador del Real Madrid comenzó la jugada por la banda.

Las bandas de música más conocidas de la historia son The Rolling Stones y Metallica.

La valla protege la casa de los animales salvajes.

El ladrón casi se escapa, pero al final se chocó contra la valla.

La barriga de ese hombre casi no cabe en la puerta.

La basura acumulada en el Océano Atlántico es una pasada.

La parte más difícil de ser padre es tener que aprenderlo.

Vimos las primeras dos películas y ahora tenemos ganas de ver la última parte.

La pandilla de mi barrio es muy conflictiva.

La pasta de dientes que compramos no sirve para niños.

La baraja española tiene cuarenta cartas con cuatro palos.

La página web del gobierno anunció la erupción del volcán de La Palma.

Para acceder a más información sobre las clases online tienes que ir a la página web.

Las papas arrugadas son una comida típica de las Islas Canarias.

Vivo en Las Palmas, pero soy de Gáldar.

Las Palmas de Gran Canaria es la capital de la isla.

Las Vacas Locas es una banda de música de Tenerife.

Las bases aéreas han parado de operar a causa de la pandemia.

La palanca de cambios de mi coche dejó de funcionar.

Con una palabra lo dijo todo.

Está mal dicho decir "la padre".

En esta celebración se suelta al final la paloma de la paz.

Me temo que la paz mundial es imposible de conseguir.

Se llama panza de burro cuando está nublado en verano.

La camisa es muy pequeña y se le ve la panza.

La batata del potaje no era muy dulce.

La barrera estaba mal colocada y el portero no veía.

Estoy a dieta y tengo miedo de subirme en la báscula.

La paga mensual es más baja de lo que pensaba Paco.

Tuve un accidente muy grave y por eso pedí una paga.

Después de cuatro horas en el quirófano los médicos lograron sacarle la bala.

El gato pardo siempre sabe cómo llenarse la panza.

2. Examples of TextGrid annotations
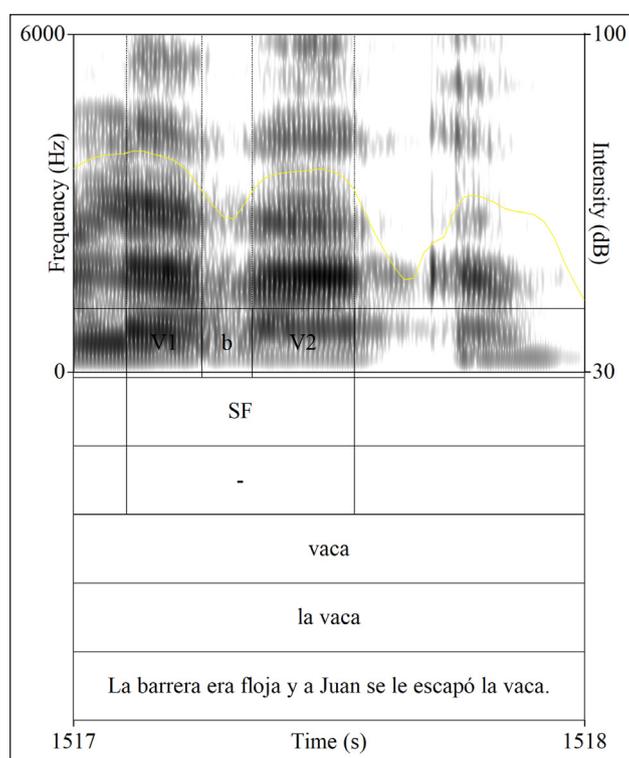
Figure 15: Annotation of a VCV sequence in the word *vaca* (cow) (condition: SF).
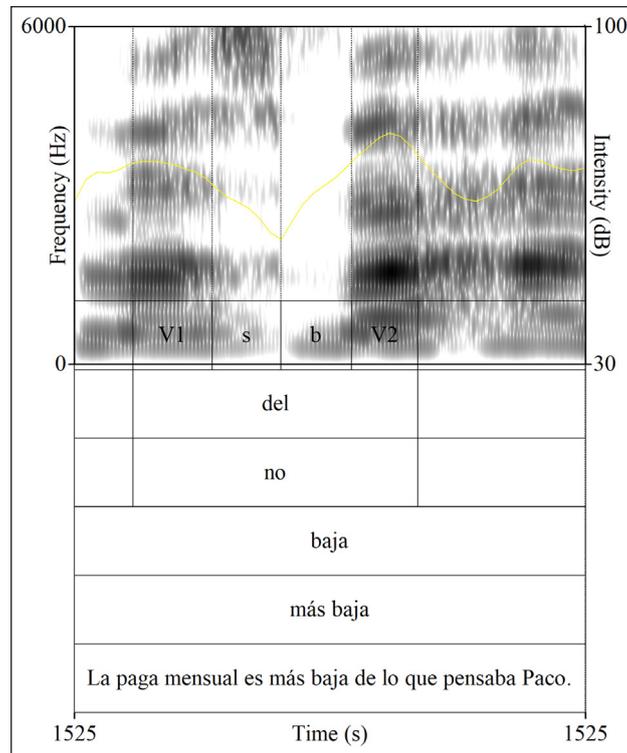
Figure 16: Annotation of a VsCV sequence in the word *baja* (low) with /s/ retention (condition: DEL).
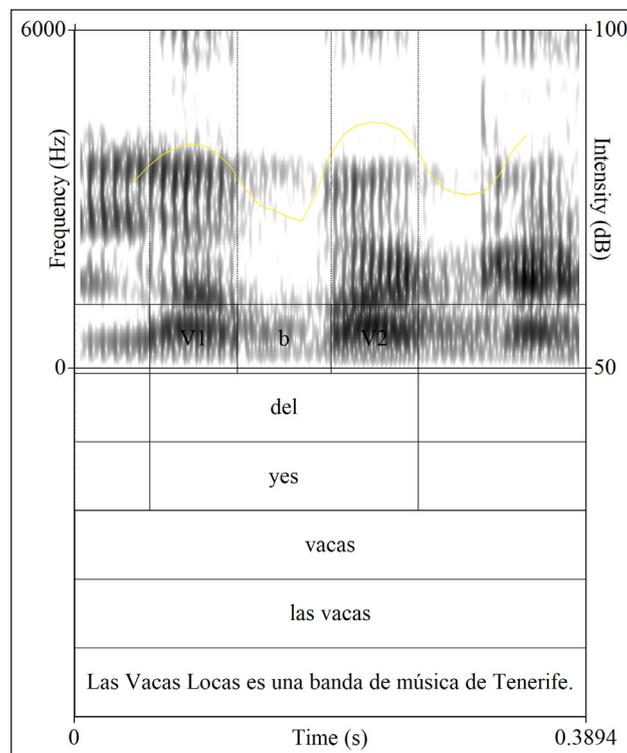
Figure 17: Annotation of a V(*s*)CV sequence in the word *vacas* (cows) with /s/ deletion (condition: DEL).

Acknowledgements

Competing Interests

The authors have no competing interests to declare.

References

Almeida, M. (1982). En torno a las oclusivas tensas grancanarias. *Revista de Filología de la Universidad de La Laguna, 1,* 77–88.

Almeida, M., & Díaz Alayón, C. (1988). *El español de Canarias.* Santa Cruz de Tenerife.

Alvar, M. (1972). *Niveles socio-culturales en el habla de las Palmas de Gran Canaria.* Las Palmas de Gran Canaria: Cabildo Insular de Gran Canaria.

Amastae, J. (1989). The intersection of s aspiration/deletion and spirantization in Honduran Spanish. *Language variation and change, 1,* 169–183. DOI: https://doi.org/10.1017/S0954394500000053

Baltrusaitis, T., Zadeh, A., Lim, Y. C., & Morency, L.-P. (2018). Openface 2. 0: Facial behavior analysis toolkit. *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, 59–66. DOI: https://doi.org/10.1109/FG.2018.00019

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2018). lme4: linear mixed-effects models using 'Eigen' and S4. R package (version 1.1-17). cran.r-project.org/web/packages/lme4.

Beckman, M. (1986). *Stress and non-stress accent.* Dordrecht, Holland/ Riverton, USA: Foris Publications. DOI: https://doi.org/10.1515/9783110874020

Beckman, M., Díaz-Campos, M., McGory, J. T., & Terrell, A. M. (2002). Intonation across Spanish, in the Tones and Break Indices framework. *Probus, 14,* 9–36. DOI: https://doi.org/10.1515/prbs.2002.008

Beckman, M., & Pierrehumbert, J. (1986). Intonational structure in English and Japanese. *Phonology Yearbook, 3*, 255–310. DOI: https://doi.org/10.1017/S095267570000066X

Beskow, J., Granström, B., & House, D. (2006). Visual correlates to prominence in several expressive modes. *INTERSPEECH 2006 ICSLP*. 1272 75. DOI: https://doi.org/10.21437/Interspeech.2006-375

Boersma, P., & Weenink, D. (2022). Praat: Doing phonetics by computer, version 6.1.52 [Computer program]. Available at: http://www.praat.org/.

Broś, K. (2016). Stratum junctures and counterfeeding: Against the current formulation of cyclicity in Stratal OT. In C. Hammerly & B. Prickett (Eds.), *Proceedings of the Forty-Sixth Annual Meeting of the North East Linguistic Society.* Volume 1, 157–170. GLSA Publications.

Broś, K., Żygis, M., Sikorski, A., & Wołłejko, J. (2021). Phonological contrasts and gradient effects in ongoing lenition in the Spanish of Gran Canaria. *Phonology, 38*(1), 1–40. DOI: https://doi.org/10.1017/S0952675721000038

Broś, K., & Lipowska, K. (2019). Gran Canarian Spanish non-continuant voicing: gradiency, sex differences and perception. *Phonetica, 76,* 100–125. DOI: https://doi.org/10.1159/000494928

Browman, C., & Goldstein, L. (1990). Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston & M. E. Beckman (Eds.), *Papers in laboratory phonology*, 341–376. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511627736.019

Canfield, D. L. (1962). *La pronunciación del español en América*. Bogotá: Instituto Caro y Cuervo.

Carrasco, P., Hualde, J. I, & Simonet, M. (2012). Dialectal Differences in Spanish Voiced Obstruent Allophony: Costa Rican versus Iberian Spanish. *Phonetica, 69,* 149–179. DOI: https://doi.org/10.1159/000345199

Cheng, Ch.-Ch., & Kisseberth, Ch. W. (1979). Ikorovere Makua tonology (part 1). *Studies in the Linguistic Sciences, 9*(1), 31–63.

Colantoni, L., & Marinescu, I. (2010). The scope of stop weakening in Argentine Spanish. In Ortega-Llebaria (Ed.), *Selected proceedings 4th Conference on Laboratory Approaches to Spanish Phonology*, 100–114. Sommerville, MA: Cascadilla Press.

Cole, J., Hualde, J. I., & Iskarous, K. (1999). Effects of prosodic and segmental context on /g/-lenition in Spanish. In Fujimura, Osamu, Joseph, Brian D. & Palek, Bohumil (Eds.), *Proceedings of LP '98: item order in language and speech. Vol. 2,* 575–589. Prague: Karolinum.

Dalcher, C. V. (2008). Consonant weakening in Florentine Italian: A cross-disciplinary approach to gradient and variable sound change. *Language Variation and Change, 20*(2), 275–316. DOI: https://doi.org/10.1017/S0954394508000021

Dorta Luis, J. (Ed.) (2013). *Estudio comparativo preliminar de la entonación de Canarias, Cuba y Venezuela.* Madrid-Santa Cruz de Tenerife: La Página ediciones S/L, Colección Universidad.

Dorta Luis, J., & Herrera Santana, J. L. (1993). Experimento sobre la discriminación auditiva de las oclusivas tensas grancanarias. *Estudios de fonética experimental, 5,* 163–188.

Eddington, D. (2011). What are the contextual phonetic variants of /b d g/ in colloquial Spanish? *Probus, 23*(1), 1–19. DOI: https://doi.org/10.1515/prbs.2011.001

Ennever, T., Meakins, F., & Round, E. R. (2017). A replicable acoustic measure of lenition and the nature of variability in Gurindji stops. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *8*(1), 20. DOI: https://doi.org/10.5334/labphon.18

Esteve-Gibert, N., & Prieto, P. (2013). Prosodic structure shapes the temporal realization of intonation and manual gesture movements. *Journal of Speech, Language, and Hearing Research*, *56*(3), 850–864. DOI: https://doi.org/10.1044/1092-4388(2012/12-0049)

Face, T. L. (2000). Prosodic manifestations of focus in Spanish. *Southwest Journal of Linguistics, 19*(1), 49–65.

Face, T. L. (2001). Focus and early peak alignment in Spanish intonation. *Probus, 13,* 223–246. DOI: https://doi.org/10.1515/prbs.2001.004

Face, T. L., & Prieto, P. (2007). Rising accents in Castilian Spanish: A revision of Sp_ToBI. In G. Elordieta & M. Vigário (Eds.), *Journal of Portuguese Linguistics, 6(1), Special Issue on Prosody of Iberian Languages*, 117–146. DOI: https://doi.org/10.5334/jpl.147

Fernandez, J. (1982). The allophones of /b, d, g/ in Costa Rican Spanish. *Orbis, 31*, 121–146.

Figueroa Candia, M. A., & Evans, B. G. (2015). Evaluation of segmentation approaches and constriction degree correlates for spirant approximant consonants. *Poster presented at the 18th International Congress of Phonetic Sciences*, Glasgow.

Goldrick, M. (1998). *Optimal opacity: Covert structure in phonology.* John Hopkins University, NJ.

Herrera Santana, J. (1997). Estudio acústico de /p, t, c, k/ y /b, d, y, g/ en Gran Canaria. In M. Almeida & J. Dorta (Eds.), *Contribuciones al estudio de la lingüística hispánica (Homenaje al profesor Ramón Trujillo),* 73–86. Barcelona: Montesinos.

Holbrook, B. B., Kawamoto, A. H., & Liu, Q. (2019). Task demands and segment priming effects in the naming task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(5), 807–821. DOI: https://doi.org/10.1037/xlm0000631

Hualde, J. I. (2005). *The sounds of Spanish*. Cambridge University Press.

Hualde, J. I. (2013). Intervocalic lenition and word-boundary effects: Evidence from Judeo-Spanish. *Diachronica, 30*(2)*,* 232–266. DOI: https://doi.org/10.1075/dia.30.2.04hua

Hualde, J. I., & Nadeu, M. (2012). Lenition and phonemic overlap in Rome Italian. *Phonetica*, *68*(4), 215–242. DOI: https://doi.org/10.1159/000334303

Hualde, J. I., & Prieto, P. (2015). Intonational variation in Spanish: European and American varieties. In S. Frota & P. Prieto (Eds.), *Intonation in Romance,* 350–391. Oxford: Oxford University Press, DOI: https://doi.org/10.1093/acprof:oso/9780199685332.003.0010

Hualde, J. I., Nadeu, M., & Simonet, M. (2010). Lenition and phonemic contrast in Majorcan Catalan. In S. Colina, A. Olarrea & A. M. Carvalho (Eds.), *Romance linguistics 2009. Selected papers from the 39th Linguistic Symposium on Romance Languages*, 63–79. Amsterdam, NL: Benjamins. DOI: https://doi.org/10.1075/cilt.315.04hua

Hualde, J. I., Simonet, M., & Nadeu, M. (2011). Consonant lenition and phonological recategorization. *Laboratory Phonology*, *2*(2). DOI: https://doi.org/10.1515/labphon.2011.011

Hualde, J. I., Shosted, R., & Scarpace, D. (2011). Acoustics and articulation of Spanish /d/ articulation. *Proceedings of ICPhS XVII*, pp. 906–909.

Hyman, L. M. (1977). On the nature of linguistic stress. In Larry M. Hyman (Ed.), *Studies in stress and accent,* 37–82.

Katz, J. (2016). Lenition, perception and neutralisation. *Phonology*, *33*(1), 43–85. DOI: https://doi.org/10.1017/S0952675716000038

Katz, J., & Pitzanti, G. (2019). The phonetics and phonology of lenition: A Campidanese Sardinian case study. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *10*(1), 16. DOI: https://doi.org/10.5334/labphon.184

Krause, P. A., Kay, C. A., & Kawamoto, A. H. (2020). Automatic motion tracking of lips using digital video and openface 2.0. *Laboratory Phonology: Journal of the Association for Laboratory Phonology, 11*(1), 9. DOI: https://doi.org/10.5334/labphon.232

Ladd, D. R. (1980). *The structure of intonational meaning.* Indiana University Press.

Lavoie, L. M. (2001). *Consonant strength: Phonological patterns and phonetic manifestations.* Routledge, Taylor & Francis Group.

Lenth, R.V. (2019). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package. https://cran.r-project.org/web/packages/emmeans/index.html.

Lipski, J. M. (1996). *El español de América.* Madrid: Cátedra.

Lozano, M. d. C. (1979). *Stop and spirant alternations: Fortition and spirantization processes in Spanish phonology.* Bloomington, IN: Indiana University Linguistics Club.

Lüdecke, D. (2018). Ggeffects: Tidy data frames of marginal effects from regression models. *Journal of Open Source Software, 3*(26), 772. DOI: https://doi.org/10.21105/joss.00772

Machuca Ayuso, M. J. (1997). *Las obstruyentes no continuas del español: relación entre las categorías fonéticas y fonológicas en habla espontánea.* PhD dissertation, Universitat Autònoma de Barcelona.

Martínez-Celdrán, E. (2008). Some chimeras of traditional Spanish phonetics. In: L. Colantoni & J. Steele (Eds.), *3rd conference on laboratory approaches to Spanish phonology,* 32–46. Somerville, MA: Cascadilla Proceedings Project.

Martínez-Celdrán, E., & Regueira, X. L. (2008). Spirant approximants in Galician. *Journal of the International Phonetic Association, 38*(01). DOI: https://doi.org/10.1017/S0025100308003265

McNeill, D. (2008). *Gesture and thought.* The University of Chicago Press.

Navarro Tomás, T. (1918). *Manual de pronunciación española,* 6th ed. New York: Hafner (1967).

Oftedal, M. (1985). *Lenition in Celtic and in Insular Spanish: the secondary voicing of stops in Gran Canaria.* Oslo: Universitetsforlaget.

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, Timothy L. (Ed.), *Laboratory approaches to Spanish phonology*, 237–253. Berlin & New York: Mouton de Gruyter.

Ortega-Llebaria, M. (2006). Phonetic cues to stress and accent in Spanish. In Manuel Díaz-Campos (Ed.), *Selected Proceedings of the 2nd Conference on Laboratory Approaches to Spanish Phonetics and Phonology*, 104–118. Somerville, MA: Cascadilla Proceedings Project.

Ortega-Llebaria M., & Prieto P. (2007). Disentangling stress from accent in Spanish: Production patterns of the stress contrast in de-accented syllables, In P. Prieto, J. Mascaró, M.-J. Solé (Eds.), *Segmental and prosodic issues in Romance Phonology,* CILT, 155–176. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/cilt.282.11ort

Parrell, B. (2010). Articulation from acoustics: Estimating constriction degree from the acoustic signal. *The Journal of the Acoustical Society of America, 128*(4), 2289–2289. DOI: https://doi.org/10.1121/1.3508033

Parrell, B. (2011). Dynamical account of how /b, d, g/ differ from /p, t, k/ in Spanish: Evidence from labials. *Laboratory Phonology, 2*(2). DOI: https://doi.org/10.1515/labphon.2011.016

Prince, A., & Smolensky, P. (2004). *Optimality theory: Constraint interaction in generative grammar.* Blackwell Pub. DOI: https://doi.org/10.1002/9780470759400

Purse, R. (2019). The Articulatory Reality of Coronal Stop 'Deletion'. In *Proceedings of the 19th International Congress of the Phonetic Sciences (XIX).* Melbourne, Australia.

Quesada Pacheco, M. A. (1996). El español de América Central. In Alvar, M. (ed.), *Manual de dialectología hispánica.* 100–115. Barcelona: Ariel.

Quesada Pacheco, M. A. (2010). *El español hablado en América Central: Nivel fonético.* Madrid/ Frankfurt: Iberoamericana/Vervuert. DOI: https://doi.org/10.31819/9783865278708

R Core Team. (2020). *R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing.* https://www.R-project.org/ (accessed 1 May 2023).

Recasens, D. (2016). The effect of contextual consonants on voiced stop lenition: Evidence from Catalan. *Language and Speech, 59*(1), 139–161. DOI: https://doi.org/10.1177/0023830915581720

Romero, J. (1995). *Gestural organization in Spanish: an experimental study of spirantization and aspiration.* PhD dissertation, University of Connecticut.

Romero, J., Parrell, B., & Riera, M. (2007). What distinguishes /p/, /t/, /k/ from /b/, /d/, /g/ in Spanish? Poster presented at Phonetics and Phonology in Iberia. Braga, Portugal.

Sangari, S. (2002). Visual correlates to focal accent and their relation to the fundamental frequency contour. *The Journal of the Acoustical Society of America, 112*(5), 2441–2441. DOI: https://doi.org/10.1121/1.4780032

Shaw, J. A., & Kawahara, S. (2023). Limits on gestural reorganization following vowel deletion: The case of Tokyo Japanese. Laboratory Phonology: *Journal of the Association for Laboratory Phonology, 14*(1), 1–33. DOI: https://doi.org/10.16995/labphon.8543

Shaw, J. A., Carignan, C., Agostini, T. G., Mailhammer, R., Harvey, M., & Derrick, D. (2020). Phonological contrast and phonetic variation: The case of velars in Iwaidja. *Language, 96*(3), 578–617. DOI: https://doi.org/10.1353/lan.2020.0042

Smolensky, P., & Goldrick, M. (2016). *Gradient symbolic representations in grammar: The case of French liaison.* Ms. Johns Hopkins University and Northwestern University. Tiede, M. (2005). Mview. Software.

Trujillo, R. (1980). Sonorización de sordas en Canarias. *Anuario Letras, 18*, 247–254.

van Oostendorp, M. (2006). *Theory of morphosyntactic colours.* Mertens Institute, Amsterdam.

Vanrell, M. d. M., & Fernández Soriano, O. (2018). Language variation at the prosody-syntax interface: Focus in European Spanish. In M. García García & M. Uth (Eds.), *Focus realization in Romance and beyond,* 33–70. Amsterdam, John Benjamins. DOI: https://doi.org/10.1075/slcs.201.02van

Wagner, P., Malisz, Z., & Kopp, S. (2014). Gesture and speech in interaction: An overview. *Speech Communication, 57*, 209–232. DOI: https://doi.org/10.1016/j.specom.2013.09.008

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis* (2nd ed. 2016). Springer International Publishing: Imprint: Springer. DOI: https://doi.org/10.1007/978-3-319-24277-4

Zimmermann, E. (2019). Gradient symbolic representations and the typology of ghost segments. *Proceedings of AMP 2018.* DOI: https://doi.org/10.3765/amp.v7i0.4576

Zubizarreta, M. L. (1998). *Prosody, focus, and word order (Vol. 213).* Cambridge, MA: MIT Press.

Żygis, M., Fuchs, S., & Stoltmann, K. (2017). Orofacial expressions in German questions and statements in voiced and whispered speech. *Journal of Multimodal Communication Studies, 4*(1–2). Special issue: Gesture and Speech in interaction, 87–92.