languages

Special Issue Reprint

# Phonetics and Phonology of Ibero-Romance Languages

Edited by
Rebeka Campos-Astorkiza

mdpi.com/journal/languages

MDPI

# Phonetics and Phonology of Ibero-Romance Languages

# Phonetics and Phonology of Ibero-Romance Languages

Guest Editor

**Rebeka Campos-Astorkiza**

MDPI

Basel • Beijing • Wuhan • Barcelona • Belgrade • Novi Sad • Cluj • Manchester

*Guest Editor*
Rebeka Campos-Astorkiza
Spanish and Portuguese
The Ohio State University
Columbus, OH
USA

*Editorial Office*
MDPI AG
Grosspeteranlage 5
4052 Basel, Switzerland

This is a reprint of the Special Issue, published open access by the journal *Languages* (ISSN 2226-471X), freely accessible at: https://www.mdpi.com/journal/languages/special_issues/95IH3BKF96.

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

Lastname, Firstname, Firstname Lastname, and Firstname Lastname. Article Title. *Journal Name* **Year**, *Volume Number*, Page Range.

**ISBN 978-3-7258-6312-9 (Hbk)**
**ISBN 978-3-7258-6313-6 (PDF)**
**https://doi.org/10.3390/books978-3-7258-6313-6**

Cover image courtesy of Rebeka Campos-Astorkiza

# Contents

# About the Editor

**Rebeka Campos-Astorkiza**

Rebeka Campos-Astorkiza is an Associate Professor within the Department of Spanish and Portuguese at the Ohio State University. She works on phonetics and phonology from a myriad of perspectives, including laboratory phonology, sociophonetics, and second language acquisition. In Lab Phonology, her research combines theory with experimental methodologies and illustrates how phonetic data, from both production and perception, can help us to develop models to explain sound patterns. In terms of sociophonetics, she is interested in phonetic variation in two areas: dialectal phonetic variation and the role of bi/multilingualism, especially in Basque-Spanish contact, and accent stylization in popular music. Finally, her work on second language acquisition of phonology focuses on new methodologies and instructional implications. She leads Our Voices/Nuestras Voces, a project that combines pedagogy and research in linguistics focusing on the acquisition of Spanish pronunciation by L1 speakers of English.

*languages* MDPI

*Editorial*

# Phonetics and Phonology of Ibero-Romance Languages: An Introduction to the Special Issue

**Rebeka Campos-Astorkiza**

Department of Spanish and Portuguese, College of Arts and Sciences, The Ohio State University, Columbus, OH 43210, USA; campos-astorkiza.1@osu.edu

This Special Issue includes twelve articles that provide an insight into the phonetics and phonology of Ibero-Romance languages. It showcases the breadth of approaches to these languages and how crucial they have been in several theoretical and methodological advances related to these fields. Thus, the articles included in this issue not only speak to researchers working on Ibero-Romance languages but also to a wider audience by highlighting how research on these languages contributes to general phonological and phonetic debates. Focusing on synchronic phenomena, this Special Issue comprises studies that focus on different aspects of the Ibero-Romance sound systems, including the sub-segmental, segmental, and prosodic levels, and cover a wide range of topics from advancements in phonological theory based on new data and new theoretical approaches to long-standing phonological issues to laboratory approaches to phonology, issues concerning the phonetics–phonology interface, experimental phonetics including acoustic, articulatory, and perception experiments, and sociophonetics. While the range of contributions within this Special Issue is broad, several threads emerge in terms of their significance for Ibero-Romance phonetics and phonology. Romance scholars have long been at the forefront of laboratory phonology (Gibson & Gil, 2019), and several studies in this Issue present new data that has theoretical importance for phonological processes such as resyllabification and rhotic allophony in Ibero-Romance varieties. Other articles frame new formal analyses to phenomena within Optimality Theory, such as vowel harmony and morphophonological alternations, explored in earlier studies that crucially, as the authors highlight, lacked empirical evidence to provide comprehensive accounts. Overall, the articles in this Special Issue include a wealth of data on Ibero-Romance sounds that authors discuss in relation to new understandings of language change, sociophonetics, and bilingual perception, in addition to the topics already mentioned. In what follows, I provide a summary for each article following a framework that highlights how the contributions tie together.

Couched within Optimality Theory (OT), **Davis and Pollock** present a new analysis of Eastern Andalusian Spanish (EAS) plural laxing, which, according to traditional descriptions, occurs when the word-final plural morpheme /s/ is deleted and the preceding vowel laxes, affecting also any preceding mid vowel. The authors' approach is informed by acoustic data from previous studies and consultations with native speakers focusing on the Granada variety of EAS. Based off those empirical observations, their analysis departs from previous accounts and argues that the plural morpheme is not represented by /s/ in EAS but rather by a floating [-ATR] feature that is anchored to the right edge of a word to satisfy alignment constraints. The authors further extend their alignment analysis to nouns whose singular ends with a stressed vowel, a case that has not been systematically analyzed in previous OT approaches. To conclude, they show how their proposed analysis can account for other patterns of EAS plural laxing reported in the literature. Also working within OT, **Lozano and Bradley** present a comparative analysis of diminutive formation across three Hispano-Romance varieties: Judeo-Spanish, Colombian Spanish, and Castilian Spanish. The first two present a dissimilatory alternation between diminutive suffix allomorphs, while the last one does not. Diminutive suffixation has attracted attention within contemporary Hispano-Romance linguistics; the authors contribute by considering iterative diminutives, i.e., forms where there is an iteration of the diminutive suffix that gives rise to intensified diminutives. Lozano and Bradley, first, present a description of the alternation in Judeo-Spanish and Colombian Spanish diminutives and identify new patterns for iterated diminutives in Colombian Spanish based on corpus data. Next, the authors develop a comprehensive phonological analysis of simple and iterated diminutivization where the alternations surface from the interaction among constraints on prosodic unmarkedness, output–output correspondence, allomorph preference, and similarity avoidance and, more precisely, dissimilation is formalized as the local self-conjunction of markedness constraints. In this OT analysis, the diminutive alternations are phonologically conditioned by consonantal place dissimilation.

**Beristein** presents a study that aims to provide evidence from aerodynamic data for a phonological process. More precisely, the author examines resyllabification of consonants as onsets across words in Spanish and utilizes degree of vowel nasalization before and after a resyllabified nasal consonant as a diagnostic for syllabic structure. Previous laboratory studies have questioned complete resyllabification and provided evidence for incomplete resyllabification; Beristein adds to this debate by expanding the analysis to surrounding vowels and applying articulatory methods. Namely, the author examines the nasal airflow of Northern Peninsular speakers as they produce words with different syllabic structures, including resyllabification contexts. The study reports that anticipatory and carryover nasalization among heterosyllabic sequences is not affected by resyllabilifcation, i.e., derived onsets lead to similar nasalization as “true” onsets, leading Beristein to conclude that Spanish phonology displays complete resyllabification. Also with a goal of informing phonological accounts, **Ramsammy and Raposo de Medeiros** report acoustic and articulatory data from ultrasound tongue imaging on rhotic variation in Brazilian Portuguese (BP). The authors analyze data from speakers from São Paulo to present a thorough description of rhotic production using categorical and continuous measures, against the backdrop of a theoretical debate on the phonological analysis of rhotic allophony in Portuguese. Results from their quantitative analysis show categorical alternations among two or three types of rhotics, depending on the speaker, and acoustic and articulatory variability within those variants conditioned by prosodic context. More precisely, the authors report the strengthening of rhotic allophones at major prosodic boundaries. In conclusion, this study presents fine-grained phonetic variability that challenges previous phonological accounts of rhotics in BP, while the authors call for further research on the sociophonetics of rhotics in the language.

**Recasens** also presents ultrasound data but, in this case, to examine (co)articulatory patterns among consonant sequences in Catalan. The author examines articulatory and coarticulatory differences in such sequences based on manner of articulation, approximant vs. stop, and syllabic affiliation, complex onsets vs. heterosyllabic sequences. Focusing on velar approximants and stops preceded or followed by liquids, Recasens finds manner-related articulatory patterns such that approximants, besides being less constricted than stops as expected, also exhibit a retracted tongue body. In addition, results indicate that constriction degree is not affected by the preceding sound; however, the manner-dependent differences in tongue configuration mentioned earlier extend into a preceding liquid. This leads the author to propose that the approximant is produced with an active articulatory gesture, pre-planned during the preceding sound. Syllabic affiliation effects were also found: heterosyllabic consonants are produced with a more extreme lingual configuration and less gestural overlap. This last finding replicates observations for non-Romance languages; Recasens expands our understanding of the (co)articulatory features of approximants by bringing a Romance variety, Catalan, to the foreground.

In their article, **Barbosa and Alvarenga** examine prosody in a novel way by analyzing the connection between syllabic duration and F0 contours in BP. Through acoustic analysis of a corpus of story retellings by speakers from São Paulo and Rio de Janeiro, the authors quantify the convergence between syllabic duration maxima, derived from the normalization of V-to-V intervals, and four F0 descriptors, F0 median and range and F0 rise and fall mean rates, in three distinct prosodic functions: non-terminal and terminal boundaries and prominence. The results reveal an alignment between duration maxima and F0 events in unique ways depending on the function, with non-terminality correlating with quicker F0 rises synchronized with duration maxima, and prominence presenting a wider F0 range and an earlier association between the duration maxima and an F0 fall. In addition, preliminary dialectal and gender-based differences emerge that, as the article highlights, merit further investigation. Prosody is also explored by **Delgado**, but from the perspective of intonational patterns and language contact. This article presents a first exploration into the intonation of Basque-French speakers in Labourd by analyzing yes–no questions in French. Following the Autosegmental Metrical Model, Delgado reports that rising contours were the most common intonation pattern, as observed for other varieties of French; however, there are significant differences based on language dominance with falling contours occurring more frequently within the French-dominant group. This article is relevant for Ibero-Romance phonetics and phonology because it engages and provides a comparison with studies that explore the Basque-Spanish contact situation. Furthermore, the study is directly informed by findings and methodologies from previous research on the linguistic situation of Spanish in the Basque Country. In including Delgado's work in this Special Issue, the boundaries of Ibero-Romance are being expanded to give a broader perspective of the linguistic reality of Basque contact across national borders.

**González, Cox and Isgar** investigate the impact of prosodic structure on non-modal voicing in word-final vowels in Spanish through a reexamination of the dataset in González et al. (2022). More precisely, the authors present a fine-grained analysis of word-final vowel phonation that includes an acoustics-based categorization of vowels according to creaky voice, breathy voice or devoicing, as well as a measure of spectral tilt, i.e., the difference in relative amplitude between the first and second harmonic (H1–H2). An innovative approach in their analysis is that the authors treat phonation as including dynamic combinations and identify vowels with single, double, and triple phonation. In fact, results show that vowels with double phonation are the most common, especially vowels beginning with modal voice and ending with breathy voice. Furthermore, the article reports that prosodic structure impacts phonation: modal voice is the most frequent type of phonation at the end of intermediate phrases, while creaky and breathy voice are the most common at the end of intonational phrases. Finally, the acoustic measure of H1–H2 differentiates between types of phonation, except for female speakers' modal and creaky voice productions, validating the use of this measure to explore Spanish phonation. The study, which includes speakers from a range of Spanish varieties, paves the way for future, detailed investigations of dialectal differences in vowel phonation. **Mendes** also presents an acoustic investigation of laryngeal activity, but in this case the author explores the voicing profile of the plural morpheme <s> in BP. In traditional descriptions, this morpheme is described as a voiced fricative, especially in pre-vocalic contexts. However, Mendes presents evidence that the plural-marking fricative is undergoing a process of devoicing that departs from earlier accounts. The study is a quantitative investigation of the degree of voicing present during the fricative and what factors might be conditioning it, including surrounding phonological context, task type, and word frequency. Results confirm devoicing of the fricative before a pause and before a vowel; however, the voicing degree is higher in pre-vocalic than pre-pausal environments. In addition, a preceding voiced stop leads to more fricative voicing than a preceding voiceless one, but this effect is only present for pre-vocalic fricatives. Individual variation is conditioned by the post-fricative context with a higher degree of individual variation in pre-vowel than pre-pause contexts, leading Mendes to conclude that the voiceless fricative before a pause is relative stable compared to fricative voicing before a vowel. Within exemplar theory, the author hypothesizes that exemplars associated with traditionally prevocalic voiced fricatives are competing with an emerging sound pattern characterized by the fricative devoicing.

**Pollock** presents a sociophonetic study that analyzes an emergent alveolar variant of the Spanish post-alveolar affricate in the speech of politicians from central and southern Spain. The author conducts a corpus analysis of political speech where acoustic measures, i.e., the center of gravity of the frication period and the percent fricative duration in each affricate, as well as the auditory coding of the place of articulation, are used to determine the distribution of the two variants based on several linguistic and sociolinguistic factors. Pollock argues that the phonetic variation present in the data indeed suggests the production of two types of affricates. Results also indicate that alveolar affricates are favored before and after front vowels, suggesting a possible coarticulatory origin for these fronted affricates. In addition, alveolar realizations are associated with female speech, speakers from Madrid, scripted speeches and interviews with women, contexts, according to the author, traditionally associated with prestige and attention paid to speech. Pollock concludes that the alveolar affricate seems to be an incipient Labovian marker in the early stages of social stratification. The article ends with a call for more perceptual work to better determine the status of this new affricate among listeners from central and southern Spanish. **Boomershine and Johnson** present the only perception study in the Special Issue. Building off an earlier study (Boomershine et al., 2008), the authors examine the interaction between language experience and phonological perception by analyzing how sounds with allophonic versus phonemic status in Spanish and English are perceived by listeners from a range of English–Spanish bilingual experiences, including heritage speakers of Spanish, late bilinguals, and monolingual speakers. Results from a similarity rating task replicate previous findings that contrastive sounds are perceived as more dissimilar. The novel finding in this study is that Heritage speakers' perception is very similar to that of monolingual and late bilingual L1 Spanish speakers but distinct from English speakers, indicating that early exposure to a language has a long-lasting impact on perception. The authors couch their findings withing the Exemplar Resonance Theory (Johnson, 2006) which predicts that perceived similarity among allophones is language-specific and influenced by the listener's linguistic experience. However, Boomershine and Johnson's data, which include non-words, challenge the model, which is built on lexical items only. The authors conclude that more research is needed into the perceptual system of Heritage speakers in other languages to elucidate other relevant factors beyond allophonic status.

Finally, **Bäumler**'s article on loanword phonology presents a quantitative corpus study of the adaptation of anglicisms into Spanish. The author explores several factors that may condition the degree of adaptation of English words to the Spanish phonology, namely whether imitation of the English word is present or grapheme–phoneme correspondence takes place, among speakers from Mexico City and Madrid. Analyzing data from a reading task, Bäumler finds that orthography plays a major role in loanword phonology, since 73% of the data presents adaptation to the grapheme–phoneme correspondences of

Spanish. However, that general trend is conditioned by type of sound, with consonants being more frequently imported than vowels; country, with Mexicans imitating more than Spaniards; and language exposure and affinity, with higher scores for this variable leading to more imitation. The author calls for future work that compares speech from reading and spontaneous tasks and other Spanish varieties.

While the studies included in this Special Issue highlight the relevance of Ibero-Romance phonetics and phonology for a range of theoretical and methodological matters, the compilation is somewhat limited in its representation of less studied languages, varieties, or communities of speakers. Studies on stigmatized varieties, contact dialects, and minoritized and racialized communities would allow us to expand our theoretical models and test the boundaries of our methodological approaches. While there are efforts focused within this scope, future research would benefit from incorporating patterns from less studied varieties vis a vis dialects and groups of speakers that have been extensively researched. This observation also underscores another area that is explored to a limited degree in this Special Issue, namely comparative studies across Ibero-Romance varieties. Lozano and Bradley present an example of how such a study could be shaped, with their project signaling the relevance and fruitful prospects of research that compares, generalizes, and/or differentiates across Ibero-Romance languages and dialects. This kind of approach is important for phonological theory and would strengthen the explanatory power of any model for phonological processes. This becomes even more relevant for research within Optimality Theory, a phonological approach where typological predictions constitute its core. Ibero-Romance lends itself to OT factorial typologies, like some previous studies have demonstrated, and work in this vein is pivotal for exploring the reach of any phonological OT analysis.

Some studies in the Special Issue call for more research that connects perception and production. Empirical studies tend to compartmentalize these two systems, even though they are deeply interconnected, and several theories posit such a connection as one of their tenets. Strides in this direction have been made in sociophonetics; however, further research is necessary to fully understand the impact of production patterns in the processing and perception of speech. For example, socioacoustic studies might identify socially stratified patterns of variation; exploring the perception of that variation would help clarify questions such as whether the feature is a social marker, identifier, or stereotype. In addition, some studies find discrepancies between the production and perception of sociolinguistic speech patterns, problematizing conclusions based only on production data. In addition, examining both production and perception is not only important for sociophonetics but also for theories that model sound systems based on users' experiences, such as exemplar-based and usage-based theories.

**Funding:** This research received no external funding.

**Acknowledgments:** Thank you to all the contributors, reviewers and the editorial team.

**Conflicts of Interest:** The author declares no conflicts of interest.

## References

Boomershine, A., Hall, K. C., Hume, E., & Johnson, K. (2008). The impact of allophony versus contrast on speech perception. In *Contrast in phonology* (pp. 145–172). De Gruyter Mouton.

Gibson, M., & Gil, J. (2019). Romance sounds: New insights for old issues. In M. Gibson, & J. Gil (Eds.), *Romance phonetics and phonology* (pp. 1–20). Oxford University Press.

González, C., Weissglass, C., & Bates, D. (2022). Creaky voice and prosodic boundaries in Spanish: An acoustic study. *Studies in Hispanic and Lusophone Linguistics*, *15*(1), 33–65. [CrossRef]
Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, *34*(4), 485–499. [CrossRef]

*languages*

MDPI

*Article*

# A Feature Alignment Approach to Plural Realization in Eastern Andalusian Spanish

**Stuart Davis [1,*] and Matthew Pollock [2,*]**

[1] Department of Linguistics, Indiana University, Bloomington, IN 47405, USA
[2] Department of English and Foreign Languages, Louisiana State University Shreveport, Shreveport, LA 71115, USA
* Correspondence: davis@iu.edu (S.D.); mpollock@lsus.edu (M.P.)

**Abstract:** Using an optimality theoretic analysis, this study offers a conception of the problem of plural realization in Eastern Andalusian Spanish (EAS) where plural suffix /s/ was deleted diachronically that differs from other accounts that assign the EAS plural an underlying suffixal /s/ synchronically. Using alignment constraints, we argue that plural /s/ does not appear in the underlying form synchronically in EAS, but that instead the plural morpheme is represented by a floating [–ATR]PL feature that aligns to the right edge of the word and spreads left. The [–ATR] feature, represented phonetically as a laxing or opening of vowels, applies to all mid vowels, low vowels in word final position, and combines with vowel epenthesis to explain Eastern Andalusian pluralization tendencies in words with final consonants. We discuss the behavior of high vowels, which can be transparent to harmony, and focus in particular on the plural of words that end in a final stressed vowel that have been rarely discussed in the EAS literature. We develop an optimality-theoretic analysis on the Granada variety and extend that analysis to other varieties with somewhat different patterns.

**Keywords:** Eastern Andalusia; Peninsular Spanish; optimality theoretic analysis; [–ATR]; pluralization; alignment constraints; vowel harmony; phonology

## 1. Introduction and Background

In the variety of Spanish spoken in the eastern part of Andalusia, the southernmost region of Spain, vowel laxing has been long documented in dialectological writings going back at least to the work of Navarro Tomás (1938, 1939) and much subsequent work including Alonso et al. (1950) and Zubizarreta (1979) among many others as detailed in Herrero de Haro (2017b) and Herrero de Haro and Hajek (2022). Focusing here on the plural word forms of nouns and adjectives, when the word-final /s/ that marks the plural is deleted, the immediately preceding vowel typically laxes, as do all other preceding mid vowels in the word; and as reported in Jiménez and Lloret (2020), depending on the specific variety of Eastern Andalusian Spanish (henceforth, EAS), other preceding vowels may lax as well, as in the Jaén dialect where all vowels are lax in plural word forms when the historical word-final *s of the plural deletes. For example, in the plural of the word *nenes* /ˈne.nes/ 'boys', a prototypical EAS production would include laxed production of the mid-vowels and an absence of a final fricative consonant (i.e., [ˈnɛ.nɛ]).

Over the past thirty years, the EAS plural has fostered a variety of analyses in Optimality Theory as well as much discussion on specific data issues, especially concerning whether or not low and high vowels can be undergoers (targets) of the harmony process. While some of this research focuses on the plural (e.g., Herrero de Haro 2020), most of these works consider vowel laxing to be largely phonological, since coda /s/ in EAS generally deletes whether it is a morphological marker (as in the noun plural or 2nd person singular *tú* verb forms; e.g., Standard Peninsular Spanish (SPS) *teléfonos* /te.ˈle.fo.nos/ 'telephones' produced as EAS [tɛ.ˈlɛ.fɔ.nɔ]) or a lexical /s/ that is underlying in a monomorphemic word (e.g., SPS *jueves* /ˈhwe.ßes/ 'Thursday' produced as EAS [ˈhwɛ.ßɛ]).

Along these lines, previous optimality-theoretic analyses, such as those conducted by Jiménez and Lloret (2007, 2020), Walker (2011), Soriano (2012), and Kaplan (2021a, 2021b), have assumed that the EAS plural synchronically contains an underlying suffixal /s/. Vowel laxing is then viewed as a process whereby coda /s/ deletes, leaving behind a [spread glottis] feature that coalesces with the immediately preceding tense [+ATR] vowel resulting in a lax [–ATR] vowel; the [–ATR] feature then spreads regressively as laxing harmony for reasons of feature perceptibility. Jiménez and Lloret rely on licensing and anchoring constraints to achieve spreading. Soriano achieves laxing spreading by an AlignSuffix/ProsodicWord constraint that has the effect of preserving the laxing that is the trace of the intended consonantal gesture of the underlying /s/ plural, which enhances the perceptibility of this plural marker. Kaplan (2021b) using harmonic serialism achieves the spreading of the [–ATR] feature from the deleted plural /s/ by the notion of persistent licensing which has the effect of extending the domain of the licensed [–ATR] feature. Regarding the precise domain of the regressive spreading of [–ATR] in plural word forms, Jiménez and Lloret (2020) note the variation that occurs in Granada EAS. Spreading of [–ATR] leftward minimally goes to the stress vowel (i.e., a foot domain) and maximally to the beginning of the lexical word (word domain), the specifics depending on the specific variety (and perhaps the speech rate).

While previous optimality-theoretic analyses of EAS plural laxing harmony have conceived of it through licensing type constraints that help to maximize the perceptibility of the [–ATR] laxing feature that comes about through the deletion of coda /s/, in this paper we offer and formalize a very different conception of EAS vowel harmony. First, we focus only on the plural under the view that harmony involving the feature [–ATR], which is the exponence of the plural, is different from purely phonological harmony triggered by the deletion of coda /s/.[1] Support for this view comes from Henriksen (2017, pp. 110–11), who shows that the phonetics is not the same between the two situations: laxing is actually stronger when a word-final lexical /s/ deletes (as in *jueves* 'Thursday') than in the morphological laxing of the plural (as in *nenes* 'boys'). Moreover, Herrero de Haro (2020, p. 19) posits a scenario of EAS plural formation where plural marking is becoming perceptible by "an unidentified suprasegmental element" that relates to vowel opening. This suggests to us that the "unidentified suprasegmental element" synchronically is a floating [–ATR] feature that is the exponence of the plural rather than a suffixal -s, as has been assumed in all the previous analyses that we are aware of. Henriksen (2017, p. 122) comments "because /s/ has been lost, EAS appears to have co-opted the independently motivated laxing pattern as a plural marker, most likely in tandem with the right-to-left harmony process". Given this, it is not surprising that there can be phonetic differences between laxing harmony triggered by the deletion of lexical /s/ in coda position and laxing as part of the exponence of the plural. We then view EAS laxing as an instantiation of what has been termed by Bermúdez-Otero (2016, and references cited therein) as 'rule scattering', whereby a single process diachronically evolves into two (or more) separate rules, especially if an environment can be morphologically interpreted. Consequently, we consider the EAS plural laxing as having a separate analysis that is conceptually different from the laxing triggered by lexical /s/ deletion in coda position. Furthermore, following the perspective of Construction Morphology (Booij 2010), we view the plural as being different than other morphological constructions that might witness laxing, such as the 2nd person singular verbal conjugation of the tú form of the present tense verb. Different morphological constructions that seemingly have homophonous exponence nonetheless can have different characteristics, as with English suffixal -er, which has a length restriction in the comparative (e.g., intelligent—*intelligenter) but not as an agentive (e.g., interrogate—interrogator).

Based on synchronic EAS, we maintain that suffixal /s/ no longer represents the underlying plural in this variety. Instead, the plural morpheme is represented by a floating [–ATR] feature that aligns to the right edge of the word and spreads left. This [–ATR] feature is represented phonetically as a laxing or opening (i.e., lowering) that occurs most regularly in the mid vowels /e/ and /o/. Because [–ATR] can only be realized on vowels, final vowel epenthesis is triggered in the plural of EAS words where the singular ends in a lexical consonant so that the plural can be realized (*carbón/carbones* [karbon]—[karbɔnɛ] 'carbon (sg.)/(pl.)') where the mid front vowel acts as a default vowel generally in Spanish phonology as it occurs elsewhere in epenthetic contexts. Additionally, our analysis accounts for word-final high vowels /u/ and /i/, which can be transparent to harmony (Henriksen 2017), as well as the low vowel /a/, which undergoes laxing in word-final position but may or may not elsewhere depending on the subvariety. Furthermore, we analyze the plurals of words that end in final stressed vowels that can differ from the plural pattern of more typical words that end in stressless vowels. As far as we are aware, these have not been systematically discussed in the EAS laxing literature.

The remainder of this paper is organized as follows. In Section 2, we provide an overview of the data we will be analyzing based on Granada EAS. Using the phonetic analysis of Henriksen (2017) to pinpoint vowel height differences, we provide words both studied by Henriksen as well as representative examples of longer words discussed in other sources and from native speaker consultation. We discuss data disagreements especially with respect to whether high vowels actually undergo harmony and the matter of the domain of spreading. In Section 3 we develop an OT alignment analysis of the EAS plural where the plural word form is marked with a floating [−ATR] autosegment aligned to the right edge of the word that spreads leftward. We consider words with vowels of various heights including borrowed words that end in a high vowel, as well as words where the final lexical segment is a consonant. While the data we analyze in Section 3 all involve nouns (and adjectives) that have typical stress patterns, in Section 4 we consider the plural nouns whose singular ends in a final stressed vowel. These do not seem to have been systematically considered in the optimality-theoretic literature on EAS and we extend our alignment account to these forms. In Section 5, we briefly show how our alignment analysis can be extended to other patterns of EAS plural laxing that have been reported in the literature. Section 6 concludes.

## 2. Presentation of EAS Plural Data

In a typical EAS plural word form (focusing on nouns and adjectives), the word final vowel is usually laxed; then, depending on various factors, other vowels in the word may also surface as lax. We see two issues in describing the plural word forms. First, do EAS plural words underlyingly have a suffixal /-s/ like in other varieties of Spanish? Second, what exactly is the vowel laxing pattern in EAS plural words? Concerning the first issue, almost all previous analyses during the past three decades have assumed that EAS plural word forms synchronically have an underlying final /-s/, even though it does not surface in plural nouns and adjectives (i.e., *bañeras* 'bathtubs' as /ba.ˈɲe.ɾas/). This is the case whether the analyses overtly show the plural morpheme in their derivation or optimality-theoretic tableaux (as in Jiménez and Lloret 2007, 2020; Soriano 2012) or just assume it without showing it (as in Kaplan 2021b). Here, though, we follow Herrero de Haro (2020) who makes clear that in EAS the plural -s never surfaces as [s], which makes it unlike Caribbean varieties that have coda /s/ deletion but where the /s/ of the plural can resurface depending on phonological context among other factors (e.g., *meses* /me.ses/ as [ˈme.ses], [ˈme.seh], or [ˈme.seː] 'months'). It also makes it unlike EAS words that end in a lexical /s/ such as *mes* /mes/ 'month' pronounced as [mɛ] in EAS but where the /s/ shows up in the plural form *meses* [ˈmɛ.sɛ] 'months'.

Further evidence that there is no need to posit an underlying plural /-s/ in EAS comes from Herrero de Haro's (2020) perceptual study that shows that the singular and plural forms of many lexemes can be distinguished from the left edge of the word rather than the right edge (e.g., *coche* [ˈko-] 'car' vs. *coches* [ˈkɔ-] 'cars'). Consequently, Herrero de Haro (2020, p. 19) suggests that plural marking is becoming identifiable by "an unidentified suprasegmental element", which we identify in our analysis as a floating [−ATR] feature that marks plural word forms. Thus, in our presentation and analysis of the EAS plural, we will not include a suffixal /-s/ as part of the underlying representation of plural words. Instead, we will present a feature alignment analysis of the floating morphological [–ATR]PL feature, which begins for the plural in Section 3.

As for the second issue—namely, which vowels are considered lax in plural word forms—there is some disagreement on this subject even for vowels within the same variety of EAS. This particularly pertains to the Granada variety, where there are small disagreements in the data presentation of plurals between Jiménez and Lloret (2020) and Henriksen (2017). The disagreements concern whether high vowels and nonfinal low vowels undergo laxing in plural word forms. This disagreement, which goes back many decades, is reviewed by Jiménez and Lloret (2020, pp. 101–2) and more thoroughly by Herrero de Haro and Hajek (2022, pp. 144–48). The disagreement is mentioned in Ranson (1992, p. 308) where locals transcribing her EAS data and the author differed with respect to the coding of certain words as marking plurality. While the disagreement to a certain extent reflects variation in different subvarieties (and perhaps idiolects), it also reflects an unstated disagreement as to what an acoustic threshold would be for a vowel to be considered lax. In their presentation of the Granada dialect, Jiménez and Lloret (2020, p. 106) seem to consider word-final high vowels and stressed low vowels to always be lax and non-final unstressed low vowels to be optionally laxed in plural word forms, while Henriksen's (2017) analysis of speech data shows them not to be lax. However, these researchers agree that mid vowels always undergo laxing in plural word forms.

To be clear, even researchers who do not consider high vowels or non-final low vowels to be lax in plural word forms may nonetheless note minor phonetic differences of such vowels. So, for example, as discussed by Herrero de Haro and Hajek (2022), several researchers such as Zubizarreta (1979) consider high vowels in laxing contexts to be phonetically, but not phonologically, lax; consequently, such vowels are not viewed as undergoing laxing. Zubizarreta, in particular, distinguishes high-level (phonological) laxing that affects all mid vowels in the harmony domain as well as word-final low vowels from low-level (phonetic) laxing that can affect non-final low vowels and all high vowels in a harmony domain. Those undergoing low-level laxing are not considered real targets of the harmony process and are either transparent or opaque to it. A main reason for the data disagreement is that there does not seem to be an established agreed upon acoustic threshold in the EAS literature that determines when a vowel should be considered lax. As a somewhat related comparison, studies on English diphthong raising whereby /ay/ is raised to [ʌy] before voiceless consonants has an established threshold (Labov et al. 2006) of a 60 Hz difference in F1 lowering for /ay/ to be considered perceptually raised before a voiceless consonant. Speakers showing a 30 Hz difference are not considered to have perceptible raising, despite the slight difference in production this difference suggests. No such acoustic threshold has been agreed upon in the literature on EAS laxing, as pointed out by Pollock (2023, p. 186) in an analysis of vowel harmony among Granada politicians. This author found an average difference of only 15 Hz for F1 between vowels in plural and non-plural wordforms, much lower than the threshold of 60 to 100 Hz identified by Herrero de Haro (2017a) for perceptibility in vowel height identification preceding /s/, /r/, and /θ/ deletion.

Given the above discussion regarding the interpretation of the laxing data in plural word forms from Granada EAS, we must make certain decisions in our data presentation regarding when high vowels and non-final low vowels are considered lax. We follow Henriksen (2017) and Kaplan (2021b) (and also Zubizarreta 1979) in considering high vowels resistant to laxing. However, one pattern that these researchers did not explicitly consider, perhaps because of its rarity, concerns the plural of words ending in a word-final stressed high vowel, although analyses drawing on the work of Jiménez and Lloret predict lax word-final high vowels regardless of stress. Based on Herrero de Haro (2020, p. 4) and our own native speaker consultation, we consider the plural of such words to indeed undergo final laxing. We will discuss and analyze plurals of words ending in final stressed vowels in Section 4 after developing our analysis in Section 3. With respect to low vowels, we follow Henriksen (2017) and Zubizarreta (1979) in considering low vowels as being lax only in word-final position in plural word forms but not in other positions. Note that this differs from the claims made by Jiménez and Lloret, who state that non-final unstressed vowels in plural word forms can fail to undergo laxing, while stressed low vowels always undergo laxing. It is interesting to note that the view of Henriksen and Zubizarreta that leftward laxing harmony only affects mid vowels finds support in the experimental work of Herrero de Haro (2020) where EAS listeners are much better at identifying a plural word form by hearing just the first vowel of a two syllable paroxytone (i.e., a trochaic word form) if that vowel is mid as opposed to a non-mid vowel. Given all the caveats discussed above, we will present the Granada EAS plural data below in a way that is largely consistent with Henriksen (2017), and our optimality-theoretic analysis in Section 3 will be largely based on that description.

In a comparison of two varieties of Iberian Spanish, Henriksen (2017) maintains that EAS, which has phonological coda /s/ deletion and laxing in plural word forms, has eight vowels phonetically that can be transcribed as [a æ e ɛ i o ɔ u], whereas speakers of the more standard North-Central Peninsular Spanish (NCPS), which tends to retain coda /s/, lack [æ ɛ ɔ]. In Henriksen's perceptual study, speakers of EAS were able to distinguish plural word forms with laxed or [−ATR] vowels from their singular counterpart with tense or [+ATR] vowels (e.g., *jefe* 'chief' [jefe] versus *jefes* 'chiefs' [jɛfɛ]), whereas those who spoke NCPS were slower to identify the difference between plural and singular, and responded at closer to chance. An issue that arises in the analysis of EAS is the phonemic status of the lax vowels [æ ɛ ɔ].[2] For the purposes of this paper, we view the lax mid vowels [ɛ ɔ] as having emerging phonemic status, given that they can appear in all pertinent word positions as well as their perceptual saliency to native speakers, which makes them different from the lax variants of the other vowels. That said, we assume here following the presentations of Henriksen (2017, p. 107), Jiménez and Lloret (2020, p. 106) and Herrero de Haro and Hajek (2022, p. 146) that the underlying vowels in plural forms are the tense vowels that are found in the corresponding singular and that the lax vowels that surface in plural word forms are derived.

In what follows, we present data on the plural that represent several key segmental contrasts. Recall that we do not consider the plural as having an underlying word-final suffixal /s/. As mentioned, our data presentation will largely follow the pattern of laxing in Henriksen (2017), which, in turn, is very similar to the pattern of "high-level laxing" in Zubizarreta (1979). First, as Henriksen (2017) and many other researchers on EAS have found, plural word forms where all the vowels are mid, as in (1), surface with [−ATR] on all vowels. (In the data presentation, the orthographic representation is in the leftmost column, singular word forms are in the second column, the corresponding plural is in the third column, and the English gloss of the singular is in the rightmost column. The transcriptions indicate the location of stress and the syllable boundary; allophonic variation not related to the plural will not always be transcribed.)

(1) Mid Vowel Data

| | Orthography | Singular | Plural | Gloss |
|---|---|---|---|---|
| a. | *nene* | [ˈne.ne] | [ˈnɛ.nɛ] | boy |
| | *pomo* | [ˈpo.mo] | [ˈpɔ.mɔ] | doorknob |
| b. | *coche* | [ˈko.tʃe] | [ˈkɔ.tʃɛ] | car |
| | *velo* | [ˈbe.lo] | [ˈbɛ.lɔ] | veil |
| c. | *celebre* | [θe.ˈle.βɾe] | [θɛ.lɛ.βɾɛ] | famous |
| | *monólogo* | [mo.ˈno.lo.ɣo] | [mɔ.ˈnɔ.lɔ.ɣɔ] | monologue |
| d. | *pomelo* | [ˈpo.me.lo] | [ˈpɔ.mɛ.lɔ] | grapefruit |
| | *teléfono* | [te.ˈle.fo.no] | [tɛ.ˈlɛ.fɔ.nɔ] | telephone |

In the above data, it can be seen that mid-vowels always acquire [−ATR] when pluralized in EAS, with harmony spreading across the entirety of the word. Word length and stress do not seem to affect the spread of the plural [−ATR] feature as shown, but Jiménez and Lloret note that spreading to the left of the stressed syllable may be optional and that harmony to the penultimate syllable is optional in words with antepenultimate stress, provided that the vowels are different.

Next, low vowels have a more complicated relationship with harmony. While they are transparent to it, allowing harmony to extend past them word-medially, low vowels, as in (2), only manifest [–ATR] in word-final position:

(2) Low Vowel Data

| | Orthography | Singular | Plural | Gloss |
|---|---|---|---|---|
| a. | *casa* | [ˈka.sa] | [ˈka.sæ] | house |
| | *mapa* | [ˈma.pa] | [ˈma.pæ] | veil |
| b. | *tela* | [ˈte.la] | [ˈtɛ.læ] | fabric |
| | *poema* | [po.ˈe.ma] | [pɔ.ˈɛ.mæ] | poem |
| c. | *bañera* | [ba.ˈɲe.ɾa] | [ba.ˈɲɛ.ɾæ] | bathtub |
| | *colorada* | [ko.lo.ˈra.ða] | [kɔ.lɔ.ˈɾa.ðæ] | red |

In these data, it can be seen that word-final /a/ acquires the plural morpheme [–ATR], which spreads across the word. But word-internal /a/ neither undergoes harmony nor blocks it, as seen in the plural of colorada. That non-final /a/ can be transparent to harmony is one of the optional patterns noted by Jiménez and Lloret (2020, p. 106). However, these researchers indicate a change in /a/ when in the stressed syllable of plural word forms as in the initial syllable of *casas* 'houses' whereas Henriksen (2017, p. 123) is clear that the low vowel in a word like *casas* is not the target of harmony, and in the experimental study of Herrero de Haro (2020), the quality of the low vowel in the first syllable of plural words like *casas* was not used as a cue that the word was plural, while the quality of mid vowels were used in this way.

The pattern of high vowels appears similar to /a/ if we focus on word-internal (i.e., non-final) position. Here, as seen in (3), non-final high vowels do not undergo harmonic spreading when in the first syllable of bisyllabic trochees as in (3a) and they are transparent to it in the middle of a harmonic domain as in (3b) and (3c).

(3) High Vowel Data (non-final position)

| | Orthography | Singular | Plural | Gloss |
|---|---|---|---|---|
| a. | *nube* | [ˈnu.βe] | [ˈnu.βɛ] | cloud |
| | *vida* | [ˈvi.ða] | [ˈvi.ðæ] | life |
| b. | *poliza* | [po.ˈli.θa] | [pɔ.ˈli.θæ] | policy |
| | *seguro* | [se.ˈɣu.ɾo] | [sɛ.ˈɣu.ɾɔ] | certain |
| c. | *película* | [pe.ˈli.ku.la] | [pɛ.ˈli.ku.læ] | movie |
| | *adjetivo* | [ad.xe.ˈti.βo] | [ad.xɛ.ˈti.βɔ] | adjective |

A complication arises in plural forms of words that end in a high vowel, which are not common in Spanish nouns and adjectives. In (4a), we see four examples of word forms ending in a stressless high vowel. There is disagreement on the data in (4a) for Granada EAS. Henriksen (2017) and Zubizarreta (1979) do not consider the final high vowel in these plural word forms to be phonologically lax. On the other hand, Jiménez and Lloret (2020) indicate that they are. Here we follow Henriksen. Regarding the non-final vowels in words like those in (4a), for Jiménez and Lloret they reflect the regular harmony pattern in that non-final high vowels fail to lax while mid vowels can undergo laxing. Henriksen (2017) does not discuss vowel laxing of non-final vowels in words ending in a high vowel like those in (4a). Following Jiménez and Lloret (2020, p. 106) we consider the pre-final mid vowels to undergo laxing while the high vowels do not. The last example in (4a) *tribu* 'tribe' is interesting in that neither of the high vowels laxes in the plural.

(4) High Vowel Data (final position)

| | Orthography | Singular | Plural | Gloss |
|---|---|---|---|---|
| a. | *poli* | [ˈpo.li] | [ˈpɔ.li] | cop |
| | *ímpetu* | [ˈim.pe.tu] | [ˈim.pɛ.tu] | violent |
| | *espíritu* | [es.ˈpi.ri.tu] | [ɛs.ˈpi.ri.tu] | spirit |
| | *tribu* | [ˈtri.bu] | [ˈtri.bu] | tribe |
| b. | *club* | [ˈklu] (/klu/) | [ˈklʊ] | club |
| | *clip* | [ˈkli] (/kli/) | [ˈklɪ] | clip |
| | *menú* | [me.ˈnu] | [me.ˈnʊ] | menu |
| | *hindú* | [in.ˈdu] | [in.ˈdʊ] | Hindu |

The data in (4a) should be contrasted with (4b) where the word-final high vowel is stressed. (Note that the first two words in (4b) are borrowed from English without the final obstruent being lexicalized. See Bermúdez-Otero 2007 for discussion on such forms.) The data in (4) have been little discussed in the literature on EAS plurals. The occurrence of a word-final high vowel with stress on a noun or adjective as in (4b) is unusual in Spanish and is typically found in borrowed words. The EAS plural form of these words follows an observation by Herrero de Haro (2020, p. 4) that "words that have stress on the last syllable only lower one vowel when [plural] /-s/ is deleted in EAS". One example would be *bebés* 'babies', where the word-final /e/ of the EAS plural laxes but the unstressed first vowel does not experience optional harmonization. This would contrast with a word having stress on the initial syllable, where both vowels lax in the plural. Words like *bebés* and those in (4b) will be discussed in more detail in Section 4 after we develop our optimality-theoretic analysis of the EAS plural data based on the patterns shown in (1)–(4a) above.

Having now seen a variety of nouns with all types of vowels, two questions remain: how are consonant-final nouns treated when pluralized, and what happens along morpheme boundaries? The data in (5) provide examples of consonant final words.

(5) Consonant-Final Noun Data

| | Orthography | Singular | Plural | Gloss |
|---|---|---|---|---|
| a. | *mes* | [ˈmɛ] (/mes/) | [ˈmɛ.sɛ] | month |
| | *gol* | [ˈgol] | [ˈgɔ.lɛ] | goal |
| b. | *empujón* | [em.pu.ˈxon] | [ɛm.pu.ˈxɔ.nɛ] | push |
| | *carbón* | [kar.ˈbon] | [kar.ˈbɔ.nɛ] | carbon |
| | *seguidor* | [se.ɣi.ˈðor] | [sɛ.ɣi.ˈðɔ.ɾɛ] | follower |
| c. | *contiguidad* | [kon.ti.ɣui.ˈðad] | [kɔn.ti.ɣui.ˈða.ðɛ] | contiguity |
| | *observacional* | [ob.seɾ.βa.θio.ˈnal] | [ɔb.sɛɾ.βa.ciɔ.ˈna.lɛ] | observational |
| d. | *departamental* | [de.par.ta.men.ˈtal] | [dɛ.paɾ.ta.mɛn.ˈta.lɛ] | departmental |

Although diachronically these data could be described as taking on /-es/ as a plural allomorph, in our synchronic analysis, we do not posit any segmental suffix; rather, as mentioned at the beginning of this section, we posit the plural to be a floating [–ATR]PL feature. Because of the need for the plural to be expressed at the right edge of the word, word-final vowel epenthesis is triggered to take on the [–ATR] specification, as it cannot be taken on by consonants in Spanish. That it is the front mid vowel that epenthesizes is consistent with Spanish phonology, as /e/ surfaces as the epenthetic vowel in other contexts (e.g., to break up impermissible initial consonant clusters in borrowed words). The [–ATR] feature that is realized on the epenthetic vowel then spreads leftward as in the other EAS plural word forms.

A final matter concerning the data is the effect of certain morpheme boundaries in blocking or limiting the spread of the [–ATR] plural feature. First, as discussed by Soriano (2012, p. 301), harmony is restricted to the rightmost prosodic word adjacent to the plural morpheme. There must be some type of prosodic word domain where the domain of a secondary stress functions as its own prosodic word. Therefore, in compound words (and others with word-medial prosodic word boundaries which we do not elaborate on), we would not expect to see harmony spreading to the left of that edge, but rather to remain restricted to the rightmost prosodic word, as in the examples in (6):

(6) Word-medial prosodic boundaries

| | Orthography | Singular | Plural | Gloss |
|---|---|---|---|---|
| a. | *fotomultiplicador* | [fo.to+mul.ti.pli.ka.ˈðoɾ] | [fo.to+mul.ti.pli.ka.ˈdɔ.ɾɛ] | photomultiplier |
| | *autodecremento* | [au.to+de.kɾe.ˈmen.to] | [au.to+dɛ.krɛ.ˈmɛn.tɔ] | self-decrease |

Second, as noted by Jiménez and Lloret (2020, p. 108), the word-final vowel of the plural does not diphthongize over the word boundary unlike a word-final vowel of the singular. They contrast *tomate y* [e̯i] *clavel* 'tomato and carnation', where diphthongization takes place, with *claveles y* [ɛ.i] *tomates* 'carnations and tomatoes', where diphthongization is prevented with a clear boundary between the [ɛ] of the plural and the [i] of the conjunction. This is consistent with our view of the EAS plural having a floating [–ATR] autosegment which is aligned to the right edge of the grammatical word, thus preventing diphthongization over a word boundary.[3]

Having presented the EAS plural data, in the next section we offer an optimality-theoretic analysis that conceptualizes the plural morpheme as a floating morphological autosegment that is realized through various alignment constraints along the lines of morphemic vowel harmony as described by Akinlabi (1996) and Finley (2009).

## 3. Optimality Theoretic Analysis: Feature Alignment without Suffixal /-s/

In this section we will sketch an optimality-theoretic analysis of the Granada EAS plural data as presented in Section 2. Unlike previous analyses, we will not assume that the plural is marked by suffixal /-s/; rather we maintain that the underlying forms of plural words have a floating morphemic feature [–ATR] that marks the plural and that this feature is realized by alignment constraints (Akinlabi 1996; Finley 2009). Previous analyses such as Jiménez and Lloret (2007, 2020) have assumed that EAS plural words are marked with a word-final /-s/ because EAS has coda /-s/ deletion generally, making word-final plural /-s/ deletion another instance of general coda deletion. Along these lines, Soriano (2012, pp. 302–3) assumes a word-final plural /-s/ for Jaén EAS, though it is never realized, because "the realization of the consonant is intended though it is not finally produced, and the special restrictions that this pseudo-articulation provokes in the oral cavity makes these differences in height [i.e., laxing] appear in vowels". Vowel laxing is viewed as a process whereby coda /s/ deletes, leaving behind a [spread glottis] feature that coalesces with the immediately preceding tense [+ATR] vowel resulting in a lax [–ATR] vowel; the [–ATR] feature then spreads regressively as laxing harmony for reasons of feature perceptibility.

However, given that the plural /-s/ is never pronounced in EAS, making it unlike words in this variety that end in a lexical /s/ (such as *mes* /mes/ 'month' pronounced as [ˈmɛ] in EAS but where the /s/ shows up in the plural form *meses* [ˈmɛ.sɛ] 'months'), we posit that from a synchronic perspective the plural is simply marked by a morphemic [–ATR] floating feature. Our perspective regarding the representation of the plural is also supported by Henriksen (2017, pp. 110–1) who shows that the phonetics of laxing in Granada EAS is different when a word-final lexical /s/ deletes (as in *jueves* 'Thursday') as opposed to the morphological laxing of the plural (as in *nenes* 'boys'). Specifically, Henriksen (2017, p. 111) notes "that the trend toward laxing is greater in /s/-final monomorphemic words than in /s/-final plurals". To us, this suggests that laxing triggered by lexical /s/-deletion is conceptually different than plural laxing: the former may indeed be conceptualized through positional licensing and feature perceptibility as in the analyses of Jiménez and Lloret, Walker, and Kaplan, whereas the latter can be conceptualized as morpheme realization via alignment constraints.

Relatedly, and what has not been previously discussed in the Granada EAS laxing literature, is a difference in laxing in oxytones (i.e., final stress) between plural forms and singular forms with lexical /s/ deletion with respect to the leftward spreading of [–ATR]. This can be seen in the contrast between the singular *revés* 'other side' pronounced as [rɛ.ˈvɛ] and the plural of *bebé* 'baby' pronounced as [be.ˈbɛ] where [–ATR] is only realized on the final vowel in plural of oxytones ending in a stressed vowel. This will be discussed in detail in Section 4. Further, as mentioned earlier, our view that the EAS plural does not have a suffixal /s/ also finds support from Herrero de Haro's (2020, p. 19) perceptual study that the plural forms of at least some lexemes can be distinguished from the left edge of the word rather than the right edge, leading him to suggest that plural marking is becoming identifiable by "an unidentified suprasegmental element"; the element we identify as a floating [−ATR] feature.

In our analysis, the plural morpheme represented by [−ATR]PL is aligned to both the right and left edge of the word. The Align-Right [−ATR]PL is categorical and must be manifested on the right edge of the grammatical or lexical word, lest it not be manifested at all; leftward spreading of [−ATR]PL is due to a gradient Align-Left feature constraint that aligns it to the left edge of the prosodic word. Alignment constraints can either be gradient or categorical. While an alignment constraint that aligns a smaller entity with a larger one is evaluated gradiently, an alignment constraint that aligns a larger entity with a smaller one is evaluated categorically. As an example of the latter, in an analysis of the distribution of the feature [spread glottis] in Korean, Davis and Cho (2003) use the constraint Align (Word, Left, [spread glottis], Left) that requires the beginning of the word (i.e., the first phoneme of the word) to have the feature [spread glottis]. If the initial sound of the word does not have the feature [spread glottis], the constraint is violated. There is no gradient evaluation. Both types of alignment constraints will be used in our analysis of EAS plural laxing below and it will be seen that these constraints interact with others, including feature cooccurrence constraints. In what follows, we return to the presented data from Section 2, provide OT constraints that help describe the process of plural vowel laxing, and analyze possible candidates in tableaux, yielding an OT ranking that describes EAS harmonic tendencies. Later, in Section 5, we suggest extensions of our analysis to cover other somewhat different patterns of EAS plural laxing that are reported in the literature.

### *3.1. Treatment of Plural Word Forms with Mid Vowels*

When working with words that contain exclusively mid vowels, there is a clear pattern. For example, in a word such as *nene* 'boy' from (1b), the plural always has [−ATR] applied to both vowels. We will use a categorical alignment constraint that establishes that the right edge of the word will have the plural morpheme (i.e., [−ATR]PL), as described in (7):

(7) AlignPL-R: Align (Word, Right, [−ATR]PL, Right)
The right edge of the grammatical word is aligned with the right edge of the plural morpheme [−ATR]PL.

The alignment constraint in (7) is categorical in that the [−ATR]PL feature has to surface at the right edge of the grammatical word or category, lest it not be manifested at all. The use of grammatical word or category as part of morpheme alignment constraints can be found in use by Kager (1999), Akinlabi (1996, 2011), or by Finley (2009), who prefers to use Anchoring constraints rather than Alignment constraints (using the term "lexical domain"), but notes that the two types are almost identical in accounting for morphological feature realization. None of these researchers discuss the EAS plural. We note that while the constraint in (7) would be trivially (nonfatally) violated in any word that does not have the floating [−ATR]PL feature in its underlying representation, the constraint plays a central role in plural word forms requiring that plural exponence be manifested in word-final position. The constraint in (7) is similar to a RealizeMorpheme type constraint (Kurisu 2001); in Section 3.3 we will incorporate such a constraint in accounting for the laxing pattern shown by the data in (4a).

In order to ensure that this [−ATR]PL feature spreads leftward, we will set up a second alignment constraint, a gradient one, that will align the left edge of the feature [−ATR]PL to the left edge of the prosodic word, as in (8):

(8) AlignPL-L: Align ([−ATR]PL, Left, Word, Left)
The plural morpheme [−ATR]PL is aligned with the left edge of the prosodic word.

We refer here to prosodic word in order to account for the data in (6) where spreading only occurs to the rightmost member of a plural compound. As the left alignment constraint is evaluated in a gradient manner, a candidate with the first vowel in the word having the [–ATR]PL feature will completely respect the constraint, but if the second vowel of the word is [–ATR] but not the first one then the constraint is violated once; a candidate where the first [–ATR] vowel in the word is two positions from the leftmost vowel will violate the constraint twice, etc. Since consonants in general do not take the feature [–ATR], distinguishing it from [RTR] which can affect both consonants and vowels (see Goad 1991; Davis 1995), we only consider vowels in determining the gradiency of the constraint in (8).

Further, in order to account for the feature change when [–ATR] is realized on a vowel, we need a faithfulness constraint that will militate against changing the input [+ATR] value in the output; this is shown in (9) and has been used in the analysis of Jiménez and Lloret as well as in other optimality-theoretic analyses of EAS laxing:

(9) ID[ATR]: Identity [ATR]
A segment does not change its feature value for [ATR].

Using these three constraints, we can examine four potential candidates for the underlying representation of the plural form of *nene* /ˈne.ne/ 'boy' and explain why the optimal candidate succeeds, in Tableau 1 where the plural morpheme in the input is shown by the floating morphemic feature [–ATR]PL.

$[–ATR]_{PL}$

**Tableau 1.** *nenes* /ˈne.ne/ 'boys'.

| $[–ATR]_{PL}$ /ˈne.ne/ | AlignPL-R | AlignPL-L | ID[ATR] |
|---|---|---|---|
| a. ˈne.ne | *! | | |
| b. ˈne.nɛ | | *! | * |
| ☞c. ˈnɛ.nɛ | | | ** |
| d. ˈnɛ.ne | *! | | * |

Ranking from Tableau 1: *AlignPL-R*, *AlignPL-L* >> *ID[ATR]*.

This tableau demonstrates a hierarchical ranking, placing both AlignPL-R and AlignPL-L above ID[ATR]. If the ranking were to be reversed such that AlignPL-R was the lowest ranked constraint, candidate (a) would win. If AlignPL-L was the lowest ranked constraint, candidate (b) would win. Therefore, we need to have ID[ATR] as the lowest ranked of these three, although the ranking between the two alignment constraints is as yet undetermined. Regarding the interpretation of the alignment constraints, candidates (a) and (d) in the tableau violate the categorical constraint AlignPL-R because there is no manifestation of the floating [–ATR]PL feature at the right edge of the word. Candidate (a) does not violate the gradient constraint AlignPL-L, since there is no [–ATR]PL vowel in the output word and so it is vacuously satisfied. Candidate (b) violates the AlignPL-L constraint because the [–ATR]PL feature is not extended to the first vowel of the word. Note that, because candidate (d) also violates ID[ATR], it is harmonically bounded by candidate (a), meaning that candidate (d) could only win if another constraint were introduced that (d) respects and (a) violates.

The next step for the analysis of mid vowels is to describe the selection of the winning candidate for words with more than two syllables. In this case, as we will see in Tableau 2 for *telefono* 'telephone' from (1d), we need an additional constraint to select a candidate with the feature [–ATR] spreading to word-internal vowels. We use the NoGap constraint after Archangeli and Pulleyblank (1994), McCarthy (1997), and used for EAS by Soriano (2012). For purposes of this paper, we relativize NoGap to [–ATR] spreading described in (10):

(10) NoGap
In a string of more than two vowels, the feature [–ATR] cannot skip over a medial vowel.

[–ATR]$_{PL}$

**Tableau 2.** *teléfonos* /te.'le.fo.no/ 'telephones'.

| [–ATR]$_{PL}$ /te.'le.fo.no/ | AlignPL-R | AlignPL-L | NoGap | ID[ATR] |
|---|---|---|---|---|
| a. te.'le.fo.no | *! | | | |
| b. tɛ.'le.fo.nɔ | | | *!* | ** |
| ☞c. tɛ.'lɛ.fɔ.nɔ | | | | **** |
| d. te.'lɛ.fɔ.nɔ | | *! | | *** |

Ranking from Tableau 2: *AlignPL-R*, *AlignPL-L*, *NoGap* >> *ID[ATR]*.

In Tableau 2, each of the three constraints must be ranked above ID[ATR] to ensure that candidate (c), in which all vowels are [–ATR], is selected as the winning candidate. If ID[ATR] were to outrank any other constraint, including NoGap, one of the other candidates would be incorrectly selected as winner.

*3.2. Treatment of Low Vowels*

In order to explain the tendency of EAS low vowels to not become [–ATR] in non-word-final position, we need a markedness constraint that militates against [–ATR] in all low vowels. This constraint is described in (11) and is crucially ranked with respect to the two alignment constraints as shown by the evaluation of the plural form of /mapa/ from the data in (2a) in Tableau 3 below:

(11) *[–ATR, +low][4]
The feature combination of [–ATR] and [+low] cannot cooccur on the same phoneme.

[–ATR]$_{PL}$

**Tableau 3.** *mapas* /'ma.pa/ 'maps'.

| [–ATR]$_{PL}$ /'ma.pa/ | AlignPL-R | *[–ATR] [+low] | AlignPL-L | NoGap | ID[ATR] |
|---|---|---|---|---|---|
| a. 'ma.pa | *! | | | | |
| b. 'mæ.pæ | | **! | | | ** |
| c. 'mæ.pa | *! | * | | | * |
| ☞ d. 'ma.pæ | | * | * | | * |

Ranking emerging from Tableau 3: *AlignPL-R* >> *[*–ATR +low*] >> *AlignPL-L*.

Tableau 3 provides evidence for the crucial ranking of various constraints. The comparison of candidate (a) with the winning candidate (d) shows that AlignPL-R must outrank *[–ATR +low] since the reverse ranking would wrongly result in (a) being the winner. Moreover, we see that AlignPL-L must be crucially ranked below the low vowel markedness constraint *[–ATR +low], or else candidate (b) would wrongly be the winner. As a result, the only unranked constraints in the above tableau are AlignPL-L in relation to NoGap; we know from Tableau 2 that NoGap must outrank ID[ATR] placing ID[ATR] at the lowest spot in the hierarchy so far.

Let us extend this analysis now to a word with three syllables and observe the role a high-ranked faithfulness constraint must play to prevent against changing the height of vowels to create permissible [–ATR] vowels in non-word-final position. This constraint is described in (12) and militates against a change in the vowel height features [high] and [low]:

(12) IdentIO(VowelHeight)
The vowel height feature(s) of an input segment is unchanged in the output.

In Tableau 4, we see in the plural form of *bañera* 'bathtub' from (2c) that the vowel height faithfulness constraint must come into play to prevent candidate (f) from winning out. This candidate violates IdentIO(VowelHeight) because the [+low] vowel in the initial syllable of the input is realized as [–low] in the output. The IdentIO(VowelHeight) constraint is undominated because there are no instances of vowel height feature change reported in the research centered around EAS. In subsequent tableaux we will not show this constraint. Regarding the evaluation of the alignment constraints, both candidates (a) and (d) in Tableau 4 violate AlignPL-R given its formulation in (7) since there is no [–ATR]$_{PL}$ feature aligned with the right edge of the word. Candidate (a), though, vacuously satisfies the AlignPL-L constraint in (8) in that there is no [–ATR]$_{PL}$ feature to left-align while candidate (d) has one violation of AlignPL-L since the [–ATR]$_{PL}$ feature surfaces only on the second vowel in the word thus incurring one violation of the AlignPL-L constraint as shown.

[–ATR]$_{PL}$

**Tableau 4.** *bañeras* /ba.ˈɲe.ɾa/ 'bathtubs'.

| [–ATR]$_{PL}$ /ba.ˈɲe.ɾa/ | IdentIO (Vowel Height) | AlignPL-R | *[–ATR] [+low] | AlignPL-L | NoGap | ID[ATR] |
|---|---|---|---|---|---|---|
| a. ba.ˈɲe.ɾa | | *! | | | | |
| b. ba.ˈɲe.ɾæ | | | * | **! | | * |
| c. bæ.ˈɲɛ.ɾæ | | | **! | | | *** |
| d. ba.ˈɲɛ.ɾa | | *! | | * | | * |
| ☞ e. ba.ˈɲɛ.ɾæ | | | * | * | | ** |
| f. bɛ.ˈɲɛ.ɾɛ | *! | | | | | *** |

Ranking emerging from Tableau 4: *IdentIO(VowelHeight) >> *[–ATR +low]*.

### *3.3. Treatment of High Vowels*

Here we discuss the high vowel data that were shown in (3) and (4a), where word forms ended in a stressless high vowel. The discussion of the data in (4b) where words end in a stressed high vowel will be considered in Section 4. Regarding the high vowel data in (3) and (4a), we will need a markedness constraint that militates against high vowels having the feature [–ATR], which we provide in (13) based on Archangeli and Pulleyblank (1994). This constraint, which we consider to be highly ranked, is commonly used in analyses of EAS vowel harmony. Consider the role of this constraint in Tableau 5, where we first examine its role on word-internal high vowels.

(13) *[–ATR, +high]
The feature combination of [–ATR] and [+high] cannot cooccur on the same phoneme.

[–ATR]$_{PL}$

**Tableau 5.** adjetivos /ad.xe.ˈti.vo/ 'adjectives'.

| [–ATR]$_{PL}$ /ad.xe.ˈti.vo/ | *[–ATR] [+high] | AlignPL-R | *[–ATR] [+low] | AlignPL-L | NoGap | ID[ATR] |
|---|---|---|---|---|---|---|
| a. ad.xe.ˈti.vo | | *! | | | | |
| b. æd.xɛ.ˈtɪ.vɔ | *! | | * | | | **** |
| c. æd.xɛ.ˈti.vɔ | | | *! | | * | *** |
| ☞ d. ad.xɛ.ˈti.vɔ | | | | * | * | ** |
| e. ad.xe.ˈti.vɔ | | | | **!* | | * |
| f. ad.xɛ.ˈtɪ.vɔ | *! | | | * | | *** |

Ranking emerging from Tableau 5: **[–ATR +high] >> NoGap*; *AlignPL-L >> NoGap*.

For the plural of *adjetivo* 'adjective' above, from the data in (3c), we see that the gradient nature of AlignPL-L is necessary to ensure that a candidate like (e) be distinguished from the winning candidate. If AlignPL-L were not evaluated gradiently then (e) would wrongly be the winner. Further, the comparison between the winning candidate (d) with (e) also shows that AlignPL-L must outrank NoGap or else (e) would wrongly be the winner.

While, just based on the above tableau, it may be unclear where to rank *[–ATR +high], it must be higher-ranked than NoGap lest (f) be the winning candidate. Moreover, *[–ATR +high] must be higher ranked than AlignPL-R because of the example of *tribu* 'tribe' in (4a), where the singular and plural are the same, ['tri.bu], with no laxing in the plural, meaning that AlignPL-R is violated in order to satisfy *[–ATR +high]. Additionally, this also entails that *[–ATR +high] is higher ranked than a Realize Morpheme constraint (Kurisu 2001), specifically RealizeMorpheme-PL (RM-PL, a plural morpheme in the input must be realized in the output) given that there is no overt exponence of the plural in ['tri.bu]. As shown in Tableau 6 for *poli* 'cops', RM-PL plays an important role in the evaluation of the words in (4a). We place it alongside AlignPL-R in the tableau, but it only needs to be critically ranked below *[–ATR +high] because of *tribus* ['tɾi.bu] 'tribes' (in Tableau 6, we do not show *[–ATR +low], as it does not apply).

$[–ATR]_{PL}$

**Tableau 6.** *poli* /'poli/ 'cops'.

| $[–ATR]_{PL}$ /'po.li/ | *[–ATR] [+high] | AlignPL-R | RM-PL | AlignPL-L | NoGap | ID[ATR] |
|---|---|---|---|---|---|---|
| a. 'po.li | | * | *! | | | |
| b. 'po.lɪ | *! | | | * | | * |
| c. 'pɔ.lɪ | *! | | | | | ** |
| ☞d. 'pɔ.li | | * | | | | * |

Ranking emerging from Tableau 6: **[–ATR +high] >> AlignPL-R*; *RM-PL >> ID[ATR]*.

This tableau demonstrates the crucial ranking of *[–ATR +high] over AlignPL-R. If the ranking of these two constraints were reversed, (c) would wrongly be the winner. Thus, *[–ATR +high] crucially outranks AlignPL-R, making it the highest-ranked constraint along with the undominated faithfulness constraint IDIO(Vowel Height).[5] Crucially, the RM-PL constraint must outrank the ID[ATR] constraint or else candidate (a) would be the winner. Consequently, the evaluation of *polis* 'cops' shows a necessity for the RM-PL constraint in addition to AlignPL-R, but since RM-PL only plays a role for data like those in (4a) we do not otherwise show its function.

### *3.4. Treatment of Words Ending in Final Consonants*

In the case of nouns (and adjectives) ending in a word-final consonant like those in (5), we posit that vowel epenthesis is triggered in the plural so that the floating [–ATR]PL feature can be realized on the right edge of the word, given that it cannot be realized on a consonant. In the example *carbón/carbones* [karbon]—[karbɔnɛ] 'carbon (sg.)/(pl.)'), there is no underlying plural allomorph /s/ or /es/. Rather, the mid front vowel, which acts as a default vowel generally in Spanish phonology, is epenthesized at the end of the word so that AlignPL-R can be satisfied. This entails that AlignPL-R outranks the constraint Dep-V, which militates against vowel insertion. The constraint is shown in (14).

(14) Dep-V
A vowel in the output has a correspondent in the input.

A complication comes up with singular words ending in a lexical /s/ such as *mes* 'month' in (5a). The singular form of the word *mes* /mes/ 'month' is [mɛ] with a deleted /s/ and not with an epenthetic vowel after the /s/. Given that researchers such as Soriano (2012) have noted that EAS has a coda condition that disallows the surfacing of /s/ in coda position (a constraint that we will refer to as NoCoda-s), the fact that the singular is [mɛ] and not the hypothetical [me.se] with epenthesis means that the Dep-V constraint in (14) must outrank the Max-C constraint in (15)

(15) Max-C
A consonant in the input must be realized in the output

That the vowel in the actual output of the singular [mɛ] is laxed is understood by Soriano (2012), and other researchers, most recently by Walker (2024), as coming from the deleted /s/. That is, when the /s/ deletes, it leaves behind a [spread glottis] gesture that coalesces with the immediately preceding tense [+ATR] vowel, resulting in a lax [–ATR] vowel. These researchers posit a high-ranked Max-Gesture constraint that the surfacing candidate [mɛ] respects but would be violated by the alternative *[me] where there is no remnant of the /s/ gesture. Given this, we posit the following tableau, Tableau 7, for the singular *mes* 'month'.

**Tableau 7.** *mes* /ˈmes/ 'month'.

| /ˈmes/ | Max-Gesture | NoCoda-s | Dep-V | Max-C | ID[ATR] |
|---|---|---|---|---|---|
| a. ˈmes | | *! | | | |
| b. ˈme | *! | | | * | |
| c. ˈme.se | | | *! | | |
| ☞d. ˈmɛ | | | | * | * |

Ranking emerging from Tableau 7: *Dep-V >> Max-C; Max-Gesture >> ID[ATR], NoCoda-s >> Max-C, ID[ATR].*

With this ranking, along with the ranking previously established with respect to the plural forms, we now consider the plural of *mes* 'month' which is realized as [mɛsɛ] with the [–ATR]PL feature on the final epenthetic vowel and not as *[mɛ] with deletion like the singular. The hypothetical plural *[mɛ] is a realistic alternative for the plural since it would seem to satisfy AlignPL-R given the ranking of Dep-V over Max-C. Tableau 8 below illustrates the analysis of the realization of the plural of *mes* 'month' where the winning candidate does indeed have a final epenthetic vowel.

[–ATR]$_{PL}$

**Tableau 8.** *meses* /ˈmes/ 'months'.

| [–ATR]$_{PL}$ /ˈmes/ | Max-Gesture | NoCoda-s | AlignPL-R | Dep-V | AlignPL-L | Max-C | ID[ATR] |
|---|---|---|---|---|---|---|---|
| a. ˈmes | | *! | * | | | | |
| b. ˈme | *! | | * | | | * | |
| c. ˈmɛ | *! | | | | | * | * |
| d. ˈmɛ | | | *! | | | * | * |
| e. ˈme.se | | | *! | * | | | |
| f. ˈme.sɛ | | | | * | *! | | |
| ☞g. ˈmɛ.sɛ | | | | * | | | * |
| h. ˈɛmɛsɛ | | | | **! | | | * |

Ranking emerging from Tableau 8: *Max-Gesture* >> *Dep-V*; *AlignPL-R* >> *Dep-V*; (*Dep-V* >> *AlignPL-L*).

The first two candidates in Tableau 8 are eliminated because of the violation of AlignPL-R with the first candidate also violating NoCoda-s and the second violating Max-Gesture. The interesting candidates are (c) and (d), both phonetically [mɛ] but different in that the [–ATR] feature on the vowel in candidate (c) comes from the realization of the input [–ATR]$_{PL}$ feature while in candidate (d) the [–ATR] feature is the remnant gesture of the deleted lexical /s/. The former critically violates Max-Gesture while the latter critically violates AlignPL-R. (We assume an independent constraint would rule out a potential candidate where two different [–ATR] features are realized on the same vowel.) The winning candidate (g), violating Dep-V, surfaces with a final epenthetic vowel, on which the plural [–ATR]$_{PL}$ feature is realized.

Tableau 8 illustrates three crucial rankings. First, Max-Gesture outranks Dep-V, since the reverse ranking would result in candidate (c), [mɛ], wrongly being the winner. Second, in our synchronic analysis, the inserted plural vowel is the mid front vowel because it is generally the epenthetic vowel in Spanish that, for example, breaks up impermissible word-initial consonant clusters. It is inserted so that AlignPL-R can be satisfied. This shows the crucial ranking of AlignPL-R over Dep-V or else candidate (d) would be the winner. Finally, note that from a technical perspective Dep-V should be ranked above AlignPL-L to avoid the possibility of initial vowel insertion, as in candidate (h). While we have not been considering consonant-initial plural words to violate AlignPL-L when the [–ATR] plural feature is realized on the vowel that is immediately after the word-initial onset, as seen for the winning candidate in (g), if it does violate it, then the Dep-V constraint would rule out (h) as the winning candidate. Further, as a reviewer has noted, this ranking of Dep-V over AlignPL-L would be needed in the plural of any word that begins with a vowel that cannot host [–ATR] so that initial epenthesis of [ɛ] would not occur.

In Tableau 9, for the plural of *empujón* 'push' from (5b), we see that the same ranking also works when the constraints NoGap and *[–ATR +high] are brought into play.

[–ATR]$_{PL}$

**Tableau 9.** *empujones* /em.pu.'xon/ 'pushes'.

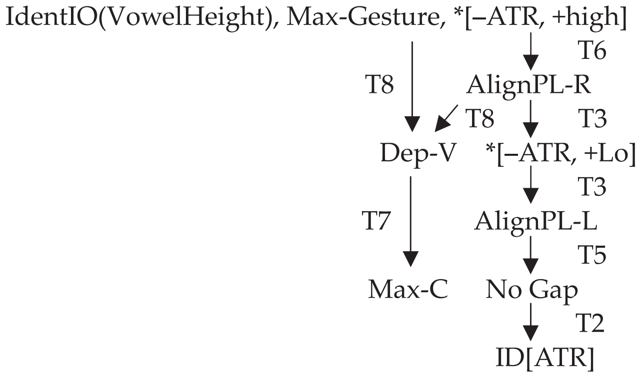
| [–ATR]$_{PL}$ /em.pu.'xon/ | Max-Gesture | *[–ATR] [+high] | Align PL-R | Dep-V | Align PL-L | NoGap | Max-C | ID[ATR] |
|---|---|---|---|---|---|---|---|---|
| a. em.pu.'xon | | | *! | | | | | |
| b. ɛm.pʊ.'xɔn | | *! | * | | | | | *** |
| c. ɛm.pu.'xɔ | *! | | | | | * | * | ** |
| d. ɛm.pu.'xɔn | | | *! | | | * | | ** |
| e. ɛm.pu.'xɔ.ne | | | *! | * | | * | | ** |
| ☞f. ɛm.pu.'xɔ.nɛ | | | | * | | * | | ** |
| g. em.pu.'xo.nɛ | | | | * | *!** | | | |
| h. ɛm.pu.'xo.nɛ | | | | * | | **! | | * |

In this tableau, we see that the constraints established from the other tableaux account for candidate (f), [ɛm.pu.'xɔ.nɛ], being the winning candidate where the [–ATR]$_{PL}$ floating feature is realized on the epenthetic vowel and then spreads leftward but skips the high vowel /u/ because of the constraint militating against high lax vowels. The evaluation also provides further evidence for the ranking of AlignPL-L over NoGap that emerged from Tableau 5. If this ranking were reversed, then candidate (g) in Tableau 9 without leftward spread would wrongly be the winning candidate. (A modified candidate (c) with nasalization on the final vowel, thus respecting the Max-Gesture constraint, would be ruled out by a high-ranked Max-NasalConsonant constraint, since nasal consonants never delete in EAS, even in coda position.)

Given this analysis of consonant final nouns and adjectives where the floating plural [–ATR]$_{PL}$ feature triggers right edge epenthesis so that the AlignPL-R constraint can be satisfied, we now must reconsider the plural of words like those in (4a), such as /es.'pi.ɾi.tu/ and /poli/, analyzed in Tableau 6 above, where the winning candidate for the plural *polis* 'cops' ['pɔ.li] in Tableau 6 violates AlignPL-R, since the final vowel is not lax. A possible alternative with vowel epenthesis, such as ['pɔ.liɛ] or [ɛs.'pi.ɾi.tu̯ɛ], is hypothetically possible given that Spanish words can end in a vocalic high-mid sequence, as in the words *pie* [pi̯ɛ] 'foot' and *fue* ['fu̯e] 'he went.' While we will not show a revised tableau with such plural candidates, we suggest that what rules them out is a high-ranked constraint militating against word-final vocalic sequences. If this is higher ranked than AlignPL-R, then plural forms with final diphthongs would not surface. A form like ['fu̯e] from underlying /'fue/ would surface with a final vocalic sequence because of a higher ranked Max-Vowel constraint that would prevent the deletion of an underlying vocalic element. The underlying plural for the words in (4a), like /poli/ and /es.'pi.ɾi.tu/, do not have final vowel sequences in their underlying representation.

We can summarize our optimality-theoretic analysis in this section by the Hasse diagram below in (16). In the diagram, the top row comprises the undominated constraints with respect to the analysis of the Granada EAS plural presented. The lower rows are the dominated constraints with reference to the Tableau (T) from which the ranking arguments emerge. (We do not show RM-PL which, as mentioned, would be ranked somewhere lower than *[−ATR, +high] and higher than ID[ATR].)

(16) Hasse Diagram of the Constraint Rankings




While there are other rankings from the tableaux such as the Align-PL constraints outranking ID[ATR] as in Tableau 1, the Hasse diagram in (16) makes clear the strict domination of the constraints in our alignment-based analysis of the EAS plural.

## 4. Treatment of Words Ending in Final Stressed Vowels

In this section we consider nouns (and adjectives) that end in a final vowel that carries the main stress. Most of these words are borrowed and many are rare in the plural form. The set of words ending in a stressed high vowel, presented in (4b), are repeated in (17), along with examples of words that end in non-high stressed vowels.

(17) Plural of words Ending in a Stressed Vowel (final position)

| | Orthography | Singular | Plural | Gloss |
|---|---|---|---|---|
| a. | *club* | ['klu] | ['klʊ] | club |
| b. | *clip* | ['kli] | ['klɪ] | clip |
| c. | *menú* | [me.'nu] | [me.'nʊ] | menu |
| d. | *hindú* | [in.'du] | [in.'dʊ] | Hindu |
| e. | *papá* | [pa.'pa] | [pa.'pæ] | father |
| f. | *bebé* | [be.'be] | [be.'bɛ] | baby |
| g. | *olé* | [o.'le] | [o.'lɛ] | olé |
| h. | *yoyo* | [yo.'yo] | [yo.'yɔ] | yoyo |

The data in (17) are interesting and are little discussed in the literature on EAS plurals. For example, Henriksen (2017) did not examine such words. The occurrence of nouns (or adjectives) ending in a single final stressed vowel is unusual in Spanish and is typically found in borrowed words. According to Martínez-Paricio (2021), this pattern occurs in less than one percent of Spanish non-verbal forms. It is assumed here that such words would come with an underlying stress indicated on the final vowel. The EAS plural of such words follows an observation by Herrero de Haro (2020, p. 4) that "words that have stress on the last syllable only lower one vowel when [plural] /-s/ is deleted in EAS". He gives the example of *papas* 'fathers' where the word-final /a/ of the EAS plural laxes but the unstressed first vowel does not (optionally) harmonize. Based on mentions of this in the literature, as well as on our own native speaker consultation, we view these words as having only final laxing, including the laxing of a word-final high vowel. Thus, the plural of the words in (17a-d) with a final stressed high vowel is different from the plural of the words in (4a) (such as *poli*) which end in a stressless high vowel in that a final stressed high vowel laxes in the plural whereas an unstressed one does not. The two plural word forms in (17c) and (17d) have been discussed for a non-Andalusian variety of Peninsular Spanish by

Bermúdez-Otero (2007), who distinguishes the plural of these two forms as *menús* [me.'nus] and *hindúes* [in.'du.es], differentiating them by showing that [me.'nu] is athematic where [in.'du] is viewed as being an e-stem adjective; thus [e] appears in the plural. While we are aware that some EAS speakers can have the alternative plural [in.'du.ɛ] for [in.'du], we are not certain whether EAS makes this distinction, as this alternative could be influenced by normative Spanish given the infrequency of the word to most speakers. The matter is in need of further research.

What is truly unique about the plurals in (17), as is most clearly seen by (17c) and (17f–h), is that when the final stressed vowel laxes in the plural, [−ATR] spreading to prior vowels is prevented even if the vowel is mid. As far as we are aware, nouns ending in stressed mid vowels have not been systematically discussed in the analysis of EAS plurals. Moreover, we believe that these data present a problem for the previous optimality-theoretic analyses, though we will not explore it in detail. As one example, Granada EAS plural forms like *bebés* [be.'bɛ] 'babies' are different from cases of final lexical coda deletion, where, as noted by Jiménez and Lloret (2020, p. 103), a word like *revés* 'other side' can be pronounced with both vowels laxed. This distinguishes laxing due to phonological coda deletion from laxing due to plural exponence.

Further, the analysis we developed in Section (3) can neither account for the plural data ending in a stressed high vowel that undergoes laxing, nor can it account for the plurals in (17f-h), where the final mid vowel is lax but there is no subsequent spreading. Under the analysis we developed, the evaluation of the plural of *bebé* should be exactly like the plural of *néne* shown in Tableau 1 where the winning candidate has laxing in both vowels. Here, we will make a first attempt at developing an alignment analysis to account for the data in (17). To see how the detailed analysis that we have developed in Section 3 with strict dominance of constraints predicts the wrong winner for a form like the plural of *bebé* [be.'be] 'baby' in (17f), consider Tableau 10 where * indicates the unintended winner.

[−ATR]$_{PL}$

**Tableau 10.** *bebés* /be.'be/ 'babies' (intended winner [be.'bɛ]).

| [−ATR]$_{PL}$ /be.'be/ | Max-Gesture | *[−ATR] [+high] | Align PL-R | Dep-V | Align PL-L | NoGap | ID[ATR] |
|---|---|---|---|---|---|---|---|
| a. be.'be | | | *! | | | | |
| b. be.'bɛ | | | | | *! | | * |
| *c. bɛ.'bɛ | | | | | | | ** |
| d. bɛ.'be | | | *! | | | | * |

As we see in Tableau 10, the actual plural (b) violates AlignPL-L, which we have shown to be higher ranked than ID[ATR] in Section 3. Moreover, our current analysis cannot account for the plural of the word forms in (17a–d) that end in a high stressed vowel where the final high vowel laxes given the ranking of *[−ATR +high] over AlignPL-R. This would seem to contradict the ranking that was established in Tableau 6.

What the data in (17) really show is that plurals of words that end in a final stressed vowel have a special status, perhaps because of their unusualness as a nonverbal stress pattern in Spanish. From a metrical perspective, these words end in a monomoraic degenerate foot, under the assumption that Spanish coda consonants are moraic at some level of analysis. Spanish foot structure is typically viewed as being trochaic (e.g., Piñeros 2016). In order to account for the observation of the plural data in (17) that the floating plural feature [−ATR]PL appears on a word-final stressed vowel regardless of its vowel height, we posit a high-ranked alignment constraint that forces the floating plural feature [−ATR]PL to appear on a word-final stressed vowel regardless of its vowel height (and so outranks *[−ATR, +high]). This is shown in (18).

(18) AlignFootμ, PL-R: Align (Monomoraic Foot, Right, [–ATR]PL, Right)
The right edge of the monomoraic foot is aligned with the right edge of the plural morpheme [–ATR]PL.

The constraint in (18) is only applicable to nouns (or adjectives) that have a monomoraic foot. Since a monomoraic foot in Spanish would only occur in words ending in a stressed vowel, the constraint is applicable to the plural of all the words in (17) but is not pertinent to any of the data presented earlier in (1)–(6) (4b excluded), since none of those forms have a monomoraic foot. The constraint in (18) would be higher ranked than the feature cooccurrence constraint *[–ATR, +high] so that a final vowel would become lax even if it is a high vowel.

While the constraint in (18) accounts for the appearance of [–ATR] on any final stressed vowel, we need another constraint to account for the lack of leftward spreading of the [–ATR]PL feature to a prior syllable. Here we reference and slightly modify a crisp edge constraint proposed in Kaplan (2018) in an analysis of centralization (laxing) in the Romance variety Tudanca Montañés. His constraint CrispEdge ([–ATR], stressed syllable, L) militates against the extension of [–ATR]PL beyond the left edge of a stressed syllable. We make use of a similar crisp edge constraint in (19), but relativize it to a monomoraic foot.

(19) CrispEdge ([–ATR]PL, Ftμ, L)
A monomoraic foot's [–ATR]PL feature cannot extend beyond the left edge of that foot.

The constraint in (19) prevents leftward spreading in the plural forms of the words in (17). The constraint would be high-ranked and crucially needs to reference the plural morpheme feature ([–ATR]PL and not [–ATR] generally. This must be the case to explain why, as noted by Jiménez and Lloret (2020, p. 103), a word like *revés* 'other side' can be pronounced with both vowels laxed due to the deletion of the lexical /s/ in what would be coda position. Tableaux 11 and 12 show our evaluation of the plural forms of *menú* and *bebé*, respectively.

$[–ATR]_{PL}$

**Tableau 11.** *menús* /me.'nu/ 'menus'.

| $[–ATR]_{PL}$/ me.'nu/ | AlignFoot$_{\mu}$,PL-R | CrispEdge $[–ATR]_{PL}$, Ft$_{\mu}$, L | *[–ATR [+high] | Align PL-R | *[ATR] [+low] | Align PL-L | NoGap | ID[ATR] |
|---|---|---|---|---|---|---|---|---|
| a. me.'nu | *! | | | * | | | | |
| ☞ b. me.'nʊ | | | * | | | * | | * |
| c. mɛ.'nʊ | | *! | * | | | | | ** |
| d. mɛ.'nu | *! | | | * | | | | * |

Ranking emerging from Tableau 10: *AlignFoot$_{\mu}$, PL-R >> *[–ATR +High]; CrispEdge ($[–ATR]_{PL}$, Ft$_{\mu}$, L) >> AlignPL-L.*

$[–ATR]_{PL}$

**Tableau 12.** *bebés* /be.'be/ 'babies'.

| $[–ATR]_{PL}$/ be.'be/ | AlignFoot$_{\mu}$, PL-R | CrispEdge $[–ATR]_{PL}$, Ft$_{\mu}$, L | *[ATR] [+high] | Align PL-R | *[ATR] [+low] | Align PL-L | NoGap | ID[ATR] |
|---|---|---|---|---|---|---|---|---|
| a. be.'be | *! | | | * | | | | |
| ☞ b. be.'bɛ | | | | | | * | | * |
| c. bɛ.'bɛ | | *! | | | | | | ** |
| d. bɛ.'be | *! | | | * | | | | * |

In Tableau 11, we see that the ranking of AlignFootμ, PL-R above *[–ATR, +high] forces a final high vowel to become lax when it is stressed. Candidate (a) in each tableau violates AlignFootμ, PL-R since the monomoraic foot does not have the [–ATR]PL feature on its right edge. In both tableaux, we see that spreading to the first vowel is prevented by the

Crisp Edge constraint, which militates against the sharing of the plural [–ATR]PL feature with any segment to the left of the monomoraic foot boundary. The addition of these two constraints focusing on a word-final stressed vowel expressed in terms of monomoraic foot structure does not affect the ranking of the other constraints and seems to us to be a realistic way of dealing with the rare nouns (and adjectives) that have vowel-final stress.[6]

In this section, we have extended the alignment-based analysis of the Granada EAS plural developed in Section 3 to the plural of nouns (and adjectives) that end in a monomoraic foot (i.e., end in a stressed vowel). As far as we are aware, these data as reflected by the word forms in (17) have not previously been systematically analyzed. The hierarchy that we have developed still applies, but with the addition of the two high-ranked constraints in (18) and (19) that have applicability just to words ending in a monomoraic foot. In the next section, we briefly discuss how our analysis can be extended to some of the different reported varieties of plural laxing in EAS.

## 5. Extension to Other Plural Patterns in EAS

The data patterns that we presented have largely followed Henriksen (2017). However, we extended our analysis to account for the plural of words ending in stressed vowels, which were not considered by Henriksen. Our analysis is different than previous analyses in not assuming that there is a suffixal plural /s/ in the underlying representation of plural word forms; rather, from the synchronic perspective, we have posited that the EAS plural is marked by a floating [–ATR]PL autosegment that is realized at the right edge of the word by the alignment constraint in (7) AlignPL-R and that spreads leftward due to the gradient alignment constraint AlignPL-L in (8). The alignment constraints interact with the feature co-occurrence constraints, the NoGap constraint, and the various faithfulness constraints as indicated by the Hasse diagram in (16) to produce the Granada EAS pattern consistent with Henriksen (2017). We now discuss how our analysis can be applied to the other patterns of plural laxing reported for EAS.

While the plural laxing in Granada EAS had not been previously analyzed independent of coda deletion and thus not viewed as morphologically conditioned, the Jaén variety of EAS as analyzed by Soriano (2012) and Jiménez and Lloret (2020) is agreed to be a subdialect with morphologically conditioned laxing. This is because the pattern found with vowel laxing specifically in the plural in Jaén is different than what is found for lexical coda /s/ deletion. In plural word forms, all vowels are laxed regardless of their height. Two examples from Jaén mentioned by Jiménez and Lloret (2020, p. 105) include the plural of *cómico* 'comic, masc.' as [ˈkɔ.mɪ.kɔ] with all vowels lax, including the high vowel, and the plural of *asa* 'handle' as [ˈæ.sæ] with both low vowels laxed. Plural laxing is different than lexical /s/ coda deletion in Jaén in that the plural laxing domain is the (prosodic) word while laxing due to coda deletion is local, so that only the vowel immediately preceding the deleted coda is laxed, thus providing a compelling case that the two types of laxing are quite different. This difference is exemplified by the comparison between *nenes* 'boys' pronounced as [ˈnɛ.nɛ] with both vowels [–ATR] and the word *jueves* 'Thursday' pronounced as [ˈhwe.ßɛ], where only the last vowel is lax due to the deletion of lexical /s/ in word-final coda position. In Jaén, its expected production in the plural would be with the last vowel lax and the first vowel tense. Laxing would occur because of the Max-Gesture constraint, while Ident-ATR would be higher ranked than any constraint that would induce spreading or copying and lower-ranked than Max-Gesture. In our analysis of plural laxing, which differs from Soriano's analysis in that we have a floating [–ATR]PL feature rather than a lexical /s/ as the plural morpheme, our alignment constraints developed in Section 3 in (7) and (8) repeated below as (20) and (21) for convenience would be high-ranked, outranking any of the feature cooccurrence constraints, thus making all vowels in the plural domain lax.

(20) AlignPL-R: Align (Word, Right, [–ATR]PL, Right)
The right edge of the grammatical word is aligned with the right edge of the plural morpheme [–ATR]PL.

(21) AlignPL-L: Align ([−ATR]PL, Left, Word, Left)
The plural morpheme [−ATR]PL is aligned with the left edge of the prosodic word.

Turning to Granada EAS, in Section 3 we analyzed this variety based on the description in Henriksen (2017), then extended the analysis to account for plural forms ending in stressed vowels in Section 4. An important observation about these latter plurals, as shown by the data in (17), is that, when the final vowel is stressed, laxing is limited to that final vowel. We interpreted this as a limitation on leftward spreading from a degenerate (monomoraic) foot. One of the subpatterns of Grenada EAS plural laxing mentioned by Jiménez and Lloret (2020) is that laxing does not extend leftward beyond the stressed syllable, meaning that the first two vowels need not be laxed in *económicos* 'economic, masc. pl.' (i.e., [e.ko.ˈnɔ.mi.kɔ]) and spreading of [−ATR] does not go beyond the stressed syllable. As mentioned in Endnote 5, we account for this using the more general crisp edge constraint CrispEdge [−ATR]PL, Ft, L] that prevents spreading to pretonic syllables along with its variable ranking with respect to the AlignPL-L constraint.

A different case of variation in the Granada variety of EAS discussed by Jiménez and Lloret occurs in proparoxytone plurals (i.e., plurals with antepenultimate stress). In these forms, they note that the penultimate vowel does not have to be lax. This is exemplified by the plural of *trébol* 'clover', which can be pronounced as [ˈtrɛ.ßo.lɛ]. Given our alignment approach to the EAS plural, the final vowel of the plural form [ˈtrɛ.ßo.lɛ] has to lax because of the high-ranked AlignPL-R constraint. The antepenultimate vowel undergoes laxing due to the AlignPL-L constraint; the lack of laxing in the penultimate vowel would result from the ID[ATR] constraint being higher-ranked than NoGap, thus preventing the laxing of /o/. The variation reported for the plural of *trébol* as either [ˈtrɛ.ßo.lɛ] or [ˈtrɛ.ßɔ.lɛ] could then be accounted for by the indeterminacy of the two lowest ranking constraints ID[ATR] and NoGap, given our hierarchy of constraints shown in (16). A similar analysis of this variation is given in Jiménez and Lloret, whereby there is variable ranking between ID[ATR] and a LICENSE (−ATR, Ft) constraint, which entails a different conception of EAS laxing.

Finally, we consider how our alignment analysis of the EAS plural can be extended to the pattern found in the Murcian variety reported by Jiménez and Lloret (2020). While plural laxing in this variety is considered to be phonological, in that any non-nasal coda consonant can delete and triggers laxing in a way that is similar to the plural, we will consider the plural to be different in that it is marked by a floating [−ATR]PL feature, rather than a suffixal -s that deletes. In Murcian plural forms, laxing extends throughout the word. The low vowel undergoes laxing but high vowels never undergo laxing. Moreover, unlike Granada EAS, high vowels in Murcian are always opaque in that they block leftward spreading of the [−ATR] feature. An example showing leftward spreading to a low vowel is the word *abeto* 'fir', whose plural form in Murcian is [æ.ˈßɛ.tɔ] with all vowels laxed. Considering the hierarchy of constraints that we established for Granada EAS shown in (16), the change necessary to account for the leftward spreading of [−ATR] to low vowels in Murcian would be to demote the feature cooccurrence constraint *[−ATR +Low] below the AlignPL-L constraint so that low vowels would undergo leftward spreading. An example showing the blocking effect of high vowels in Murcian is the plural of *cómico* 'comic, masc.', whose plural form is [ˈko.mi.kɔ] with only the last vowel laxed. The blocking effect of high vowels can be accounted for by adjusting the constraint ranking shown in (16) to place NoGap higher than AlignPL-L. A plural form like *[ˈkɔ.mi.kɔ], which occurs in other varieties of EAS, would be ruled out in Murcian by the NoGap constraint. This shows how our alignment conception of the EAS plural without suffixal /s/ can be extended to patterns of plural laxing reported in other varieties described in the EAS literature.

## 6. Conclusions

In this paper, we have developed a feature alignment approach to the problem of vowel laxing in plural word forms in Eastern Andalusian Spanish with an analysis in

Optimality Theory. Our analysis differs in conception from previous analyses, in that we do not assume overtly or covertly that plural forms have a lexical suffixal /s/ in their underlying representation. Rather, we conceptualize the plural as being marked by a floating [–ATR]PL feature that is realized through morphological alignment constraints that interact with other constraints to account for the various patterns of plural laxing found in EAS. Previous accounts maintain a suffixal plural /s/ because of the similarity found in EAS between plural laxing and laxing due to lexical /s/ deletion in codas.

Diachronically, the EAS plural almost certainly has its origins in lexical coda deletion of /s/ whereby the deleted /s/ leaves behind a [spread glottis] feature that coalesces with the immediately preceding tense [+ATR] vowel, resulting in a lax [–ATR] vowel. This feature then spreads regressively as laxing harmony for reasons of feature perceptibility. While this may be the historical source of plural laxing, we maintain in this paper that this phonological process has morphologized in current EAS so that [–ATR] (as a floating feature) marks the plural. We have argued for this based on synchronic evidence that plural laxing is different from laxing due to lexical /s/ deletion in coda position. As we have detailed, evidence for our view is supported by the phonetic study of Herrero de Haro and Hajek (2022, p. 141), who "found little evidence of so-called aspiration as a pronunciation of coda /s/ in [their] corpus", and by Herrero de Haro (2020, p. 19), who concludes that the EAS plural is coming to be marked by "an unidentified suprasegmental element". This view is further supported by Henriksen (2017, pp. 110–11), who finds a difference between the phonetics of morphological laxing in the plural and laxing triggered by lexical coda /s/ deletion. Importantly, this view is also supported by the observation that there is a difference in laxing in oxytones (i.e., final stress) between plural forms and singular forms with lexical /s/ deletion. This can be seen in the contrast between the singular *revés* 'other side' pronounced as [rɛ.ˈvɛ] and the plural of *bebé* 'baby' pronounced as [be.ˈbɛ] where [–ATR] is only realized on the final vowel in plural of oxytones ending in a stressed vowel. Finally, this view is supported by the Jaen variety of EAS, where plural laxing is clearly distinct in its phonological pattern from coda /s/ deletion. This suggests that EAS laxing exemplifies the notion of rule scattering (e.g., Bermúdez-Otero 2016) where a single diachronic process can evolve synchronically into two similar but separate rules, especially if an environment can be interpreted morphologically.

All these reasons justify an analysis of the EAS plural as entailing a floating [–ATR] feature, represented as [–ATR]PL. Finally, we note that the optimality-theoretic analysis that we have developed in Sections 3 and 4 is consistent with Finley's (2009, p. 478) description of morphemic harmony as "the result of a feature or subset of features functioning as a morpheme on their own" that "follows from faithfulness constraints to the harmonic feature associated with the harmony-inducing morpheme". While Finley uses Anchor constraints rather than Alignment constraints in her work, she notes that Alignment and Anchor have very similar outcomes. Diachronically, then, the change of plural laxing from phonological to morphological would constitute an instance of a rule life cycle (Janda 1987; Bermúdez-Otero 2016) where a phonological rule becomes morphologized.

There are still many remaining problems concerning plural laxing in Eastern Andalusian Spanish. We mention two. One problem is to determine the exact acoustic criteria for low and high vowels to distinguish between lax and tense productions. There is not yet an established acoustic threshold for the degree of F1 raising (and/or F2 change) necessary to determine if a vowel should be considered lax. While Herrero de Haro (2017a) has identified general differences in perceptibility between vowels in different contexts before a diachronically-elided consonant, the exact boundaries of perceptibility merit further investigation. Work regarding the English diphthong /ay/ being raised to [ʌy] before voiceless consonants, for example, has identified an established threshold (Labov et al. 2006) of a 60 Hz difference in F1 lowering for /ay/ to be considered perceptually raised. Speakers showing a difference below this level are not considered to have raising. While this intuition was probably behind Zubizarreta's distinction between low-level and high-level laxing in EAS, no such acoustic threshold has been agreed upon in the literature.

A second problem addressed in Section 4 is the laxing pattern found in the plural of words with the uncommon pattern of having a word-final stressed vowel. Given our native speaker consultation regarding the plural of words like *menú* and *bebé* shown in (17), we concur with Herrero de Haro (2020, p. 4), when he observes that "words that have stress on the last syllable only lower one vowel when /-s/ is deleted in EAS", based on the example of *papás* 'fathers' being produced as [pa.ˈpæ]. However, this claim is in need of further testing and verification. It would be interesting to know if this is the case in all varieties of plural laxing in EAS, including the Jaén variety that has been described as having laxing throughout the word in plural forms. We suspect systematic investigation of such oxytones would further support the morphological alignment analysis of EAS plural laxing put forward in this paper.

**Author Contributions:** Conceptualization, S.D.; methodology, M.P. and S.D.; formal analysis, S.D. and M.P.; investigation, M.P. and S.D.; resources, S.D. and M.P.; data curation, M.P.; writing— original draft preparation, M.P.; writing—review and editing, S.D. and M.P.; visualization, M.P. and S.D.; supervision, S.D.; project administration, S.D. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** This study did not require ethical approval.

**Data Availability Statement:** Data available on request from the authors.

**Acknowledgments:** This paper has benefited from discussion with Ricardo Bermúdez-Otero, Nick Henriksen, Alfredo Herrero de Haro, Maria-Rosa Lloret, and Estefania Valenzuela-Mochón. The analysis in this paper was developed by the authors through various presentations at Indiana University. We thank our colleagues and our students for their feedback, and we especially thank the anonymous reviewers. The usual disclaimers apply.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Notes

1 Previous research has suggested that there is minimal presence of aspiration in Granada Spanish. For example, Herrero de Haro and Hajek (2022, p. 141) state that they "found little evidence of so-called aspiration as a pronunciation of coda /s/ in [their] corpus". As a result, we do not consider aspiration in the subvariety of Granada Spanish described in this paper.

2 Following Henriksen (2017), we treat [æ] as a [–ATR] vowel alternating with the [+ATR] /a/, due to the acoustic analogy with /e/ and /ɛ/. These three vowels (i.e., /æ/, /ɛ/, /ɔ/) share a raised F1 in comparison to their tense counterparts, which we view as characteristic of a [±ATR] contrast. We note that the low vowel also fronts when it laxes, but we do not know to what extent fronting is salient for the perception of the plural. In languages with one low vowel phoneme, such as Arabic, there is much front-back variation in its realization.

3 While the data on Granada EAS presented in this section have emphasized the word as the domain of plural [–ATR] realization, we note examples with final high stressed vowels such as *menús* [me.ˈnʊ] 'menus' in (4b) where only the final syllable is [–ATR]. Similarly, Jiménez and Lloret (2020, p. 106) maintain that a plural word form like *monólogos* [mɔ.ˈnɔ.lɔ.ɣɔ] 'monologues' in (1c) could optionally be pronounced as [mo.ˈno.lɔ.ɣɔ] where the spreading of [–ATR] does not extend to pretonic position (before the stressed syllable). This latter pronunciation shows that there can be a foot domain to the realization of the plural [–ATR] feature. We suggest that this is more likely in slower speech, but nonetheless the foot as a domain for plural [–ATR] realization seems to be what is occurring with final stressed high vowels as in *menús* [me.ˈnʊ] 'menus'. We discuss this in detail in Section 4. In contrast, in the Jaén dialect of EAS, as described by Jiménez and Lloret, the domain of plural [–ATR] is the entire word, with all vowels becoming [–ATR] regardless of their underlying quality.

4 As a reviewer points out, the constraint *[–ATR, +Low] is somewhat odd, since there tends to be an incompatibility with [+Low] and [+ATR]. However, it is known that languages may differ as to whether they treat certain back and low vowels as [–ATR] or [+ATR] when there is only a single low vowel phoneme. As an example, Hantgan and Davis (2012) show that, in the [ATR] harmony system of the Dogon (Niger-Congo) language Bondu-so spoken in Mali, the single low vowel phoneme /a/ sometimes triggers [–ATR] harmony and sometimes [+ATR] harmony. For Spanish, one could maintain that all five vowel phonemes are by default [+ATR], given that it is not underlyingly contrastive. Thus, with respect to Spanish, the constraint in (11) is plausible, especially given that one of the acoustic correlates found by Henriksen for low vowel laxing is the lowering of the vowel (i.e., raising of F1). We also note that Kaplan (2021a, p. 706) considers /a/ to be [+ATR].

5 While we do not offer a tableau of the words *impetus* 'violence (pl.)' and *espíritu* 'spirits' from (4a), an issue arises in their analysis as to whether the non-final mid vowel is lax or not when the constraint RM-PL is incorporated. Based on the analysis of Jiménez and Lloret (2020), its laxing is optional. Our constraint ranking in Tableau 6 predicts that it should be lax so as to satisfy RM-PL. However, Jiménez and Lloret consider the final high vowel of the plural words in (4a) to undergo laxing (contrary to Henrikesen and Zubizarreta). If the final high vowels of words in (4a) do undergo plural laxing, then the mid vowels in words like *impetus* 'violence (pl.)' and *espíritu* 'spirits' need not be lax, since RM-PL would be satisfied by the laxing of the final vowel (though the first vowel in *espíritu* may be lax because of coda /s/ deletion). In general, the plural of words ending in high vowels in EAS is in need of more detailed empirical study.

6 The CrispEdge constraint in (19) that is relativized to a monomoraic foot prevents spreading of [–ATR]PL to the left of the stressed syllable. We note that a similar crisp edge constraint that is relativized to a foot more generally (i.e., CrispEdge ([–ATR]PL, Ft, L) can be used to account for the Granada variety in Jiménez and Lloret's (2020) analysis, where spreading of [–ATR] to pretonic syllables is optional. In their example of *económicos* 'economic, masc. pl.', the first two vowels are optionally laxed and so can be pronounced as [e.ko.ˈnɔ.mi.kɔ]. The variable ranking of this CrispEdge constraint with respect to the AlignPL-L constraint can account for the variable pronunciation between [e.ko.ˈnɔ.mi.kɔ] and [ɛ.kɔ.ˈnɔ.mi.kɔ], where the former has CrispEdge higher-ranked and the latter has AlignPL-L higher ranked. (See Kaplan 2021a for the use of the CrispEdge constraint in an analysis of EAS using Noisy Harmonic Grammar that has a different conception of the EAS laxing problem).

## References

Akinlabi, Akinbiyi. 1996. Featural affixation. *Journal of Linguistics* 32: 239–89. [CrossRef]

Akinlabi, Akinbiyi. 2011. Featural Affixes. In *The Blackwell Companion to Phonology*. Edited by Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume and Keren Rice. Chichester: Wiley-Blackwell, pp. 1945–71.

Alonso, Dámaso, Alonso Zamora Vicente, and María Josefa Canellada de Zamora. 1950. Vocales Andaluzas. *Nueva revista de Filología Hispánica* 3: 209–30. [CrossRef]

Archangeli, Diana, and Douglas Pulleyblank. 1994. *Grounded Phonology*. Cambridge, MA: MIT Press.

Bermúdez-Otero, Ricardo. 2007. Spanish pseudopluaral: Phonological cues in the acquisition of a syntax-morphology mismatch. In *Deponency and Morphological Mismatches*. Edited by Matthew Baerman, Greville C. Corbett, Dunstan Brown and Andrew Hippisley. Oxford: Oxford University Press, pp. 231–69.

Bermúdez-Otero, Ricardo. 2016. Amphichronic explanation and the life cycle of phonological processes. In *The Oxford Handbook of Historical Phonology*. Edited by Patrick Honeybone and Joseph Salmons. Oxford: Oxford University Press, pp. 374–99.

Booij, Geert. 2010. *Construction Morphology*. Oxford: Oxford University Press.

Davis, Stuart. 1995. Emphasis Spread in Arabic and Grounded Phonology. *Linguistic Inquiry* 26: 465–498.

Davis, Stuart, and Mihui Cho. 2003. The distribution of aspirated stops and /h/ in American English and Korean: An alignment approach with typological implications. *Linguistics* 41: 607–652. [CrossRef]

Finley, Sara. 2009. Morphemic harmony as featural correspondence. *Lingua* 119: 478–501. [CrossRef]

Goad, Heather. 1991. [Atr] and [Rtr] are different features. In *Proceedings of the Tenth West Coast Conference on Formal Linguistics*. Edited by Dawn Bates. Stanford: CSLI Publications, pp. 163–73.

Hantgan, Abbie, and Stuart Davis. 2012. Bondu vowel harmony: A descriptive analysis with theoretical implications. *Studies in African Linguistics* 41: 187–212. [CrossRef]

Henriksen, Nicholas. 2017. Patterns of vowel laxing and harmony in Iberian Spanish: Data from production and perception. *Journal of Phonetics* 63: 106–26. [CrossRef]

Herrero de Haro, Alfredo. 2017a. Four mid back vowels in Eastern Andalusian Spanish. *Zeitschrift fur romanische Philologie* 133: 82–114. [CrossRef]

Herrero de Haro, Alfredo. 2017b. The phonetics and phonology of eastern Andalusian Spanish: A review of literature from 1881 to 2016. *Ikalá, Revista de Lenguaje y Cultura* 22: 313–57. [CrossRef]

Herrero de Haro, Alfredo. 2020. Morpheme dislocation in Eastern Andalusian Spanish plurals. *Lingua* 238: 1–23. [CrossRef]

Herrero de Haro, Alfredo, and John Hajek. 2022. Eastern Andalusian Spanish. *Journal of the International Phonetics Association* 52: 135–55. [CrossRef]

Janda, Richard D. 1987. On the Motivation for an Evolutionary Typology of Sound-Structural Rules. Doctoral dissertation, University of California, Los Angeles, CA, USA. Available online: https://linguistics.ucla.edu/general/dissertations/Janda.1987.pdf (accessed on 25 April 2024).

Jiménez, Jesús, and María-Rosa Lloret. 2007. Andalusian vowel harmony: Weak triggers and perceptibility. Presented at the 4th Old World Conference in Phonology, Workshop on Harmony in the Languages of the Mediterranean, Rhodes, Greece, January 18–21. Available online: http://roa.rutgers.edu/files/901-0207/901-JIMENEZ-0-0.PDF (accessed on 3 October 2023).

Jiménez, Jesús, and María-Rosa Lloret. 2020. Vowel harmony. In *The Routledge Handbook of Spanish Phonology*. Edited by Sonia Colina and Fernando Martínez-Gil. New York: Routledge, pp. 100–28.

Kager, Rene. 1999. *Optimality Theory*. Cambridge: Cambridge University Press.

Kaplan, Aaron F. 2018. Asymmetric crisp edge. In *Hana-Bana: A Festschrift for Junko Ito and Armin Mester*. Edited by Ryan Bennett, Andrew Angeles, Adrian Brasoveanu, Dhyana Buckley, Nick Kalivoda, Shigeto Kawahara, Grant McGuire and Jaye Padgett. Santa Cruz: Department of Linguistics, University of California. Available online: https://itomestercelebration.sites.ucsc.edu/ (accessed on 3 October 2023).

Kaplan, Aaron F. 2021a. Categorical and gradient ungrammaticality in optional processes. *Language* 97: 703–31. [CrossRef]

Kaplan, Aaron F. 2021b. Persistence and opacity in Eastern Andalusian harmony. In *Proceedings of the Annual Meeting on Phonology (AMP) 2020*. Available online: https://journals.linguisticsociety.org/proceedings/index.php/amphonology/article/view/4899 (accessed on 3 October 2023).

Kurisu, Kazutaka. 2001. The Phonology of Morpheme Realization. Ph.D. dissertation, University of California, Santa Cruz, CA, USA.

Labov, William, Sharon Ash, and Charles Boberg. 2006. *The Atlas of North American English: Phonetics, Phonology and Sound Change*. Berlin: Walter de Gruyter.

Martínez-Paricio, Violeta. 2021. Stress in morphologically simple and complex Spanish words. In *The Routledge Handbook of Spanish Phonology*. Edited by Antonio Fábregas, Víctor Acedo-Matellán, Grant Armstrong, María Cristina Cuervo and Isabel Pujol Payet. New York: Routledge, pp. 375–86.

McCarthy, John J. 1997. Process-specific constraints in Optimality Theory. *Linguistic Inquiry* 28: 231–51.

Navarro Tomás, Tomás. 1938. Dédoublement de phonemes dans le dialecte andalou. *Travaux du Cercle Linguistique de Prague* 8: 184–86.

Navarro Tomás, Tomás. 1939. Desdoblamiento de fonemas vocálicos (Reduplication of vowel phonemes). *Revista de filologia hispánica* 1: 15–167.

Piñeros, Carlos. 2016. The phonological weight of Spanish syllables. In *The Syllable and Stress: Studies in Honor of James W. Harris*. Edited by Rafael Nuñez-Cedeño. Berlin: De Gruyter, pp. 271–314.

Pollock, Matthew. 2023. Performing Andalusian in Political Speech: Political Party And Sociophonetic Patterns Across Production And Perception. Doctoral dissertation, Indiana University, Bloomington, IN, USA.

Ranson, Diana L. 1992. Nominal number marking in Andalusian Spanish in the wake of /s/ deletion. *Hispanic Linguistics* 4: 301–27.

Soriano, Bárbara. 2012. Andalusian vowel harmony and morphology-phonology interface. *International Journal of Basque Linguistics and Philology* 46: 295–307.

Walker, Rachel. 2011. *Vowel Patterns in Language*. New York: Cambridge University Press.

Walker, Rachel. 2024. Stressed syllables are not singled out for transparency: A gestural account. Invited talk presented at the 21st Old World Conference on Phonology (OCP 21), Leipzig, Germany, February 13–16. Available online: https://home.uni-leipzig.de/ocp21/OCP21_Walker.pdf (accessed on 22 February 2024).

Zubizarreta, María L. 1979. Vowel harmony in Andalusian Spanish. *MIT Working Papers in Linguistics* 1: 1–11.

*languages*

MDPI

*Article*

# Dissimilation in Hispano-Romance Diminutive Suffixation

**Claire Julia Lozano [1] and Travis G. Bradley [2,*]**

[1] Department of Spanish and Portuguese, Macalester College, St. Paul, MN 55105, USA
[2] Department of Spanish and Portuguese, University of California, Davis, CA 95616, USA
* Correspondence: tgbradley@ucdavis.edu

**Abstract:** A highly productive derivational process, diminutive suffixation in Spanish (e.g., *gatito* ~ *gatiko/gatico* 'little/well-known/beloved/awful cat' < *gato* 'cat') has received much attention in the morphology–phonology interface literature. The present study contributes a novel comparative analysis of a dissimilatory alternation between diminutive suffix allomorphs *-ito/a* and *-ico/a (-iko/a)* across three Hispano-Romance varieties. In Judeo-Spanish, the voiceless dorsal stop [k] of default *-iko/a* dissimilates to coronal [t] after any dorsal segment [k, ɡ, ɡʷ, x, w] in the base-final syllable. In Colombian Spanish, the voiceless coronal stop [t] of default *-ito/a* dissimilates to dorsal [k] after only an identical [t] in the base-final syllable. By contrast, Castilian Spanish *-ito/a* does not dissimilate, thereby providing a baseline for comparison. All three varieties allow for optional iteration of the suffix, which conveys greater smallness or endearment than the simple diminutive, e.g., Castilian Spanish *gatitito* 'little/beloved kitty', without dissimilation. Iterated diminutives in Colombian Spanish show two patterns of dissimilation, which have not been fully acknowledged in the previous literature. For example, either (i) [it] and [ik] alternate to avoid adjacent identical syllable onsets, e.g., *gat*[ikitíko], or (ii) [it] is iterated until alternating with word-final [ik], e.g., *gat*[ititíko]. In all three Hispano-Romance varieties, base-final unstressed vowels are deleted before a vowel-initial diminutive suffix, followed by unstressed *-o/a*, and stress (indicated by an acute accent) is shifted rightward onto the penultimate syllable of the diminutive word. Vowel deletion and stress shift apply recursively in iterated diminutives. We propose an Optimality Theory analysis of these alternations in terms of suffix allomorphy that is phonologically conditioned by consonantal place dissimilation. The analysis is formalized as an interaction among constraints that enforce prosodic unmarkedness, output–output correspondence, allomorph preference, and similarity avoidance. We consider theoretical alternatives and compare our analysis to other recent proposals.

**Keywords:** dissimilation; diminutive suffixation; Judeo-Spanish; Colombian Spanish; Castilian Spanish; optimality theory; constraint conjunction; obligatory contour principle

## 1. Introduction

Diminutive suffixation is a well-known process of Hispano-Romance derivational morphology that can semantically express smallness, familiarity, affection, or even disdain. Several different diminutive suffixes are attested in contemporary Spanish varieties, e.g., *-ito/a*, *-(e)cito/a*, *-illo/a*, *-(e)cillo/a*, *-(z)uelo/a*, *-(e)cico/a*, *-uco/a*, and *-ín/ina*. The most common is *-ito/a*, which also has a longer, morpho-phonologically conditioned allomorph *-(e)cito/a.* In Judeo-Spanish, the preferred diminutive suffix *-iko/a* alternates with both *-ito/a* and a longer allomorph *-eziko/a.*[1] Diminutive suffixation has been described and analyzed in contemporary Hispano-Romance linguistics (Jaeggli 1980; Nieuwenhuis 1985; P. Prieto 1992; Crowhurst 1992; Harris 1994; Ambadiang 1996; Ohannesian 1996, 2020; Ohannesian and Pons 2009; Lázaro Mora 1999; Miranda 1999; Eguren 2001; Bunis 2003; Colina 2003, 2009; V. M. Prieto 2005; Bermúdez-Otero 2006, 2013; Zacarías 2006; Bradley and Smith 2011; Smith 2011; Ambadiang and Bergareche 2012; Normann-Vigil 2012; Fábregas 2013; Vadella 2017; Camus Bergareche 2018; Tarazona 2021). In simple diminutives, the suffix [it]/[ik]

attaches to a base word, e.g., *gatito* ~ *gatiko/gatico* 'little/well-known/beloved/awful cat' < *gato* 'cat'. The highly productive nature of diminutivization can even give rise to what Nieuwenhuis (1985, pp. 73–74) calls an 'intensified diminutive,' which conveys even greater smallness or endearment and is generally expressed in Romance by the iteration, or repetition, of the suffix, e.g., *gatitito* ~ *gatikito/gatiquito* ~ *gatitico* 'little/beloved kitty' < *gatito* ~ *gatiko/gatico* 'kitty' < *gato* 'cat'. We henceforth refer to such forms as *iterated diminutives.*

The present article is about a dissimilatory alternation between diminutive suffix allomorphs [ik] and [it], as observed in Judeo-Spanish (JS) and Colombian Spanish (CoS) and compared to non-alternating [it] in Castilian Spanish (CaS). Table 1 illustrates the main patterns.[2] We will argue that dissimilation in shorter allomorphs avoids the repetition of sufficiently similar onset-initial segments across two adjacent syllables. Each Hispano-Romance variety also includes a longer allomorph, as attested in examples such as JS *paneziko* and CoS/CaS *panecito*, both derived from *pan* 'bread', as well as CoS/CaS *pintorcito* < *pintor* 'painter'. The longer allomorphs do not participate in the dissimilation alternation. In CoS, *gatico* is the preferred diminutive of *gato* 'cat', even though hypothetical **gatecito* would have equally avoided the repetition of onset-initial [t] in **gatito.* However, the longer allomorphs are important to mention here because they provide a clue as to which of the two shorter allomorphs is the default suffix, which appears to the left in each row of Table 1. For example, default *-iko/a* matches *-eziko/a* in JS, and default *-ito/a* matches *-(e)cito/a* in CoS (and in CaS, which altogether lacks an [ik] allomorph).[3]

**Table 1.** Diminutive suffix allomorphy across three Hispano-Romance varieties. Orthographic forms include an unstressed terminal element *-o* or *-a.* Phonetic forms of the diminutive allomorph appear beneath.

| Variety | Shorter Allomorphs | | | | Longer Allmorphs |
|---|---|---|---|---|---|
| Judeo-Spanish (JS) | *-iko/a* | ~ | *-ito/a* | ~ | *-eziko/a* |
| | [ik] | | [it] | | [ezik] |
| Colombian Spanish (CoS) | *-ito/a* | ~ | *-ico/a* | ~ | *-(e)cito/a* |
| | [it] | | [ik] | | [(e)sit] |
| Castilian Spanish (CaS) | *-ito/a* | | ~ | | *-(e)cito/a* |
| | [it] | | | | [(e)θit] |

This article has two main goals, one empirical and one theoretical. The empirical goal is to give a thorough description of the [ik]~[it] alternation in JS and CoS diminutives. We identify novel generalizations about the patterning of iterated diminutives in CoS. We show that diminutives in JS, CoS, and CaS also involve processes of base-final unstressed vowel deletion and rightward stress shift, which apply recursively in iterated diminutives. Our theoretical goal is to develop a comprehensive phonological analysis of simple and iterated diminutivization within the constraint-based framework of Optimality Theory.

This article is structured as follows. In Section 2, we present data from JS, CoS, and CaS to motivate generalizations about the morpho-phonological domains of alternation between the diminutive suffix allomorphs [ik] and [it]. In Section 3, we propose a formal model of simple and iterated diminutives in classic, monostratal Optimality Theory (OT; Prince and Smolensky 1993/2004; McCarthy and Prince 1999), which assumes that inputs are mapped to optimal outputs by a phonological grammar consisting of a single constraint ranking. In Section 4, we analyze the [ik] ~ [it] alternation in terms of suffix allomorphy that is phonologically conditioned by consonantal place dissimilation, formalized as an interaction among constraints that enforce prosodic unmarkedness, output–output correspondence, allomorph preference, and similarity avoidance. Section 5 compares our analysis with theoretical alternatives. Section 6 summarizes and concludes.

## 2. Dissimilation in Hispano-Romance Diminutive Suffixation

### *2.1. JS*

Bunis (2003) provides, to our knowledge, the most comprehensive description of JS diminutives to date. Such forms were amply attested in texts since the early 18th century and involve "lexemes derived from all source components, including Ottoman Turkish and other local languages, Hebrew and Aramaic, and, from the second half of the 19th century, Western European prestige languages such as Italian, French and German" (p. 205). In earlier Ottoman Jewish-letter texts from the 16th century, "the use of *-ito* is always governed by phonological constraints: it is *only* attracted to stems exhibiting the velars *k* or *g*, especially word-medially, thus appearing to effect a kind of dissimilation" (p. 198).

In modern JS simple diminutives, [ik] appears after syllable-initial labials [p, b, v, m] (1a), coronals [t, d, z, n, ɾ, r, l] (1b), and (pre)palatals [ʧ, ʃ, ʒ, j] (1c). To simplify the presentation, we give only a few representative examples for each phonological context, which is defined here in terms of the base-final syllable. A period denotes a syllable boundary, braces enclose possible syllable-initial segments, and an underscore shows where the diminutive suffix attaches, assuming deletion of the base-final unstressed vowel, or terminal element. Page numbers indicate the location of each example cited in Bunis's (2003) study, including the diminutive, base, and English gloss. See Appendix A for an exhaustive list of JS simple diminutives extracted from Bunis (2003).

| (1) | | Context | Diminutive | Base | | |
|---|---|---|---|---|---|---|
| | a. | .{p, b, v, m}_ | *kapika* | *kapa* | 'cap' | 213 |
| | | | *gulubika* | *guluba* | 'pigeon' | 214 |
| | | | *manseviko* | *mansevo* | 'young man' | 214 |
| | | | *ramika* | *rama* | 'branch' | 213 |
| | b. | .{t, d, z, n, ɾ, r, l}_ | *kartika* | *karta* | 'card' | 213 |
| | | | *adjuntadiku* | *adjuntadu* | 'united' | 214 |
| | | | *ermoziko* | *ermozo* | 'beautiful' | 214 |
| | | | *kadenika* | *kadena* | 'chain' | 214 |
| | | | *orika* | *ora* | 'hour' | 213 |
| | | | *karriko* | *karro* | 'wagon' | 213 |
| | | | *Solika* | *Sol* | (name) | 207 |
| | c. | .{ʧ, ʃ, ʒ, j}_ | *biskochiko* | *biskocho* | 'biscuit' | 214 |
| | | | *kashika* | *kasha* | 'box, drawer' | 213 |
| | | | *navajika* | *navaja* | 'razor' | 214 |
| | | | *bo(y)iko* | *boyo* | 'bun' | 213 |

On the other hand, in modern JS simple diminutives, [it] appears after any dorsal, not only [k] (2a) and [ɡ, $ɡ^w$] (2b) but also [x] (2c) and [w] (2d):

| (2) | | Context | Diminutive | Base | | |
|---|---|---|---|---|---|---|
| | a. | .k_ | *freskito* | *fresko* | 'fresh' | 211 |
| | b. | .{ɡ, $ɡ^w$}_ | *bragita* | *braga* | 'trousers' | 210 |
| | | | *agwita* | *agwa* | 'water' | 211 |
| | c. | .x_ | *blahitu* | *blahu* | 'Christian' | 211 |
| | d. | .w_ | *Elyawito* | *Elyaw* | (name) | 211 |

Besides being reflected in the longer allomorph *-eziko/a* in Table 1, [ik] also has a wider surface distribution in simple diminutives (1) and can, therefore, be considered the default allomorph of the shorter suffix in JS. By contrast, the distribution of [it] is restricted to the specific context of a preceding dorsal onset (2).

Regarding iterated diminutives, Bunis (2003) writes that in

> "endings such as *-ikito (< -iko + -ito)* and *-etiko/-itiko (< -ito + -iko)*, the first element probably constituted the earlier suffix, originally employed with the stem in a preceding stage of the language; the second suffix was perhaps added after the earlier diminutive had lost its diminutive force, or had acquired a specialized sense, no longer functioning as a mere diminutive of the stem, or to express a higher degree of diminution" (p. 228)

Examples of JS iterated evaluative (e.g., diminutive or augmentative) suffixes are given in (3), along with the number of such suffixes that appear in the longest form. The examples in (3a) show that adjacent identical syllable onsets are avoided, giving rise to recursive dissimilation between diminutive suffix allomorphs [it] and [ik]. In *saltikón* (3b), *-iko* combines with *salto*, to which the second, augmentative suffix *-ón* then attaches. Although not technically an iterated diminutive, *saltikón* '(big little) jump' shows that evaluative suffixes are productively iterated in JS:

| (3) | | # of suffixes | Diminutives | | |
|---|---|---|---|---|---|
| | a. | 2 | *chikitiko* < *chikito* < *chiko* | 'small' | 228 |
| | | | *pokitiko* < *pokito* < *poko* | 'tiny bit' | 228 |
| | b. | 2 | *saltikón* < *saltiko* < *salto* | 'jump' | 229 |

The pair in (3a) are the only concrete examples of iterated diminutives cited in Bunis's (2003) description of modern Jewish-letter texts. In general, iterated diminutives in Hispano-Romance seem to be much more common in the spoken language, where they serve many pragmatic purposes across a range of different interactional contexts, all of which deserve further detailed study. Corpora that are limited to more formal, written registers of JS may fail to capture the true extent of iterated diminutives in actual speech. Their high frequency in the spoken language of contemporary non-Sephardic Spanish suggests that iterated diminutives were likely just as productive in spoken JS. For discussion on this point, we give thanks to Aldina Quintana (personal communication).

### *2.2. CoS*

According to Lipski (1999, p. 19), the Spanish diminutive suffix *-ico/a* is attested in Colombia, Costa Rica, Venezuela, Cuba, and regions of eastern and southern Spain (see also Kany 1960, p. 156), and its use presents a striking parallel with the JS diminutive suffix *-iko/a.* Historically, Peninsular Spanish *-ico/a* derives from Proto-Romance *-iccu/a.* Although its origins prior to Vulgar Latin are not known for certain, this suffix is associated primarily, although not exclusively, with the regions of Aragon and Murcia in eastern Spain (Alvar and Pottier 1983, pp. 367–8; Nieuwenhuis 1985, p. 186–7). According to previous theories, its possible origins may be Iberian, Celtic, Basque, African, Roman, or German (González Ollé 1962, pp. 319–26). Nebrija's (1492) *Gramática Castellana* gives evidence that *-ico/a* was one of the three most popular diminutive suffixes in the 15th century, along with *-illo/a* and *-ito/a.* In the Golden Age literature, *-ico/a* enjoyed widespread use and prestige. Even though it is no longer used in normative CaS, *-ico/a* is still attested in present-day Navarra, Aragon, and Murcia, as well as in eastern regions of Andalusia in southern Spain (Penny 2014, p. 319).

The presence of *-ico/a* in 15th-century Spain aligns historically with the expulsion of the Sephardic Jews from the Iberian Peninsula starting in 1492. As with other regional linguistic features, such as syllable-final /s/ aspiration, *yeísmo*, *seseo*, etc., it is possible to trace the existence of *-ico/a* diminutives in the American colonies back to settlers from the eastern and southern Iberian Peninsula, many of whom were Sephardic Jews. Lipski (1999) draws an explicit historical connection between Latin American Spanish *-ico/a* and JS *-iko/a*, arguing that the presence of the same suffix in Colombia and Cuba is unlikely to be a "casual coincidence" (p. 31). He proposes that the contributions from JS and eastern and southern Peninsular Spanish represent one of the most notable "missing links" (p. 32) in the evolution of Latin American Spanish varieties. In particular, Colombia is the "apparent epicenter" (p. 19) of South American *-ico/a*, and in Costa Rica, the extensive use of *-ico/a* in iterated diminutives like *hermanitico* < *hermano* 'brother' has given rise to the hypocoristic "Ticos" that is commonly used to refer to Costa Ricans. Lipski claims that in the Peninsular Spanish varieties that use *-ico/a*, this suffix can be added to any nominal root, regardless of the configuration of final consonants and vowels, e.g., *angelico* < *ángel* 'angel', *casica* < *casa* 'house', *gordico* < *gordo* 'fat'. However, the Latin American varieties have a dissimilatory phonotactic restriction whereby *-ico/a* can attach only to base words whose final syllable begins with a voiceless coronal stop [t], e.g., *momentico* < *momento* 'moment', *ratico* < *rato* 'a while', *maestrico/a* < *maestro/a* 'teacher'.

Evidence for such a restriction comes from Fontanella's (1962) foundational descriptive study of CoS diminutives, drawn from a corpus of over 200 such forms that she observed during informal conversations over a two-month period around the city of Bogotá, Colombia. In CoS simple diminutives as used in the spoken language, [it] appears after labials [p, b, m] (4a), coronals [d, s, n, ɾ, l] (4b), palatals [ɲ, ʝ] (4c), and dorsals [k, ɡ, x] (4d). We give representative examples below, but see (39) and (40) of Appendix B for an exhaustive list of simple diminutives extracted from Fontanella (1962).

| (4) | | Context | Diminutive | Base | | |
|---|---|---|---|---|---|---|
| | a. | .{p, b, m}_ | *papito* | *papi* | 'daddy' | 560 |
| | | | *Albita* | *Alba* | (name) | 560 |
| | | | *climita* | *clima* | 'climate' | 567 |
| | b. | .{d, s, n, ɾ, l}_ | *caldito* | *caldo* | 'soup, broth' | 563 |
| | | | *bracito* | *brazo* | 'arm' | 563 |
| | | | *hermanito* | *hermano* | 'brother' | 563 |
| | | | *ahorita* | *ahora* | 'now' | 557 |
| | | | *arbolito* | *árbol* | 'tree' | 566 |
| | c. | .{ɲ, ʝ}_ | *niñito* | *niño* | 'child' | 557 |
| | | | *Estrellita* | *Estrella* | (name) | 560 |
| | d. | .{k, ɡ, x}_ | *cerquita* | *cerca* | 'near' | 557, 569 |
| | | | *alguito* | *algo* | 'something' | 568, 570 |
| | | | *cajita* | *caja* | 'box' | 570 |

However, [ik] appears after the voiceless coronal stop [t] (5):

| (5) | Context | Diminutive | Base | | |
|---|---|---|---|---|---|
| | .t_ | *abiertico* | *abierto* | 'open' | 566 |
| | | *galletica* | *galleta* | 'cookie' | 557, 566 |
| | | *tantica* | *tanta* | 'so much' | 566 |

Besides being reflected in the longer allomorph *-(e)cito/a* in Table 1, [it] also has a wider surface distribution in simple diminutives (4) and can, therefore, be considered the default allomorph of the shorter suffix in CoS. By contrast, the distribution of [ik] is restricted to the specific context of a preceding onset [t] (5).

However, there is some more recent empirical evidence that dissimilation in CoS simple diminutives is not an obligatory process but instead displays variation. Data from a search of the *Corpus del Español: Web/Dialects* (Davies 2016–) carried out in March 2024, limited to Colombia and based on the wildcard string **tit?*, are given in (6). Although these representative forms are not the only examples we found, they serve to show that [it] can appear after a voiceless coronal stop [t], where Fontanella's description of the Bogotá variety as spoken in 1962 would lead us to expect [ik]. Further research is necessary to determine the sources (diachronic, geographic, social, stylistic, etc.) and full extent of such variation in present-day CoS:

| (6) | Context | Diminutive | Base | |
|---|---|---|---|---|
| | .t_ | *calentito/a* | *caliente* | 'hot' |
| | | *gatito/a* | *gato/a* | 'cat' |
| | | *momentito* | *momento* | 'moment' |

CoS iterated diminutives show three distinct patterns. In the first pattern, [it] iterates but does not alternate with [ik], i.e., there is no dissimilation. A search of the *Corpus del Español*, limited to Colombia and based on the string **itit?*, returned examples such as those in (7), which mirror the simple diminutives presented in (6):

| (7) | | # of suffixes | Diminutive | Base | |
|---|---|---|---|---|---|
| | a. | 2 | *ahoritita* | *ahora* | 'now' |
| | | | *chiquitito/a* | *chico/a* | 'small' |
| | | | *poquitito* | *poco* | 'few' |
| | | | *puritito/a* | *puro/a* | 'pure' |
| | | | *queditito* | *quedo* | 'soft' |
| | | | *toditito/a* | *todo/a* | 'all' |
| | b. | 3 | *puntitita* | *punta* | 'tip' |

In the second pattern, [it] ~ [ik] alternate to avoid identical onsets in adjacent syllables, as the examples from Fontanella (1962) show in (8) (see (41) of Appendix B for an exhaustive list). This pattern is identical to the JS iterated diminutives in (3a):

| (8) | | # of suffixes | Diminutive | Base | | |
|---|---|---|---|---|---|---|
| | a. | 2 | *arribitica* | *arriba* | 'above' | 561 |
| | | | *poquitico* | *poco* | 'few' | 561 |
| | | | *toditica* | *toda* | 'all' | 568 |
| | b. | 3 | *cortiquitica* | *corta* | 'short' | 558, 561, 568 |
| | c. | 4 | *chiquitiquitico* | *chico* | 'small' | 558 |
| | | | *toditiquitica* | *toda* | 'all' | 568 |
| | d. | 5 | *cortiquitiquitico* | *corto* | 'short' | 558 |

In the third pattern, [it] is iterated until alternating with word-final [ik] (9) and (10a). However, [ik] never iterates without alternating (10b). A search of the *Corpus del Español* based on **ititic?* returned the CoS examples in (9), while those in (10) come from an article published in 2018 by Ana María Díaz Collazos in the *Gaceta Dominical* of the online newspaper *El País*, based in Cali, Colombia.[4]

| (9) | | # of suffixes | Diminutive | Base | |
|---|---|---|---|---|---|
| | a. | 3 | *chiquititica* | *chica* | 'small' |
| | | | *enterititica* | *entera* | 'entire' |
| | | | *livianititica* | *liviana* | 'light' |
| | | | *quebradititica* | *quebrada* | 'uneven, rough' |
| | | | *sequititica* | *seca* | 'dry' |
| | b. | 7 | *igualitititititica* | *igual* | 'equal' |
| | | | *poquitititititico* | *poco* | 'few' |

| (10) | | # of suffixes | Diminutive | Base | |
|---|---|---|---|---|---|
| | a. | 6 | *ahorititititica* | *ahora* | 'now' |
| | b. | 6 | **ahoritiquiquiquiquica* | | |

The difference between (8) and (9)–(10) has gone unnoticed in the previous literature. We provide the first theoretical account of these patterns in Section 4.

### 2.3. CaS

CaS *-ito/a* does not dissimilate, thereby providing a baseline against which to compare the dissimilatory patterns of JS and CoS. To facilitate the comparison, we follow the same presentational format as in previous sections. In CaS, the diminutive suffix [it] appears to the total exclusion of [ik], both in simple (11) and iterated (12) diminutives. These data come from a *Corpus del Español* search limited to Spain based on the strings **pit?*, **bit?*, **vit?*, **fit?*, **mit?* (11a), **tit?*, **dit?*, **cit?*, **sit?*, **nit?*, **rit?*, **rrit?*, **lit?* (11b), **chit?*, **ñit?*, **llit?*, **yit?* (11c), **quit?*, **guit?*, **güit?*, **git?*, **jit?* (11d) and for iterated diminutives, **itit?* (12a) and **ititit?* (12b,c).

| (11) | | Context | Diminutive | Base | |
|---|---|---|---|---|---|
| | a. | .{p, b, f, m}_ | *principito* | *príncipe* | 'prince' |
| | | | *bebito* | *bebé* | 'baby'[5] |
| | | | *vivito* | *vivo* | 'alive' |
| | | | *triunfito* | *triunfo* | 'triumph' |
| | | | *mismito/-a* | *mismo/-a* | 'same' |
| | b. | .{t, d, θ, s, n ,ɾ, r, l}_ | *ratito* | *rato* | 'a while' |
| | | | *cucharadita* | *cucharada* | 'tablespoon' |
| | | | *trocito* | *trozo* | 'slice' |
| | | | *besito* | *beso* | 'kiss' |
| | | | *granito* | *grano* | 'grain, seed' |
| | | | *señorita* | *señora* | 'lady' |
| | | | *perrito* | *perro* | 'dog' |
| | | | *arbolito* | *árbol* | 'tree' |
| | c. | .{ʧ, ɲ, ʎ, ʝ}_ | *bichito* | *bicho* | 'bug, creature' |
| | | | *pequeñito/-a* | *pequeño/-a* | 'small' |
| | | | *rollito* | *rollo* | 'roll, reel, a drag' |
| | | | *joyita* | *joya* | 'jewel' |
| | d. | .{k, ɡ, ɡ$^{w}$, x}_ | *poquito/-a* | *poco/-a* | 'few |
| | | | *barriguita* | *barriga* | 'belly' |
| | | | *agüita* | *agua* | 'water' |
| | | | *Jorgito* | *Jorge* | (name) |
| | | | *cajita* | *caja* | 'box' |

| (12) | | # of suffixes | Diminutive | Base | |
|---|---|---|---|---|---|
| | a. | 2 | *chiquitito/-a* | *chico/-a* | 'small' |
| | | | *poquitito/-a* | *poco/-a* | 'few' |
| | | | *puritito/-a* | *puro/-a* | 'pure' |
| | b. | 3 | *ahoritititita* | *ahora* | 'now' |
| | | | *perritititito* | *perro* | 'dog' |
| | | | *todititito/-a* | *todo/-a* | 'all' |
| | c. | 5 | *chiquitititititito/-a* | *chico/-a* | 'small' |

### *2.4. Generalizations to Be Accounted for*

The data presented in the previous sections motivate several morpho-phonological generalizations about diminutive suffixation in Hispano-Romance and, in particular, the dissimilatory alternation between allomorphs [ik] and [it] in JS and CoS. The generalizations in (13) directly inform the analysis to be developed in the rest of this article.

(13) a. JS and CoS differ in terms of whether [ik] or [it] is considered the unmarked (default) allomorph of the diminutive suffix (DIM), while [ik] is absent from CaS. (See Table 1).

b. The [k] of JS DIM [ik] dissimilates to [t] after any syllable-initial dorsal consonant (2) and (3a), whereas the [t] of CoS DIM [it] dissimilates to [k] after only an identical syllable-initial [t] (5) and (8)–(10).

c. DIM [it]/[ik] can be optionally iterated, showing three patterns:
  (i) only [it] iterates, as in CoS (6) and (7), similarly to CaS (12).
  (ii) [it]~[ik] alternate, as in JS (3a) and CoS (5) and (8).
  (iii) [it] iterates until alternating with word-final [ik], as in CoS (9) and (10a). However, [ik] never iterates without alternating (10b).

d. Unstressed terminal element (TE) suffix vowels of the base are deleted before vowel-initial DIM [it] and [ik], and the stress of the base is shifted rightward onto the penultimate syllable of the diminutive, which must end in either TE [o] or [a].

## 3. Setting the Stage: Base-Final Vowel Deletion and Rightward Stress Shift

The constraint-based framework of OT (Prince and Smolensky 1993/2004; McCarthy and Prince 1999) makes possible an explicit, formal account of the generalizations in (13). Before addressing allomorphy and dissimilation (13a–c), it will help to first explain base-final vowel deletion and rightward stress shift (13d), as both phenomena are common to all three Hispano-Romance varieties. For simplicity, we use the non-dissimilating, baseline variety CaS to model the two patterns. Following work on transderivational similarity (Benua 1997; Steriade 1999, 2000; Crosswhite 1998; Kager 1999; Ohannesian and Pons 2009; among others), we assume a model of diminutive formation that evaluates correspondence relations between the output forms of morphologically related words, in this case [base+DIM+TE] ↔ [diminutive]. As shown in Figure 1, the derivation of iterated diminutives is a recursive procedure of attaching DIM+TE to the latest base output. The procedure is terminated at the speaker's will. As the CaS simple (a) and iterated (b) diminutives in Table 2 make clear, base-final TE vowels delete before the vowel-initial DIM, and the stress foot (Σ) of the prosodic word (ω) of the base shifts rightward in the ω of the diminutive.

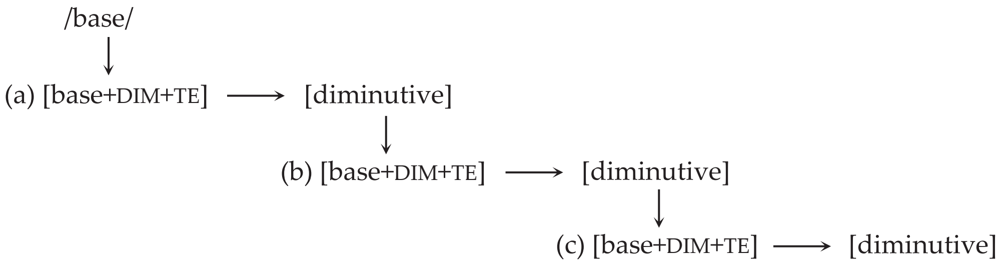



**Figure 1.** Output–output correspondence between [base+DIM+TE] and [diminutive] forms.

**Table 2.** CaS simple diminutive *besito* < *beso* 'kiss' (a) and iterated diminutive *ahoritita* < *ahorita* < *ahora* 'now' (b).

| | **[Base+DIM+TE]** | | **[Diminutive]** | |
|---|---|---|---|---|
| (a) | $((\text{bé.so})_{\Sigma})_{\omega}$+it+o | → | $(\text{be}(.\text{sí.to})_{\Sigma})_{\omega}$ | |
| (b) | $(\text{a}(\text{ó.ɾa})_{\Sigma})_{\omega}$+it+a | → | $(\text{a.o}(.\text{ɾí.ta})_{\Sigma})_{\omega}$ | → |
| | $(\text{a.o}(.\text{ɾí.ta})_{\Sigma})_{\omega}$+it+a | → | $(\text{a.o.ɾi}(.\text{tí.ta})_{\Sigma})_{\omega}$ | |

Attaching a vowel-initial DIM to a vowel-final base creates a hiatus, or sequence of two adjacent vowels that belong to separate syllables. A hiatus is a cross-linguistically marked structure that languages tend to avoid. A common repair strategy is to delete the first vowel. This strategy can be formalized as an interaction between a markedness constraint against a hiatus (14a) and an anti-deletion correspondence constraint (14b). (For longer constraint names, an abbreviated title appears after the hyphen, '—', here and below.) Furthermore, a positional anti-deletion correspondence constraint (14c) prohibits deletion of the second vowel in a hiatus created by DIM suffixation (see Casali 1997, 2011).

(14) a. *HIATUS
Assign a violation for every sequence of adjacent vowels that belong to separate syllables.
BASEDIMINUTIVE-MAXIMALITY—BD-MAX
b. Assign a violation for every segment in [base+DIM+TE] that has no correspondent in the diminutive. ("The diminutive contains a maximal expression of the segments of [base+DIM+TE].")
BASEDIMINUTIVE-MAXIMALITY(V2)—BD-MAX(V2)
c. Assign a violation for every postvocalic vowel in [base+DIM+TE] that has no correspondent in the diminutive. ("The diminutive contains a maximal expression of the postvocalic vowels of [base+DIM+TE]".)

We argue that *HIATUS (14a) crucially outranks BD-MAX (14b), while the relative ranking of BD-MAX(V2) (14c) is indeterminate. Tableau (15) gives the analysis of CaS *besito* < *beso* (Table 2a).[6] Corresponding segments between the base output and the diminutive candidates now include identical numerical subscripts. Hiatus in [base+DIM+TE], as reflected in [$o_4$.$í_5$] (15a), is repaired by deleting the base-final TE vowel [$o_4$] (15b). Because deleting the DIM-initial [$i_5$] will always violate both BD-MAX(V2) and BD-MAX, no matter how the two correspondence constraints are ranked, candidate (15c) is harmonically bounded by, and will always lose to, candidate (15b). Ramsammy (2017, p. 72) proposes a similar analysis of stem-final vowel deletion before the vowel-initial DIM /et/ in Italian *libretto* < *libro* 'book', although his account assumes a Stratal OT model that distinguishes among stem-, word-, and phrase-level constraint rankings, which we do not adopt here.

| (15) | | $((b_1é_2.s_3o_4)_Σ)_ω+i_5t_6+o_7$ | *HIATUS | BD-MAX | BD-MAX(V2) |
|---|---|---|---|---|---|
| | a. | $(b_1e_2.s_3o_4(.í_5.t_6o_7)_Σ)_ω$ | *! | | |
| ☞ | b. | $(b_1e_2(.s_3í_5.t_6o_7)_Σ)_ω$ | | * | |
| | c. | $(b_1e_2(.s_3ó_4.t_6o_7)_Σ)_ω$ | | * | *! |

A reviewer asks why the DIM is assumed to attach to the inflected base instead of the stem itself, which would obviate the need for *HIATUS to induce base-final vowel deletion. We first answer a related question: why is the DIM assumed to include its own TE, which must be either [o] or [a]? CaS examples like *coch(ec)ito* < *coche* 'car' (masc.) and *noch(ec)ita* < *noche* 'night' (fem.), cited by Bermúdez-Otero (2006, p. 303), show that the DIM cannot be followed by the TE [e] but, nevertheless, must be followed by some TE, as **cochite/cochecite*, **nochite/nochecite*, **cochit/cochecit*, and **nochit/nochecit* are all ungrammatical. Therefore, instead of attaching between the stem and the base-final TE, the DIM appears with its own TE, [o] for masculine or [a] for feminine nominals. Now, back to the reviewer's question. According to the output–output model that we adopt in Figure 1, Table 2, and tableau (15), the base word $((b_1é_2.s_3o_4)_Σ)_ω$ is a fully formed output that already includes a TE vowel. If the suffix sequence DIM+TE [$i_5t_6$+$o_7$] were to attach directly to the right edge of the stem after [$s_3$], then the fate of the base-final TE [$o_4$] would still need to be explained in any event. An explicit account of the mapping of [$o_4$] must show how high-ranking constraints eliminate additional output candidates that are not shown in (15). For example, inserting [$i_5t_6$+$o_7$] between stem-final [$s_3$] and TE [$o_4$] of the base yields a candidate $(b_1e_2.s_3i_5(.t_6ó_7.o_4)_Σ)_ω$, which fatally violates *HIATUS. Fusing the TE [$o_4$] of the base together with the identical TE [$o_7$] of the DIM into the surface vowel [$o_{4,7}$] violates a high-ranking base-diminutive correspondence constraint, BD-UNIFORMITY, not shown here, which penalizes the coalescence of two segments into one. Low-ranking BD-MAX ensures deletion of the TE as the optimal strategy. Ultimately, the question of where to attach DIM+TE is part of a larger debate on the status of stem-final vowel deletion in Spanish phonology, which we cannot attempt to resolve in this study. For more detailed discussion and analysis of stem-final vowel deletion from a Stratal OT perspective, see Bermúdez-Otero (2006, 2007, 2013).

We propose a novel formal explanation of rightward stress shift in diminutivization, which operates recursively in iterated diminutives. We argue that rightward stress shift in diminutive words emerges from an interaction of constraints on prosodic feet and stress, as defined in (16) (McCarthy 2003, p. 109; Alber and Arndt-Lappe 2022, pp. 17–19). McCarthy (2003, pp. 104–10) criticizes previous OT approaches to stress patterns based on gradient ALIGN constraints because they make pathological cross-linguistic predictions. We adopt his categorically defined constraints (16a,b) (see the constraint definitions in McCarthy's example (33)). Thus, (16c) is violated whenever the stressed vowel of the base is unstressed in the corresponding diminutive word. In particular, our ranking of (16a) above (16b,c) ensures that the Σ coincides with the right edge of the ω. Together, (16d,e) require the stress foot (Σ) to consist of either a single bimoraic syllable or of two monomoraic syllables, the first of which is the most prominent.

(16) a. PARSEσ$_F$
Assign a violation for every unfooted syllable in ω-final position.
b. PARSEσ$_I$
Assign a violation for every unfooted syllable in ω-initial position.
c. BASEDIMINUTIVE-MAXIMALITY(stress)—BD-MAX(str)
Assign a violation for every stressed vowel in the base whose correspondent in the diminutive is unstressed. ("The diminutive contains a maximal expression of the stressed vowel of the base.").
d. FOOTBINARITY—FTBIN
Assign a violation for every Σ that is not binary at the mora or syllable level.
e. TROCHEE—TROCH
Assign a violation for every Σ whose head syllable is not initial.

We argue that PARSEσ$_F$ (16a) outranks both PARSEσ$_I$ and BD-MAX(str) (16b,c), while the relative ranking of FTBIN and TROCH (16d,e) is indeterminate. Tableau (17) continues the analysis of CaS *besito* < *beso* from tableau (15) above, this time focusing on the optimization of foot structure. The unfooted ω-final syllable in [base+DIM+TE] is repaired by aligning the right edge of the ω with the right edge of the Σ, which can be neither monosyllabic (c) nor iambic (d) but instead, must be both disyllabic and trochaic (b). The right-aligned, trochaic stress foot in (b) is optimal because it violates only PARSEσ$_I$ and BD-MAX(str).

| (17) | | ((b$_1$é$_2$.s$_3$o$_4$)$_Σ$)$_ω$+i$_5$t$_6$+o$_7$ | PARSEσ$_F$ | PARSEσ$_I$ | BD-MAX(str) | FTBIN | TROCH |
|---|---|---|---|---|---|---|---|
| | a. | ((b$_1$é$_2$.s$_3$i$_5$)$_Σ$.t$_6$o$_7$)$_ω$ | *! | | | | |
| ☞ | b. | (b$_1$e$_2$(.s$_3$í$_5$.t$_6$o$_7$)$_Σ$)$_ω$ | | * | * | | |
| | c. | (b$_1$e$_2$.s$_3$i$_5$(.t$_6$ó$_7$)$_Σ$)$_ω$ | | * | * | *! | |
| | d. | (b$_1$e$_2$(.s$_3$i$_5$.t$_6$ó$_7$)$_Σ$)$_ω$ | | * | * | | *! |

Tableau (18) gives the analysis of the CaS iterated diminutive *ahoritita* < *ahorita* < *ahora* (Table 2b). To avoid cluttering the tableaux, we henceforth omit numerical subscripts from individual segments. The top half of the tableau shows the optimal mapping between (a(ó.ɾa)$_Σ$)$_ω$+it+a and the corresponding simple diminutive (18b). The bottom half of the tableau shows the optimal mapping between (a.o(.ɾí.ta)$_Σ$)$_ω$+it+a and the corresponding iterated diminutive (18f). In both evaluations, unfooted ω-final syllables in [base+DIM+TE] are repaired by aligning the right edge of the ω with the right edge of the Σ, which can be neither monosyllabic (c,g) nor iambic (d,h) but instead, must be both disyllabic and trochaic (b,f).

| (18) | | (a(ó.ɾa)$_Σ$)$_ω$+it+a | PARSEσ$_F$ | PARSEσ$_I$ | BD-MAX(str) | FTBIN | TROCH |
|---|---|---|---|---|---|---|---|
| | a. | (a(.ó.ɾi)$_Σ$.ta)$_ω$ | *! | * | | | |
| ☞ | b. | (a.o(.ɾí.ta)$_Σ$)$_ω$ | | * | * | | |
| | c. | (a.o.ɾi(.tá)$_Σ$)$_ω$ | | * | * | *! | |
| | d. | (a.o(.ɾi.tá)$_Σ$)$_ω$ | | * | * | | *! |
| (18b) → | | (a.o(.ɾí.ta)$_Σ$)$_ω$+it+a | | | | | |
| | e. | (a.o(.ɾí.ti)$_Σ$.ta)$_ω$ | *! | * | | | |
| ☞ | f. | (a.o.ɾi(.tí.ta)$_Σ$)$_ω$ | | * | * | | |
| | g. | (a.o.ɾi.ti(.tá)$_Σ$)$_ω$ | | * | * | *! | |
| | h. | (a.o.ɾi(.ti.tá)$_Σ$)$_ω$ | | * | * | | *! |

The same constraint rankings that account for base-final TE deletion (15b) and rightward stress shift in simple (17b), (18b) and iterated (18f) diminutives in CaS can be assumed to apply in both JS and CoS. This accounts for generalization (13d), which pertains to all three Hispano-Romance varieties. We now turn to cross-linguistic differences.

## 4. Diminutive Suffix Allomorphy Conditioned by Consonantal Place Dissimilation

In this section, we provide an OT account of the JS/CoS DIM [ik] ~ [it] alternation in terms of phonologically conditioned allomorphy. We argue that the DIM has two input allomorphs, /ik/ and /it/, and that the two Hispano-Romance varieties choose one or the other as the preferred allomorph. An argument in favor of this approach comes from the fact that neither variety has a regular, synchronic process of consonantal place dissimilation outside the morphological context of the DIM.

Recent accounts of phonologically conditioned allomorphy posit lexically ordered allomorphs and a PRIORITY constraint enforcing the lexical ordering, which interacts with other constraints in an OT grammar (Bonet et al. 2007; Mascaró 2007; McCarvel 2016, among others). Following McCarvel (2016), we define PRIORITY (19) as a markedness constraint that is violated when the unmarked (default) allomorph is not selected:

(19) PRIORITY (McCarvel 2016, p. 43; cf. Mascaró 2007, p. 726)
Assign a violation for use of any allomorph in the output other than the unmarked (default) allomorph.

Previous studies using PRIORITY (Bonet et al. 2007; Mascaró 2007; Bradley and Smith 2011; McCarvel 2016, among others) order the preferred allomorph above, '›', other allomorphs, a convention that we also adopt here. Table 3 gives the representation of the DIM across three Hispano-Romance varieties:

**Table 3.** Representations of the shorter DIM in JS, CoS, and CaS. ('›' = 'is preferred over').

| Variety | Representation |
|---|---|
| JS | DIM = {ik›it} |
| CoS | DIM = {it›ik} |
| CaS | DIM = it |

JS and CoS have both [ik] and [it] enclosed within braces { }, with the unmarked (default) allomorph ordered above the non-default allomorph. CaS has [it] but lacks [ik]. Furthermore, in the output–output model of diminutive formation in Figure 1, the DIM representations of Table 3 are technically not input but output forms, which aligns with McCarvel's (2016) definition of PRIORITY as not a faithfulness but a markedness constraint. In classic OT, input–output faithfulness evaluates corresponding inputs and outputs, while markedness and output–output faithfulness evaluate only outputs.

Given the output–output model in Figure 1, when one of the allomorphs of DIM is selected, output–output faithfulness constraints on features and segments are vacuously satisfied. Since faithfulness is irrelevant, the optimal allomorph is decided by the interaction between PRIORITY and other markedness constraints in the grammar. In the OT literature, this effect is commonly known under the rubric of The Emergence of the Unmarked (TETU): markedness constraints that are typically inactive obtain the chance to exert their effects when higher-ranking faithfulness constraints are rendered inert. Beyond the specific context in which multiple allomorphs are available, faithfulness constraints otherwise prevent changing the features and segments of morphemes. This approach to consonantal place dissimilation in JS/CoS naturally explains why the alternation emerges only in the context of diminutive suffix allomorphy.

We adopt the phonological features in Table 4, which permit a representational distinction among the relevant segments as a function of consonantal place (C-Place) and manner specifications. [coronal] consonants involve the tongue tip or blade and [dorsal] consonants, the tongue dorsum. [continuant] segments are produced with (–) or without (+) full vocal tract closure. [voice] segments are produced with (+) or without (–) vocal fold vibration (see Chomsky and Halle 1968, pp. 293–329; Hayes 2009, pp. 70–102).

**Table 4.** Relevant phonological features of coronal and dorsal consonants.

| | t | d | s | z | k | g | x | w |
|---|---|---|---|---|---|---|---|---|
| C-Place | cor | cor | cor | cor | dors | dors | dors | dors |
| [cont] | − | − | + | + | − | − | + | + |
| [voi] | − | + | − | + | − | + | − | + |

The base-diminutive constraint (20) requires corresponding segments in the outputs [base+DIM+TE] and [diminutive] to have identical C-Place features.

(20) BASEDIMINUTIVE-IDENTITY(C-Place)—BD-IDENT(C-Pl)
Assign a violation for every segment of [base+DIM+TE] whose correspondent in the diminutive does not have identical C-Place.

The next step towards an analysis of dissimilation is to define the markedness constraints that are responsible for the alternation. Dissimilation can be understood as the avoidance, within some locally defined prosodic domain, of sufficiently similar consonants, defined in terms of shared phonological features. de Lacy (2006, p. 2) formalizes a universal markedness scale for C-Place, enclosed by vertical lines '|' in (21a):

| | | |
|---|---|---|
| (21) a. | \| dorsal ▸ labial ▸ coronal ▸ glottal \| ('▸' = 'is more marked than') | |
| b. | *{dors} | Assign a violation for each [dorsal] feature. |
| c. | *{dors,lab} | Assign a violation for each [dorsal] and each [labial] feature. |
| d. | *{dors,lab,cor} | Assign a violation for each [dorsal], each [labial], and each [coronal] feature. |
| e. | *{dors,lab,cor,gl} | Assign a violation for each [dorsal], each [labial], each [coronal], and each [glottal] feature. |

We use the symbol '▸' to indicate markedness relations among C-Place features, which are opposite in directionality to the allomorph preference relations expressed by '›' in Table 3. For example, |dorsal ▸ labial| means that dorsal C-Place is more marked (and less preferred) than labial C-Place, while JS {ik›it} means that the allomorph [ik] is more preferred (and less marked) than [it]. The scale in (21a) projects the markedness constraints (21b–e), which assign a violation for each C-Place feature enclosed within braces { } in the output. These constraints form a stringency hierarchy, which means that their universal ranking need not be stipulated in order to account for cross-linguistic implications among C-Place features. For example, the dorsal stop [k] violates both *{dors} and *{dors,lab,cor}, while the coronal stop [t] violates only *{dor,lab,cor}. Therefore, [k] is inherently more marked than [t], even if *{dors} is ranked below *{dors,lab,cor}. The stringency relation exists because violations of the more specific *{dors} form a subset of the violations of the more general *{dors,lab,cor}. This specific-to-general relationship follows from the fact that each markedness constraint includes all of the more marked C-Place features that appear to the left of '▸' along the scale in (21a). See de Lacy (2006) for further explanation of how stringency makes stipulating universally fixed rankings in OT unnecessary.

A theory-internal motivation for ordering the JS DIM [ik] as the default allomorph above [it] in Table 3 comes from the fact that the C-Place hierarchy alone is incapable of selecting the correct allomorph. If we consider an unordered set of DIM allomorphs {it,ik}, as in tableau (22), then [it] is incorrectly chosen as optimal, regardless of the ranking of the three constraints. Either violation of *{dors} or *{dors,lab} by (22b) can be considered fatal, as indicated by the parentheses around the exclamation points. The symbol '💣' means that (22a) is predicted as the winner but should not be. By contrast, assuming an ordered set of DIM allomorphs {ik›it}, as in tableau (23), allows for high-ranking PRIORITY to choose the correct allomorph (23b), even though the [k] of [ik] is more marked in terms of C-Place features. The designation of [ik] as the default allomorph in JS is also motivated by the fact that [ik] appears as part of the longer allomorph *-eziko/a* (see Table 1) and by the wider surface distribution of [ik] in simple diminutives, as discussed in Section 2.1.

| (22) | {it,ik} | *{dors} | *{dors,lab} | *{dors,lab,cor} |
|---|---|---|---|---|
| 💣 a. | it | | | * |
| b. | ik | *(!) | *(!) | * |

| (23) | {ik›it} | PRIORITY | *{dors} | *{dors,lab} | *{dors,lab,cor} |
|---|---|---|---|---|---|
| a. | it | *! | | | * |
| ☞ b. | ik | | * | * | * |

Dissimilation can be formalized in OT as the local self-conjunction of markedness constraints (Alderete 1997; Ito and Mester 1998, 2003; McCarthy 2008, p. 218). The key idea behind this approach is that having multiple constraint violations in a local domain is categorically worse than having the same violations in a non-local domain. We argue that consonants in the initial position of two adjacent syllables are evaluated by the self-conjoined markedness constraints (24a,b). The local self-conjunction of *{dors} (21b) yields (24a), abbreviated as *.K$^2$, while the local self-conjunction of *{dors,lab,cor} (21d) in combination with the features [–cont,–voi] yields *.{kpt}$^2$ (24b):

(24) a. *.{dors}$^2$—*.K$^2$
Assign a violation for every pair of [dorsal] consonants in the initial position of adjacent syllables.

b. *.[{dors,lab,cor},–cont,–voi]$^2$—*.{kpt}$^2$
Assign a violation for every pair of [dorsal], [labial], or [coronal] voiceless stops in the initial position of adjacent syllables.

The superscript $^2$ means that the constraint penalizes two occurrences, i.e., a repetition, of the relevant features across the initial positions of two adjacent syllables. Thus, *.K$^2$ is violated by the repetition of any syllable-initial [dorsal] consonant and is responsible for dissimilation in JS. A question arises as to why *.{kpt}$^2$ makes reference to [dorsal] and [labial], when [coronal] is the relevant C-Place feature that triggers dissimilation in CoS. Why not define (24b) more simply as *.[cor,–cont,–voi]$^2$, abbreviated as *.t$^2$? The answer is that markedness constraints (21b–e) form a stringency hierarchy, based on how the violations of more specific constraints form subsets of the violations of more general constraints. The only way to pick out [coronal] (and exclude [glottal]) from the scale in (21a)—while maintaining stringency—is to also include the relatively more marked [dorsal] and [labial] C-Place features as part of the definition, in both (21d) and (24b). The prediction is that the repetition of syllable-initial [p] would also be avoided, even though our data sets fail to provide such evidence, as none of the alternations under analysis crucially involves bilabials. Thank you to Brechtje Post (personal communication) for helpful discussion.

In preparation for its application to the three varieties under study, we briefly summarize the key elements of our OT analysis introduced above. PRIORITY (19) enforces the lexical ordering of allomorphs in Table 3. BD-IDENT(C-Pl) (20) requires faithfulness to consonantal place features, while (21b–e) formalize C-Place markedness in terms of stringently related constraints. Further, (24a,b) encode the avoidance of marked structures that repeat sufficiently similar consonants in the initial position of adjacent syllables.

*4.1. JS*

The examples of JS simple and iterated diminutives in Table 5 show that (i) {ik›it}+TE attaches to the right edge of the base, shifting the Σ rightward in the diminutive ω, and (ii) base-final vowels delete before the vowel-initial DIM {ik›it}.

**Table 5.** JS simple diminutive *kapika* < *kapa* 'cap' (a) and iterated diminutive *pokitiko* < *pokito* < *poko* 'tiny bit' (b).

| | [Base+DIM+TE] | | [Diminutive] | |
|---|---|---|---|---|
| (a) | ((ká.pa)$_{\Sigma}$)$_{\omega}$+{ik›it}+a | → | (ka(.pí.ka)$_{\Sigma}$)$_{\omega}$ | |
| (b) | ((pó.ko)$_{\Sigma}$)$_{\omega}$+{ik›it}+o | → | (po(.kí.to)$_{\Sigma}$)$_{\omega}$ | → |
| | (po(.kí.to)$_{\Sigma}$)$_{\omega}$+{ik›it}+o | → | (po.ki(.tí.ko)$_{\Sigma}$)$_{\omega}$ | |

We argue that in the JS grammar, BD-IDENT(C-Pl) dominates *.K$^2$, which in turn dominates both PRIORITY and *.{kpt}$^2$. In JS *kapika* < *kapa* (25), BD-IDENT(C-Pl) prevents changing C-Place in the consonant of the base-final syllable (c). PRIORITY prevents selecting the non-default DIM [it] (b). The default DIM [ik] emerges as optimal after bilabial [p] (a).

| (25) | | ((ká.pa)$_{\Sigma}$)$_{\omega}$+{ik›it}+a | BD-IDENT(C-Pl) | *.K$^2$ | PRIORITY | *.{kpt}$^2$ |
|---|---|---|---|---|---|---|
| ☞ | a. | (ka(.pí.ka)$_{\Sigma}$)$_{\omega}$ | | | | |
| | b. | (ka(.pí.ta)$_{\Sigma}$)$_{\omega}$ | | | *! | |
| | c. | (ka(.tí.ka)$_{\Sigma}$)$_{\omega}$ | *! | | | |

In the analysis of JS *pokitiko* < *pokito* < *poko*, illustrated in tableau (26) below, BD-IDENT(C-Pl) prevents changing C-Place in consonants of the base (c,f,g). In the simple diminutive, *.K$^2$ prevents initial [dorsal] consonants in adjacent syllables (a), and dissimilation involves selecting the non-default DIM [it] after [k] while tolerating the violation of low-ranking PRIORITY (b). In the iterated diminutive, PRIORITY or *.{kpt} prevents selecting the non-default DIM [it] (e), and dissimilation involves selecting default [ik] after [t] (d). The relationship between PRIORITY and *.{kpt} is indeterminate because either ranking guarantees the same winners in (b,d). The ranking of *.K$^2$ above PRIORITY ensures that JS DIM allomorphs [it] in *pokito* (b) and [ik] in *pokitiko* (d) alternate across successive syllables within the diminutive ω.

Although not shown here, the input–output correspondence constraint IO-IDENT(C-Pl) ranks above both *.K$^2$ and *.{kpt}$^2$. Along with high-ranking BD-IDENT(C-Pl), this ensures that dissimilation in C-Place can emerge as an unmarked realization only in morphological environments derived by the suffixation of a set of DIM allomorphs. For example, the [k. . .k] and [t. . .t] sequences in JS *kakao* 'cocoa' and *total* 'total' (Nehama 1977, pp. 261, 560) are immune to dissimilation because they appear in non-diminutive words that do not involve allomorph competition. In this context, high-ranking IO-IDENT(C-Pl) prevents changing either consonant of the sequence, and violations of *.K$^2$ and *.{kpt}$^2$ are tolerated in the output. By contrast, C-Place correspondence is inactive when the DIM includes two allomorphs, as selecting either allomorph will vacuously satisfy correspondence. The optimal allomorph emerges instead from the interaction among lower-ranking constraints PRIORITY, *.K$^2$, and *.{kpt}$^2$, as shown in (25) and (26). We assume that IO-IDENT(C-Pl) also ranks high in CaS and CoS, as in JS.

| (26) | ((pó.ko)$_{\Sigma}$)$_{\omega}$+{ik›it}+o | BD-IDENT(C-Pl) | *.K$^2$ | PRIORITY | *.{kpt}$^2$ |
|---|---|---|---|---|---|
| a. | (po(.kí.ko)$_{\Sigma}$)$_{\omega}$ | | *! | | * |
| ☞ b. | (po(.kí.to)$_{\Sigma}$)$_{\omega}$ | | | * | |
| c. | (po(.tí.ko)$_{\Sigma}$)$_{\omega}$ | *! | | | |
| (26b) → | (po(.kí.to)$_{\Sigma}$)$_{\omega}$+{ik›it}+o | | | | |
| ☞ d. | (po.ki(.tí.ko)$_{\Sigma}$)$_{\omega}$ | | | | |
| e. | (po.ki(.tí.to)$_{\Sigma}$)$_{\omega}$ | | | *(!) | *(!) |
| f. | (po.ki(.kí.ko)$_{\Sigma}$)$_{\omega}$ | *! | ** | | ** |
| g. | (po.ki(.kí.to)$_{\Sigma}$)$_{\omega}$ | *! | * | * | * |

### 4.2. *CoS*

Dissimilation is variable in CoS simple diminutives (5) and (6). CoS iterated diminutives (7)–(10) show three distinct patterns, summarized in Table 6: (a) only the default DIM [it] iterates (cf. CaS (12)), (b) [it] ~ [ik] alternate (cf. JS (3a)), and (c) [it] iterates until alternating with word-final [ik].

**Table 6.** CoS iterated diminutives *cortitita* < *cortita* < *corta* 'short' (a), *cortiquita* < *cortica* < *corta* (b), and *cortitica* < *cortita* < *corta* (c).

| | [Base+DIM+TE] | | [Diminutive] | |
|---|---|---|---|---|
| (a) | ((kóɾ.ta)$_{\Sigma}$)$_{\omega}$+{it›ik}+a | → | (koɾ(.tí.ta)$_{\Sigma}$)$_{\omega}$ | → |
| | (koɾ(.tí.ta)$_{\Sigma}$)$_{\omega}$+{it›ik}+a | → | (koɾ.ti(.tí.ta)$_{\Sigma}$)$_{\omega}$ | |
| (b) | ((kóɾ.ta)$_{\Sigma}$)$_{\omega}$+{it›ik}+a | → | (koɾ(.tí.ka)$_{\Sigma}$)$_{\omega}$ | → |
| | (koɾ(.tí.ka)$_{\Sigma}$)$_{\omega}$+{it›ik}+a | → | (koɾ.ti(.kí.ta)$_{\Sigma}$)$_{\omega}$ | |
| (c) | ((kóɾ.ta)$_{\Sigma}$)$_{\omega}$+{it›ik}+a | → | (koɾ(.tí.ta)$_{\Sigma}$)$_{\omega}$ | → |
| | (koɾ(.tí.ta)$_{\Sigma}$)$_{\omega}$+{it›ik}+a | → | (koɾ.ti(.tí.ka)$_{\Sigma}$)$_{\omega}$ | |

The three CoS patterns differ in whether and when the ranking of PRIORITY above *.{kpt}$^2$ is reversed during the recursive evaluation of iterated diminutive outputs. Pattern (a) in Table 6 is predicted by maintaining the high-ranking of PRIORITY for both simple (27a) and iterated (27d) diminutives. Pattern (b) is predicted by a re-ranking of PRIORITY below *.{kpt}$^2$ for both simple (28b) and iterated (28d) diminutives. Finally, pattern (c) is predicted by a re-ranking of PRIORITY below *.{kpt}$^2$ at the point when the final DIM suffix is added to the base (29e). In CoS, the DIM [ik] never repeats without alternating because BD-IDENT(C-Pl) prevents changing syllable-initial [t] in the base, in both simple (29c) and iterated (29f–k) diminutives.

| (27) | ((kóɾ.ta)$_{\Sigma}$)$_{\omega}$+{it›ik}+a | BD-IDENT(C-Pl) | *.K$^2$ | PRIORITY | *.{kpt}$^2$ |
|---|---|---|---|---|---|
| ☞ a. | (koɾ(.tí.ta)$_{\Sigma}$)$_{\omega}$ | | | | * |
| b. | (koɾ(.tí.ka)$_{\Sigma}$)$_{\omega}$ | | | *! | |
| c. | (koɾ(.kí.ta)$_{\Sigma}$)$_{\omega}$ | *! | * | | * |
| (27a) → | (koɾ(.tí.ta)$_{\Sigma}$)$_{\omega}$+{it›ik}+a | | | | |
| ☞ d. | (koɾ.ti(.tí.ta)$_{\Sigma}$)$_{\omega}$ | | | | ** |
| e. | (koɾ.ti(.tí.ka)$_{\Sigma}$)$_{\omega}$ | | | *! | |
| f. | (koɾ.ti(.kí.ta)$_{\Sigma}$)$_{\omega}$ | *! | | | |

| (28) | ((kóɾ.ta)$_{\Sigma}$)$_{\omega}$+{it›ik}+a | BD-IDENT(C-Pl) | *.K$^2$ | *.{kpt}$^2$ | PRIORITY |
|---|---|---|---|---|---|
| | a. (koɾ(.tí.ta)$_{\Sigma}$)$_{\omega}$ | | | *! | |
| ☞ | b. (koɾ(.tí.ka)$_{\Sigma}$)$_{\omega}$ | | | | * |
| | c. (koɾ(.kí.ta)$_{\Sigma}$)$_{\omega}$ | *! | * | * | |
| (28b) → | (koɾ(.tí.ka)$_{\Sigma}$)$_{\omega}$+{it›ik}+a | | | | |
| ☞ | d. (koɾ.ti(.kí.ta)$_{\Sigma}$)$_{\omega}$ | | | | |
| | e. (koɾ.ti(.kí.ka)$_{\Sigma}$)$_{\omega}$ | | *(!) | *(!) | * |
| | f. (koɾ.ti(.tí.ka)$_{\Sigma}$)$_{\omega}$ | *! | | | * |

| (29) | ((kóɾ.ta)$_{\Sigma}$)$_{\omega}$+{it›ik}+a | BD-IDENT(C-Pl) | *.K$^2$ | PRIORITY | *.{kpt}$^2$ |
|---|---|---|---|---|---|
| ☞ | a. (koɾ(.tí.ta)$_{\Sigma}$)$_{\omega}$ | | | | * |
| | b. (koɾ(.tí.ka)$_{\Sigma}$)$_{\omega}$ | | | *! | |
| | c. (koɾ(.kí.ta)$_{\Sigma}$)$_{\omega}$ | *! | * | | * |
| (29a) → | (koɾ(.tí.ta)$_{\Sigma}$)$_{\omega}$+{it›ik}+a | | | *.{kpt}$^2$ | PRIORITY |
| | d. (koɾ.ti(.tí.ta)$_{\Sigma}$)$_{\omega}$ | | | **! | |
| ☞ | e. (koɾ.ti(.tí.ka)$_{\Sigma}$)$_{\omega}$ | | | * | * |
| | f. (koɾ.ti(.kí.ta)$_{\Sigma}$)$_{\omega}$ | *! | | | |
| | g. (koɾ.ti(.kí.ka)$_{\Sigma}$)$_{\omega}$ | *! | * | * | * |
| | h. (koɾ.ki(.kí.ka)$_{\Sigma}$)$_{\omega}$ | *!* | *** | *** | * |
| | i. (koɾ.ki(.tí.ka)$_{\Sigma}$)$_{\omega}$ | *! | * | * | * |
| | j. (koɾ.ki(.kí.ta)$_{\Sigma}$)$_{\omega}$ | *!* | ** | ** | |
| | k. (koɾ.ki(.tí.ta)$_{\Sigma}$)$_{\omega}$ | *! | * | * | |

More generally, the stringency relationship between *.K$^2$ and *.{kpt}$^2$ ensures that, no matter how the two self-conjoined markedness constraints are ranked, a structure such as [.ki.ka] (30b) and (31b) is inherently more marked than [.ti.ta] (30a) and (31a):

| (30) | | *.K$^2$ | *.{kpt}$^2$ |
|---|---|---|---|
| ☞ | a. .ti.ta | | * |
| | b. .ki.ka | *! | * |

| (31) | | *.{kpt}$^2$ | *.K$^2$ |
|---|---|---|---|
| ☞ | a. .ti.ta | * | |
| | b. .ki.ka | * | *! |

Although there may be differences in the language-specific ordering of DIM allomorphs and the relative ranking of PRIORITY, repeating initial [k] across two adjacent syllables will always be relatively more marked than repeating initial [t] in the same context.

Recall that the JS DIM [ik] alternates with [it] after any syllable-initial dorsal consonant (2), whereas the CoS DIM [it] alternates with [ik] only after syllable-initial [t] (5) and (8)–(10). Why do these two grammars differ in what counts as sufficiently similar? The answer has to do with how we formalized the self-conjoined markedness constraints that are responsible for dissimilation. While (24a) refers simply to the most marked C-Place feature [dorsal], (24b) refers to the less marked [coronal] in combination with two specific manner features. In JS *bragita* < *braga* 'trousers' (2b) and *blahitu* < *blahu* 'Christian' (2c), *.K$^2$ (24a) prevents initial dorsals in adjacent syllables, *even if they differ* in manner features, e.g., [voice] in *[.gí.ka] (32a) or [continuant] in *[.xí.ku] (33a):[7]

| (32) | ((bɾá.ga)Σ)ω+{ik›it}+a | *.K$^2$ | *.{kpt}$^2$ | PRIORITY |
|---|---|---|---|---|
| a. | (bɾa(.gí.ka)Σ)ω | *! | | |
| ☞ b. | (bɾa(.gí.ta)Σ)ω | | | * |

| (33) | ((blá.xu)Σ)ω+{ik›it}+u | *.K$^2$ | *.{kpt}$^2$ | PRIORITY |
|---|---|---|---|---|
| a. | (bla(.xí.ku)Σ)ω | *! | | |
| ☞ b. | (bla(.xí.tu)Σ)ω | | | * |

In CoS *caldito* < *caldo* 'soup, broth' and *bracito* < *brazo* 'arm' (4b), *.{kpt}$^2$ (24b) allows initial coronals in adjacent syllables, *as long as they differ* in manner features, e.g., [voice] in [.dí.to] (34a) or [continuant] in [.sí.to] (35a):

| (34) | ((kál.do)Σ)ω+{it›ik}+o | *.K$^2$ | *.{kpt}$^2$ | PRIORITY |
|---|---|---|---|---|
| ☞ a. | (kal(.dí.to)Σ)ω | | | |
| b. | (kal(.dí.ko)Σ)ω | | | *! |
| | | | | |

| (35) | ((bɾá.so)Σ)ω+{it›ik}+o | *.K$^2$ | *.{kpt}$^2$ | PRIORITY |
|---|---|---|---|---|
| ☞ a. | (bɾa(.sí.to)Σ)ω | | | |
| b. | (bɾa(.sí.ko)Σ)ω | | | *! |

### *4.3. CaS*

We return, at last, to the baseline variety CaS, which has the DIM [it] but lacks [ik]. If there is no dissimilatory alternation, then allomorph ordering plays no decisive role in the analysis, and the ranking of PRIORITY with respect to *.K$^2$ and *.{kpt}$^2$ is indeterminate. High-ranking faithfulness constraints IO-IDENT(C-Pl) and BD-IDENT(C-Pl) generally prevent changing C-Place features in bases and in simple and iterated diminutives. This means that violations of low-ranking similarity avoidance constraints *.K$^2$ and *.{kpt}$^2$ will be tolerated in CaS, which allows for recursive iteration of [it] without C-Place dissimilation.

### *4.4. Summary*

The OT analysis presented in this section accounts for the remaining generalizations left unaddressed in Section 3. First, (13a) is understood as cross-linguistic variation in the phonological representation of the DIM, as shown in Table 3. If there are two ordered allomorphs, then PRIORITY can potentially enforce a preference for the default allomorph. Second, the inclusion of [–cont,–voi] along with C-Place features in the definition of *.{kpt}$^2$ (24b) accounts for the difference between CoS and JS in (13b). In diminutive suffixation, the syllable-initial consonants that trigger dissimilation need share only the feature [dorsal] in JS, but in CoS, must be totally identical, sharing [coronal], [–continuant], and [–voice]. Finally, differences in the ranking of PRIORITY and *.{kpt}$^2$ account for generalization (13c), thereby predicting the three patterns of iterated diminutives as a function of whether and when re-ranking occurs during the recursive process of candidate evaluation.

## **5. Theoretical Alternatives**

To our knowledge, V. M. Prieto (2005, p. 34) is the first to suggest that the alternation between DIM [it] and [ik] in Latin American Spanish can be captured in OT by a markedness constraint based on the Obligatory Contour Principle (OCP). The OCP was originally proposed in tonal phonology (Leben 1973; Goldsmith 1976) and was later extended to segmental phonology (McCarthy 1986) as a way of prohibiting two adjacent identical elements on a given phonological tier. V. M. Prieto argues that a high-ranked OCP constraint can explain, for example, why the diminutive of *carta* 'letter' is *cartica* instead of **cartita:* *[t. . .t] has two adjacent identical segments on the consonantal tier, which [t. . .k] avoids. However, a serious problem with this approach is that a general OCP constraint against adjacent identical segments fails to explain why JS [ik] alternates with [it] after *any* dorsal consonant, regardless of manner features. Apparently, the syllable-initial consonants that trigger dissimilation need not be totally identical. A solution to this problem in JS phonology would be to allow the OCP to refer to the C-Place feature [dorsal] alone.

Bradley and Smith (2011) propose such a solution to account for the alternation between JS *-iko/a* and *-ito/a:* "Avoid [dorsal] consonants that are adjacent across an intervening vowel" (p. 22). An additional problem for both Bradley and Smith's OCP(dorsal) and V.M. Prieto's more general OCP constraint is that segmental adjacency on the consonantal tier fails to explain dissimilation in, e.g., JS *negrito* < *negro* 'bad' (see (38b) in Appendix A) and CoS *maestrico/a* < *maestro/a* 'teacher' (Lipski 1999), *letrica* < *letra* 'letter' (see (40) of Appendix B), and *atlicas* < *atlas* 'atlas' (see Note 4). In **negriko*, **maestrito/a*, **letrita*, and **atlitas*, a liquid intervenes between the two consonants in [g. . .k] and [t. . .t], which are not adjacent on the consonantal tier and, therefore, should be unable to trigger the OCP violations.

Both of these shortcomings constitute a strong argument in favor of the similarity avoidance constraints *.$K^2$ (24a) and *.$\{kpt\}^2$ (24b), which correctly assign a violation to every pair of sufficiently similar consonants in the initial position of adjacent syllables. Because adjacency on the consonantal tier is not required, this approach correctly prohibits marked sequences in which the triggering consonants appear in the initial position of adjacent syllables even when there is an intervening liquid in the second position of the first onset, rendering the relevant consonants technically non-adjacent.

Although they both invoke the OCP as a way of modeling the C-Place alternation between diminutive suffixes, V. M. Prieto (2005, p. 34) and Bradley and Smith (2011) do not attempt to explain why dissimilation is blocked from applying in non-diminutive words, as can be observed in JS *kakao* 'cocoa' and *total* 'total'. An advantage of the analysis proposed in the present article is that high-ranking IO-IDENT(C-Pl) and BD-IDENT(C-Pl) effectively restrict dissimilation to environments derived by the suffixation of a set of ordered DIM allomorphs.

Previous Stratal OT accounts of Hispano-Romance diminutive formation neither acknowledge nor attempt to model the C-Place alternation between DIM [ik] and [it]. Bermúdez-Otero (2006, 2007, 2013) proposes that the Spanish DIM /it/ is attached to the morphological stem within the Word Level (WL) domain, where it triggers a process of stem-final unstressed vowel (TE) deletion. A key argument for assuming that stems are lexically stored with TE vowels, and that such vowels are phonologically deleted before vowel-initial suffixes, comes from words that show a stem-based alternation between unstressed mid vowels /o,e/ and corresponding stressed diphthongs /we,je/. Specifically, diphthongization is argued to apply unexpectedly, i.e., to overapply, in the unstressed syllable of the diminutive form of such words. For example, both *portero* 'doorman' (36a) and *puertita* 'little door' (36b) share the same stem, /p{o,we}ɾt-a/, which includes /o/ and /we/ as part of two lexically listed allomorphs, i.e., /poɾt-a/ and /pweɾt-a/ (Bermúdez-Otero 2006, p. 286):

| (36) | | a. | *portero* | b. | *puertita* |
|---|---|---|---|---|---|
| | domain structure | | ⟦$_{WL}$ ⟦$_{SL}$ p{o,we}ɾt-a-eɾ-o ⟧⟧ | | ⟦$_{WL}$ ⟦$_{SL}$ p{o,we}ɾt-a ⟧it-a ⟧ |
| | SL | | poɾ.té.ɾo | | pwéɾ.ta |
| | WL | | poɾ.té.ɾo | | pweɾ.tí.ta |
| | | | 'doorman' | | 'little door' |

The masculine suffix *-ero* (cf. feminine *-era*) which derives, for example, names of professions, is attached to the stem /p{o,we}ɾt-a/ within the Stem Level (SL) domain, indicated by hollow brackets ⟦$_{SL}$ . . . ⟧(36a). Since stress is assigned to the penultimate syllable in (poɾ(.té.ɾo)$_{\Sigma}$)$_{\omega}$ and ((pwéɾ.ta))$_{\Sigma}$)$_{\omega}$, the SL constraint ranking (not shown here) optimizes unstressed [o] in the former versus stressed [wé] in the latter. These SL outputs then become inputs to the constraint ranking that operates over the WL domain ⟦$_{\text{WL}}$ . . . ⟧, where the diminutive suffix is attached (36b). Although stress is shifted rightward, [we] is still maintained in the unstressed antepenultimate syllable of the WL output (pweɾ(.tí.ta))$_{\Sigma}$)$_{\omega}$ in (36b). Diphthongization turns out to be opaque because it applies in the SL domain, before the DIM is attached in the WL domain.[8]

Ohannesian and Pons (2009) provide an alternative, monostratal OT analysis of opaque diphthongization in diminutives using output–output faithfulness. In their approach, morphologically related pairs such as <pwéɾ.ta, pweɾ.tí.ta> constitute "a subparadigm included in the paradigm which comprises all the words derived from the base" (p. 88). A high-ranking output–output faithfulness constraint on the base-diminutive subparadigm eliminates the whole paradigm candidate <pwéɾ.ta, poɾ.tí.ta, poɾ.té.ɾo>, in which members of the subparadigm <pwéɾ.ta, poɾ.tí.ta> differ with respect to the diphthong versus mid vowel distinction. The whole paradigm candidate <pwéɾ.ta, pweɾ.tí.ta, poɾ.té.ɾo> is optimal, in which diphthongization overapplies in the base-diminutive subparadigm. However, Ohannesian and Pons (2009), like Bermúdez-Otero (2006, 2007, 2013), do not consider C-Place dissimilation.

Because no Stratal OT account of DIM [ik] and [it] has ever been proposed in the literature, we are reluctant to hastily put together an explicit analysis only to then argue against a strawman. This would require us to commit to a number of theoretical assumptions that practitioners of Stratal OT may or may not share and that are ultimately orthogonal to debate about how many constraint strata there should be in an OT grammar. However, we can sketch out several complexities that a future Stratal OT approach would need to contend with. Since Bermúdez-Otero's account in (36) already appeals to input allomorphy for stems that show an alternation between mid vowels and corresponding diphthongs, a plausible Stratal OT approach to dissimilation in diminutive suffixation might also appeal to two allomorphs [ik] and [it], as in (22). However, the greater C-Place markedness of [ik] over [it] would still seem to require a default vs. non-default ordering of allomorphs, along with a PRIORITY constraint to enforce it, at least in JS (23). Alternatively, if one were to posit JS /ik/ versus CoS /it/ and treat C-Place dissimilation as purely phonological, then it would become necessary to restrict the alternation in some way to the domain of suffixes. Either way, the recursive process of iterated diminutivization would presumably require multiple DIM+TE sequences to be attached to the stem within the WL domain. Under this scenario, the similarity avoidance constraints (24a,b) could start out as low-ranked at the SL stratum, to allow for, e.g., JS *kakao* 'cocoa' and *total* 'total'. Similarity avoidance could then be promoted above input–output faithfulness to C-Place features at the WL stratum, thereby enforcing dissimilation between the onsets of adjacent syllables across multiple DIM+TE sequences (assuming deletion of every non-final TE).

However, it is not clear how such an approach would account for the difference between CoS *cortiquita* (b) and *cortitica* (c) in Table 6. Our understanding is that in Stratal OT, a single constraint ranking is assumed to operate over a given stratum, and constraints can be re-ranked only at the next stratum. As our analysis in tableau (29) makes clear, PRIORITY needs to be re-ranked below *.{kpt}$^2$ once the final DIM+TE is attached to the latest base. Since our OT model is monostratal, this re-ranking necessarily takes place not across strata but within a single stratum of the phonological grammar.

## 6. Conclusions

Based on the data and generalizations presented in Section 2, we have argued that JS, CoS, and CaS differ in their phonological representation of the shorter diminutive suffix, DIM. When there are two surface variants in the same grammar, the representation of the DIM encodes a language-specific preference for the default allomorph, which appears to the left of '›' in the set of allomorphs enclosed within braces: JS {ik›it} and CoS {it›ik} (cf. CaS it), as shown in Table 3.

In Sections 3 and 4, we have developed a formal account of C-Place dissimilation in diminutive suffixation across three Hispano-Romance varieties, within the classic monostratal OT framework. As summarized in Table 7, this way of understanding and explicitly comparing closely related phonological grammars reveals that two of the same constraint rankings are shared in common (a,b) and that cross-linguistic variation stems from crucial re-rankings of PRIORITY and *.{kpt}$^2$ (c–f). By contrast, PRIORITY is irrelevant and, therefore, not shown in CaS (g).

**Table 7.** Summary of constraint rankings, effects, and grammars in Hispano-Romance diminutive suffixation. '»' indicates a crucial ranking of adjacent constraints, and the comma a lack thereof.

| | Constraint Ranking | Effect | Variety |
|---|---|---|---|
| (a) | *HIATUS, BD-MAX(V2) » BD-MAX | Base-final vowel deletion in (15) | CaS, JS, CoS |
| (b) | FTBIN, TROCH, PARSE$\sigma_F$ » PARSE$\sigma_I$, BD-MAX(str) | Rightward stress shift in (17) and (18) | CaS, JS, CoS |
| (c) | IO/BD-ID(C-Pl) » *.K$^2$ » PRIORITY, *.{kpt}$^2$ | DIM [it] ~ [ik] alternation in (25) and (26) | JS |
| (d) | IO/BD-ID(C-Pl) » *.K$^2$, PRIORITY » *.{kpt}$^2$ | Iteration of DIM [it] in (27) | CoS |
| (e) | IO/BD-ID(C-Pl) » *.K$^2$, *.{kpt}$^2$ » PRIORITY | DIM [it] ~ [ik] alternation in (28) | CoS |
| (f) | IO/BD-ID(C-Pl) » *.K$^2$, PRIORITY » *.{kpt}$^2$ | Iteration of DIM [it] until. . . | CoS |
| | then: . . . *.{kpt}$^2$ » PRIORITY | . . .alternation with word-final [ik] in (29) | |
| (g) | IO/BD-ID(C-Pl) » *.K$^2$, *.{kpt}$^2$ | Iteration of DIM [it] in Section 4.3 | CaS |

Together, these representations and rankings provide an explicit, formal account of several properties of Hispano-Romance simple and iterated diminutives, which are given in (13). The same DIM allomorphs [ik] and [it] are present in JS and CoS but under the opposite ordering. The consonants that trigger dissimilation need share only the feature [dorsal] in JS but must be totally identical in CoS. Iterated diminutives show three patterns: only [it] iterates, [it] ~ [ik] alternate, or [it] iterates until alternating with word-final [ik]. In particular, the third pattern of iterated diminutives has remained unacknowledged in the literature until now. Unstressed TE vowels of the base are deleted before DIM [it] and [ik], and the stress of the base is shifted rightward in the diminutive, which must end in either TE [o] or [a].

The present study has delivered a systematic cross-linguistic comparison and analysis of diminutive formation, which would have been impossible without the empirical foundations laid by previous descriptive studies of JS (Bunis 2003) and CoS (Fontanella 1962). We have proposed to understand CaS, JS, and CoS as generative grammars in which phonological representations are optimized by specific rankings of universal, violable constraints. Such an approach makes it possible to identify, in a formally explicit way, both similarities and differences across the three Hispano-Romance varieties, and to provide explanations of these patterns in terms of interacting, violable, surface-oriented constraints. The OT framework hypothesizes that universal constraints belong to all natural human languages and that their different rankings give rise to systematic differences across possible languages and dialects.

We hope to have shown that JS and CoS deserve to have a voice in current phonological theorizing. By analyzing the interaction between morphology and phonology in diminutive suffixation, we have situated JS, CoS, and CaS within a Hispano-Romance typology of consonantal dissimilation patterns. We have also argued that similarity avoidance effects in JS and CoS that have been previously analyzed in terms of the OCP are better understood as locally self-conjoined markedness constraints, in line with the analysis of dissimilation in diminutive suffixation proposed in this article.

**Author Contributions:** Conceptualization, C.J.L. and T.G.B.; methodology, C.J.L. and T.G.B.; formal analysis, T.G.B.; investigation, T.G.B. and C.J.L.; resources, T.G.B. and C.J.L.; data curation, T.G.B.; writing—original draft preparation, C.J.L.; writing—review and editing, T.G.B.; visualization, T.G.B.; supervision, T.G.B.; project administration, T.G.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Drawn from studies in the literature, the data reported in this study are openly available within the previously published sources.

**Acknowledgments:** For helpful comments and constructive criticism, we thank two anonymous reviewers for *Languages*, the journal's editorial team, and the Editor of this Special Issue, Rebeka Campos-Astorkiza. We gratefully acknowledge the first co-author's extended family members, who provided insights into the formation of diminutives in Colombian Spanish, and Sonia Colina, who guided the first co-author in the beginning stages of this project at the University of Arizona. We recognize Robert J. Blake and the late Samuel G. Armistead, who, at the University of California, Davis, in the early 2000s, first encouraged the second co-author to explore the phonology of Judeo-Spanish. The second co-author gave a talk based on an earlier version of the present article at an international workshop on Judeo-Spanish, held in July 2024 at the Johannes Gutenberg University of Mainz, Germany. Special thanks go to the workshop organizers Christoph Gabriel and Jonas Grünke, and for helpful discussion, Javier Caro Reina, Brechtje Post, and Aldina Quintana. Any remaining shortcomings in the present article are the sole responsibility of the two co-authors.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Appendix A

The following data sets contain all of the examples of simple diminutives that we extracted from Bunis's (2003) chapter, including the allomorphs [ik] (37) and [it] (38), along with page numbers:

| (37) | | Context | Diminutive | Base | | |
|---|---|---|---|---|---|---|
| | a. | .{p, b, v, m}_ | *kapika* | *kapa* | 'cap' | 213 |
| | | | *gulubika* | *guluba* | 'pigeon' | 214 |
| | | | *manseviko* | *mansevo* | 'young man' | 214 |
| | | | *tavlika* | *tavla* | 'board' | 213 |
| | | | *livriko* | *livro* | 'book' | 213 |
| | | | *patimiko* | *pátimo* | 'step' | 214 |
| | | | *ramika* | *rama* | 'branch' | 213 |
| | | | *tilimikos* | *tilim* | 'Psalms' | 213 |
| | b. | .{t, d, z, n, ɾ, r, l}_ | *altiko* | *alto* | 'tall' | 213 |
| | | | *banketika* | *banketa* | 'gambling bank' | 214 |
| | | | *chibritiku* | *chibrit* | 'match' | 213 |
| | | | *gatiko* | *gato* | 'cat' | 211 |
| | | | *kamaretika* | *kamareta* | 'bedroom' | 214 |
| | | | *kartika* | *karta* | 'card' | 213 |
| | | | *momentiko* | *momento* | 'moment' | 215 |
| | | | *prezentiko* | *prezente* | 'gift' | 214 |
| | | | *Rutika* | *Rud* | (name) | 207 |
| | | | *salatika* | *salata* | 'salad' | 214 |
| | | | *tantiko* | *tanto* | 'so much' | 213 |
| | | | *vestimyentika* | *vestimyenta* | 'garment' | 214 |
| | | | *adjuntadiku* | *adjuntadu* | 'united' | 214 |
| | | | *apegadika* | *apegada* | 'attached' | 214 |
| | | | *awnadiku* | *awnadu* | 'united' | 214 |
| | | | *demazyadiko* | *demazyado* | 'too much' | 214 |
| | | | *dezmazaladiko* | *dezmazalado* | 'unfortunate' | 214 |
| | | | *Gadiko* | *Gad* | (name) | 207 |
| | | | *gritandiko* | *gritando* | 'shouting' | 215 |
| | | | *intindidiku* | *intindidu* | 'intelligent' | 214 |
| | | | *kriadiku* | *kriadu* | 'child' | 214 |
| | | | *kuydadiko* | *kuydado* | 'be careful!' | 214 |
| | | | *ladiko* | *lado* | 'side' | 213 |
| | | | *pekadiko* | *pekado* | 'sin' | 214 |
| | | | *Pedriko* | *Pedro* | (name) | 213 |
| | | | *vestidiko* | *vestido* | 'suit' | 214 |
| | | | *ermoziko* | *ermozo* | 'beautiful' | 214 |
| | | | *mezika* | *meza* | 'table' | 213 |
| | | | *Mwiziko* | *Mwís/-z* | (name) | 213 |
| | | | *rozika* | *roza* | 'rose' | 213 |
| | | | *chintiyaniko* | *chintiyán* | 'trousers' | 214 |

| | | | | |
|---|---|---|---|---|
| | *Daniko* | *Dan* | (name) | 207 |
| | *Djwaniko* | *Djwan* | (name) | 207 |
| | *de manyanika* | *demanyana* | 'in the morning' | 215 |
| | *en vaniko* | *en vano* | 'in vain' | 215 |
| | *ermaniko* | *ermano* | 'brother' | 214 |
| | *findjaniko* | *findján* | 'small cup' | 213 |
| | *Haniko* | *Hané* | (name) | 210 |
| | *kadenika* | *kadena* | 'chain' | 214 |
| | *koronika* | *korona* | 'crown' | 214 |
| | *Reynika* | *Reyna* | (name) | 213 |
| | *shadrivaniko* | *shadriván* | 'fountain' | 214 |
| | *tempraniko* | *temprano* | 'early' | 215 |
| | *batiriko* | *batir* | 'beating' | 214 |
| | *devagariko* | *devagar* | 'slowly, softly' | 214 |
| | *Ezriko* | *Ezrá* | (name) | 208 |
| | *folariko* | *folar* | 'egg pastry' | 213 |
| | *ger(r)eriko* | *ger(r)ero* | 'warrior' | 214 |
| | *hamoriko* | *hamor* | 'donkey' | 213 |
| | *kriaturika* | *kriatura* | 'child' | 214 |
| | *lugariko* | *lugar* | 'place' | 213 |
| | *mezameriko* | *mezamer* | 'asstistant cantor' | 214 |
| | *moriko* | *moro* | 'Moor' | 213 |
| | *orika* | *ora* | 'hour' | 213 |
| | *pashariko* | *pásharo* | 'bird' | 214 |
| | *Pyeriko* | *Pyer* | (name) | 207 |
| | *Sarika* | *Sará* | (name) | 208 |
| | *shubarika* | *shubara* | 'fur cap' | 214 |
| | *sinyoriko* | *sinyor* | 'sir' | 213 |
| | *Sterika* | *(E)ster* | (name) | 207 |
| | *karriko* | *karro* | 'wagon' | 213 |
| | *perriko* | *perro* | 'dog' | 221 |
| | *arvoliko* | *árvole* | 'tree' | 214 |
| | *bimweliko* | *bimwelo* | 'fritter' | 214 |
| | *Eliko* | *Elí* | (name) | 210 |
| | *hamaliko* | *hamal* | 'porter' | 213 |
| | *kavdaliko* | *kavdal* | 'capital' | 213 |
| | *kazaliko* | *kazal* | 'village' | 213 |
| | *paliko* | *palo* | 'stick' | 213 |
| | *papeliko* | *papel* | 'paper' | 213 |
| | *Solika* | *Sol* | (name) | 207 |
| | *soliko* | *solo* | 'alone' | 213 |
| c. .{ʧ, ʃ, ʒ, j}_ | *biskochiko* | *biskocho* | 'biscuit' | 214 |
| | *estrechika* | *estrecha* | 'narrow' | 214 |
| | *kolchika* | *kolcha* | 'blanket' | 213 |
| | *bashiko* | *basho* | 'short' | 213 |
| | *kashika* | *kasha* | 'box, drawer' | 213 |
| | *leshikos* | *leshos* | 'far' | 213 |
| | *Moshiko* | *Moshé* | (name) | 210 |
| | *yidishiko* | *yídish* | 'Ashkenazic man' | 213 |
| | *ijiko* | *ijo* | 'son' | 213 |
| | *navajika* | *navaja* | 'razor' | 214 |
| | *ojiko* | *ojo* | 'eye' | 213 |
| | *bo(y)iko* | *boyo* | 'bun' | 213 |
| | *ga(y)iko* | *gayo* | 'rooster' | 213 |
| | *Hananyiko* | *Hananyá* | (name) | 208 |
| | *Sabetayiko* | *Sabetay* | (name) | 209 |
| | *shada(y)iko* | *Shaday* | (God of a charm) | 213 |
| d. .V_ | *Djohaiko* | *Djohá* | (name) | 208 |
| | *eskrivaniika* | *eskrivanía* | 'writing case' | 214 |
| | *Leika* | *Leá* | (name) | 208 |
| | *Yudaiko* | *Yudá* | (name) | 208 |

| | | | | | |
|---|---|---|---|---|---|
| (38) a. | .k_ | *bayrakito* | *bayrak* | 'flag' | 210 |
| | | *bishlikitu* | *bishlik* | 'five-piastre coin' | 211 |
| | | *blankito* | *blanko* | 'white' | 211 |
| | | *bokita* | *boka* | 'mouth' | 211 |
| | | *burakito* | *burako* | 'hole' | 210 |
| | | *burikita* | *bur(r)eka* | 'filled pastry' | 211 |
| | | *chorekito* | *chorek* | 'round loaf' | 210 |
| | | *dikdukito* | *dikduk* | 'grammar book' | 211 |
| | | *erkyekito* | *erkyek* | 'male' | 211 |
| | | *(E)s.hakito* | *(E)s.hak* | (name) | 211 |
| | | *fis(h)kita* | *fis(h)ka* | 'pimple' | 211 |
| | | *frankito* | *franko* | 'Western European' | 211 |
| | | *freskito* | *fresko* | 'fresh' | 211 |
| | | *hendekito* | *hendek* | 'moat' | 211 |
| | | *Jakito* | *Jak* | (name) | 207 |
| | | *kantikita* | *kantika* | 'song' | 210 |
| | | *kapakito* | *kapak* | 'lid' | 210 |
| | | *kashkita* | *kashka* | 'skin, peel' | 211 |
| | | *kayikito* | *kayik* | 'rowboat' | 211 |
| | | *librikito* | *(l)ibrik* | 'water ewer' | 210 |
| | | *mankito* | *manko* | 'less' | 211 |
| | | *moshkita* | *moshka* | 'fly' | 211 |
| | | *okita* | *oka* | 'measure of weight' | 211 |
| | | *parlakito* | *parlak* | 'match' | 211 |
| | | *pasukito* | *pasuk* | '(biblical) verse' | 211 |
| | | *Rifkita* | *Rifká* | (name) | 211 |
| | | *sakito* | *sako* | 'sack' | 211 |
| | | *serkita* | *serka* | 'near' | 211 |
| | | *turkito* | *turko* | 'Turk; Muslim' | 211 |
| | | *vakita* | *vaka* | 'cow' | 210 |
| b. | .{g, $g^w$}_ | *albondigita* | *albondiga* | 'meatball' | 210 |
| | | *amigito* | *amigo* | 'friend' | 210 |
| | | *bogito* | *bogo* | 'large, soft bundle' | 210 |
| | | *bragita* | *braga* | 'trousers' | 210 |
| | | *figito* | *figo* | 'fig' | 210 |
| | | *gregito* | *grego* | 'Greek' | 210 |
| | | *hw-/fwegito* | *hw-/fwego* | 'fire' | 210 |
| | | *kantigita* | *kantiga* | 'song' | 210 |
| | | *largito* | *largo* | 'long' | 210 |
| | | *lungito* | *lungo* | 'long' | 210 |
| | | *Megita* | *Meg* | (name) | 207 |
| | | *migita* | *miga* | 'crumb' | 210 |
| | | *minagito* | *minag* | 'custom' | 210 |
| | | *negrito* | *negro* | 'bad' | 210 |
| | | *Ogito* | *Og* | (name) | 207 |
| | | *pligito* | *pligo* | 'sheaf' | 210 |
| | | *talegita* | *talega* | 'prayer article bag' | 210 |
| | | *trigito* | *trigo* | 'wheat' | 210 |
| | | *agwita* | *agwa* | 'water' | 211 |
| | | *fragwita* | *fragwa* | 'building' | 211 |
| | | *lingwita* | *lingwa* | 'tongue' | 211 |
| c. | .x_ | *blahitu* | *blahu* | 'Christian' | 211 |
| | | *djarrahito* | *djarrah* | 'surgeon' | 211 |
| | | *felahito* | *felah* | 'Arab peasant' | 211 |
| | | *grahita* | *graha* | 'bean' | 211 |
| | | *malahito* | *malah* | 'angel' | 211 |
| d. | .w_ | *Elyawito* | *Elyaw* | (name) | 211 |
| | | *Irmyawito* | *Irmyaw* | (name) | 211 |
| | | *Lyawito* | *Lyaw* | (name) | 207 |

## Appendix B

The following data sets contain all of the examples that we extracted from Fontanella's (1962) article, including the allomorphs [it] (39) and [ik] (40) in simple diminutives, as well as iterated diminutives (41), along with page numbers:

| (39) | Context | Diminutive | Base | | |
|---|---|---|---|---|---|
| a. | .{p, b, m}_ | *papito* | *papi* | 'daddy' | 560 |
| | | *Albita* | *Alba* | (name) | 560 |
| | | *climita* | *clima* | 'climate' | 567 |
| | | *mamita* | *mami* | 'mommy' | 560 |
| b. | .{d, s, n, ɾ, l}_ | *arregladita* | *arreglada* | 'fixed, repaired' | 563 |
| | | *Bayardito* | *Bayardo* | (name) | 560 |
| | | *caldito* | *caldo* | 'soup, broth' | 563 |
| | | *cerradita* | *cerrada* | 'closed' | 570 |
| | | *delgadito* | *delgado* | 'thin' | 560 |
| | | *enseguidita* | *enseguida* | 'right away' | 561 |
| | | *mojadito* | *mojado* | 'wet' | 557, 563 |
| | | *nadita* | *nada* | 'nothing' | 570 |
| | | *ordenadito* | *ordenado* | 'organized' | 570 |
| | | *Pedrito* | *Pedro* | (name) | 560 |
| | | *pescadito* | *pescado* | 'fish' | 570 |
| | | *todito* | *todo* | 'all' | 557 |
| | | *almuercito* | *almuerzo* | 'lunch' | 563 |
| | | *Beatricita* | *Beatriz* | (name) | 559 |
| | | *bracito* | *brazo* | 'arm' | 563 |
| | | *confiancita* | *confianza* | 'confidence' | 568 |
| | | *Patricita* | *Patricia* | (name) | 559 |
| | | *Alfonsito* | *Alfonso* | (name) | 560 |
| | | *atrasito* | *atrás* | 'behind' | 566 |
| | | *detrasito* | *detrás* | 'behind' | 561 |
| | | *Gladisita* | *Gladis* | (name) | 560 |
| | | *permisito* | *permiso* | 'pardon' | 566 |
| | | *vasito* | *vaso* | 'cup, glass' | 562 |
| | | *chinita* | *china* | 'girl' | 564 |
| | | *hermanito* | *hermano* | 'brother' | 563 |
| | | *limosnita* | *limosna* | 'handout' | 566 |
| | | *manito* | *mano* (f.) | 'hand' | 564 |
| | | *Marinita* | *Marina* | (name) | 559 |
| | | *personita* | *persona* | 'person' | 568 |
| | | *unito* | *uno* | 'one' | 565 |
| | | *ahorita* | *ahora* | 'now' | 557 |
| | | *curita* | *cura* (m.) | 'priest' | 566 |
| | | *arbolito* | *árbol* | 'tree' | 566 |
| | | *Consuelito* | *Consuelo* (f.) | (name) | 559 |
| | | *Julita* | *Julia* | (name) | 560 |
| | | *Lolita* | *Lola* | (name) | 560 |
| c. | .{ɲ, ʝ}_ | *niñito* | *niño* | 'child' | 557 |
| | | *pequeñito* | *pequeño* | 'little' | 562 |
| | | *Estrellita* | *Estrella* | (name) | 560 |
| | | *Mireyita* | *Mireya* | (name) | 559 |
| d. | .{k, g, x}_ | *Blanquita* | *Blanca* | (name) | 560 |
| | | *cerquita* | *cerca* | 'near' | 557, 569 |
| | | *poquito* | *poco* | 'few' | 566 |
| | | *sequito* | *seco* | 'dry' | 566 |
| | | *alguito* | *algo* | 'something' | 568, 570 |
| | | *amiguito* | *amigo* | 'friend' | 570 |
| | | *Huguito* | *Hugo* | (name) | 559 |
| | | *cajita* | *caja* | 'box' | 570 |
| | | *Jorgito* | *Jorge* | (name) | 560 |
| | | *m'hijita* | *m'hija* | 'my daughter' | 560, 566 |
| | | *m'hijito* | *m'hijo* | 'my son' | 560 |
| | | *ovejita* | *oveja* | 'sheep' | 564 |

| | | | | | |
|---|---|---|---|---|---|
| (40) | .t_ | *abiertico* | *abierto* | 'open' | 566 |
| | | *Aramintica* | *Araminta* | (name) | 560 |
| | | *calientico* | *caliente* | 'hot' | 557, 564 |
| | | *completica* | *completa* | 'complete' | 557 |
| | | *Cristico* | *Cristo* | (name) | 564 |
| | | *enantico* | *enantes* | 'recently' | 557, 561 |
| | | *envueltico* | *envuelto* | 'wrapped, covered' | 557 |
| | | *exactica* | *exacta* | 'exact' | 567 |
| | | *galletica* | *galleta* | 'cookie' | 557, 566 |
| | | *guisantico* | *guisante* | 'pea' | 567 |
| | | *letrica* | *letra* | 'letter' | 564 |
| | | *Martica* | *Marta* | (name) | 559 |
| | | *momentico* | *momento* | 'moment' | 565 |
| | | *tantica* | *tanta* | 'so much' | 566 |
| | | *vetica* | *veta* | 'vein, streak' | 562 |

| (41) | | # of suffixes | Diminutive | Base | | |
|---|---|---|---|---|---|---|
| | a. | 2 | *arribitica* | *arriba* | 'above' | 561 |
| | | | *bajitico* | *bajo* | 'short' | 561 |
| | | | *bellitico* | *bello* | 'beautiful' | 563 |
| | | | *cerquitica* | *cerca* | 'near' | 561, 569 |
| | | | *chiquitico/-a* | *chico/-a* | 'small' | 558, 568 |
| | | | *delgaditica* | *delgada* | 'thin' | 568 |
| | | | *estrechitico/-a* | *estrecho/-a* | 'narrow' | 561, 568 |
| | | | *finitico* | *fino* | 'fine' | 561 |
| | | | *fresquitico* | *fresco* | 'cool' | 569 |
| | | | *pasitico* | *despacio* | 'slowly' | 569 |
| | | | *poquitico* | *poco* | 'few' | 561 |
| | | | *seguiditico* | *seguido* | 'one after another' | 569 |
| | | | *sorbitico* | *sorbo* | 'mouthful' | 563 |
| | | | *toditica* | *toda* | 'all' | 568 |
| | | | *trisitico* | *tris* | 'a tiny bit' | 568 |
| | b. | 3 | *cortiquitica* | *corta* | 'short' | 558, 561, 568 |
| | c. | 4 | *chiquitiquitico* | *chico* | 'small' | 558 |
| | | | *toditiquitica* | *toda* | 'all' | 568 |
| | d. | 5 | *cortiquitiquitico* | *corto* | 'short' | 558 |

## Notes

1. We use the term Judeo-Spanish to denote those varieties of Spanish that have been spoken by the descendants of the Sephardic Jews, or Sephardim, who were forced to leave Spain by the edict of expulsion of King Ferdinand and Queen Isabella in 1492. The Sephardim who resettled in North Africa, Turkey, Greece, and the Balkans have retained many archaic linguistic features from Old Spanish and other Ibero-Romance languages spoken at the time of the expulsion, but internal change and contact with other languages have also produced linguistic innovations. For background on Judeo-Spanish and description of its linguistic characteristics and cross-dialectal variation, see Quintana (2006), Bunis (2008, 2011), and Bradley (2022).

2. Throughout this article, we use orthographic forms to present linguistic examples, sometimes with transcriptions in modern IPA. Following the standard romanized orthography proposed by Moshe Shaul (1979) in the inaugural issue of *Aki Yerushalayim*, our JS examples show the following grapheme-phoneme correspondences: <*dj*> = /ʤ/; <*gw*> = /ɡʷ/; <*h*> = /x/; <*j*> = /ʒ/; <*k*> = /k/; <*ny*> = /n.j/, or /ɲ/ in some varieties; intervocalic <*rr*>, word-intial <*r*> = /r/, or /ɾ/ in some varieties; non-word-initial <*r*> = /ɾ/; <*sh*> = /ʃ/; <*v*> = /v/; <*w*> = /w/; <*y*> = /j/; and <*z*> = /z/. In CoS, <*b*>, <*v*> = /b/; <*z*> = /s/; before <*i,e*>, <*c*> = /s/, <*g*> = /x/, <*gu*> = /ɡ/, <*gü*> = /ɡʷ/ <*qu*> = /k/; before <*a,o,u*>, <*c*> = /k/, <*g*> = /ɡ/; <*h*> is silent; <*j*> = /x/; <*ñ*> = /ɲ/; intervocalic <*rr*>, word-intial <*r*> = /r/; non-word-initial <*r*> = /ɾ/; <*x*> = /ks/; <*y*>, <*ll*> = /ɟ/; and <*z*> = /s/. The same correspondences of CoS apply in CaS, except that /θ/ corresponds to <*z*> and to <*c*> before <*i,e*> and that <*ll*> = /ʎ/ (at least for some older speakers).

3. As a reviewer points out in relation to diminutive suffixes, it is worth mentioning that *-ito/a*, *-illo/a*, and *-ico/a* and their longer allomorphs *-(e)cito/a*, *-(e)cillo/a*, and *-(e)cico/a* display the same distribution. Furthermore, the suffix *-zuelo/a* alternates with *-uelo/a*, whereby the former is attached to consonant-final or *e*-final vowel stems, e.g., *ladrónzuelo* 'petty thief' < *ladrón* 'thief', while the latter is attached to stems with the *-o/a* terminal element, e.g., *muchachuelo* 'youngster' < *muchacho* 'boy'; in fact, it is the same distribution of *-ito/a*, *-cito/a.* We do not attempt to account for the distribution of the longer allomorphs in the present study, whose focus is on consonantal place dissimilation involving *-ito/a* and *-ico/a* (*-iko/a*).

4. The article was re-posted on https://bloglenguaencolombia.blogspot.com/2018/11/diminutivo.html (accessed on 12 December 2024). We translate the most relevant part of the article here. The examples in (10) appear in the final paragraph:"Well, in *caleñol* (as in Caribbean Spanish) we have added a rule to the common use of diminutives: if the word ends in *-t*, *-tr* or *-tl* + vowel, we use *-ico* or *-ica*, instead of *-ito* or *-ita*, as in the rest of the Spanish-speaking world. For example, *zapato* 'shoe' ends in *-to*, and

that's why we say *zapatico; aguardiente* 'moonshine' ends in *-te*, and that's why *aguardientico*; *gato* 'cat' ends in *-to*, and that's why *gatico* and not *gatito; otro* 'other' ends in *-tro*, and that's why *otrico* and not *otrito.* A small *atlas* 'atlas' would be an *atlicas.*For a woman whose name is *Marta*, we say *Martica* and not *Martita*: because *Marta* ends in *-ta.* And that's why we say *momentico* and not *momentito: momento* 'moment' ends in *-to.* The other day we were talking about people who say *llamó estica* 'this one called' to refer to a woman whose name they don't remember. This allows us to explain why we say *estica* and not *estita*.The most interesting thing is that we use this rule to make a diminutive of a diminutive. From *ahora* 'now' we make *ahorita*, which is less time than *ahora.* But if we want to express even less time than *ahorita* we use the rule, because *ahorita* ends in *-ta*, and the word turns out to be *ahoritica.* Likewise, *poquito* is less than *poco* 'small' and *poquitico* much less than *poquito.*The only problem is that *-ica* (or *-ico*) cannot be iterated, but *-ita* can be repeated as many times as we want. So, we say *ahoritititititica* if we want to exaggerate the tiny amount of time, but we would never say *ahoritikikikikikica* (Here I use *k* to indicate the sound [k] in the hypothetical example). [our translation—CJL and TGB]"Díaz Collazos purports to describe diminutive formation in the *caleño* variety spoken in Cali, Colombia. However, the failure of *-ico/a* to iterate (10b) is likely a more general feature of CoS, as well as other dissimilating varieties in Costa Rica, Venezuela, Cuba, and eastern and southern Spain. We leave it to future research to determine the geographical extent of the *caleñol* pattern.

5 In CaS *bebito* < *bebé* 'baby' (11a), the base lacks an unstressed terminal element. Deletion of base-final stressed [é] is exceptional in this case, as base-final stressed vowels usually condition the allomorph *-cito/a*, e.g., *cafecito* < *café* 'coffee', *mamacita* < *mamá* 'mom'. JS personal names ending in stressed [á] (37d) (Appendix A) show variation between maintenance or deletion of the base-final vowel, e.g., *Djohaiko* < *Djohá*, vs. *Leika* < *Leá* (Bunis 2003, p. 208).

6 The goal of an OT analysis is to explain why one representation is mapped onto another representation. Different mappings are evaluated against a partially ranked set of constraints. Output candidates are evaluated by their violations of the constraints whose titles appear along the top row of an OT tableau, proceeding from left to right. The optimal mapping satisfies the highest-ranking constraints. In the tableaux of the present study, the output [base+DIM+TE] appears in the first cell of the top row, and relevant output candidates of the corresponding diminutive appear within the same column in the rows beneath, each denoted by a separate letter (a), (b), etc. Crucial constraint rankings are indicated by a solid line between two columns, while a broken line means that the ranking between the two adjacent constraints is indeterminate. When a diminutive output candidate violates a constraint, an asterisk appears in the corresponding cell beneath that constraint. As indicated by an exclamation point after the relevant asterisks, violations of higher-ranking constraints are fatal and eliminate candidates from the competition for optimality. The single remaining candidate that incurs no fatal violations of the higher-ranking constraints is the winner, as indicated by the manicule '☞'.

7 *blahu* 'Christian' and other such JS words are attested in varieties that have an independent process of unstressed mid vowel raising, which we do not analyze here. In these words, final [u] corresponds to the morphological TE [o] of other Hispano-Romance varieties.

8 Bermúdez-Otero (2006, 2007, 2013) does not attempt to account for the longer DIM allomorph *-ecita* in *puertecita*, which for many CaS speakers is the optimal form instead of *puertita*. Nevertheless, the argument for ordering stem-level diphthongization before word-level diminutivization still holds.

## References


Alber, Birgit, and Sabine Arndt-Lappe. 2022. Anchoring in truncation: A typological analysis. *Natural Language and Linguistic Theory* 41: 1–50. [CrossRef]

Alderete, John. 1997. Dissimilation as local conjunction. In *Proceedings of the North East Linguistic Society 27*. Edited by Kiyomi Kusumoto. Amherst: GLSA Publications, pp. 17–32.

Alvar, Manuel, and Bernard Pottier. 1983. *Morfología histórica del español*. Madrid: Gredos.

Ambadiang, Théophile. 1996. La formación de diminutivos en español: ¿fonología o morfología? *Lingüística Española Actual* 18: 175–211.

Ambadiang, Théophile, and Bruno Camus Bergareche. 2012. Morfofonología de la formación de diminutivos en español: ¿reglas morfológicas o restricciones fonológicas? In *Los Límites de la Morfología: Estudios Ofrecidos a Soledad Varela Ortega*. Edited by Antonio Fábregas Alfaro, Elena Felíu Arquiola and José Pazó Espinosa. Madrid: UAM, pp. 55–78.

Benua, Laura. 1997. Transderivational Identity: Phonological Relations Bewteen Words. Ph.D. thesis, University of Massachusetts Amherst, Amherst, MA, USA.

Bermúdez-Otero, Ricardo. 2006. Morphological structure and phonological domains in Spanish denominal derivation. In *Optimality-Theoretic Studies in Spanish Phonology*. Edited by Fernando Martínez-Gil and Sonia Colina. Amsterdam: John Benjamins, pp. 278–311. [CrossRef]

Bermúdez-Otero, Ricardo. 2007. Spanish pseudoplurals: Phonological cues in the acquisition of a syntax-morphology mismatch. In *Deponency and Morphological Mismatches, Proceedings of the British Academy Volume 145*. Edited by Matthew Baerman, Greville G. Corbett, Dunstan Brown and Andrew Hippisley. Oxford: Oxford University Press, pp. 231–69. [CrossRef]

Bermúdez-Otero, Ricardo. 2013. The Spanish lexicon stores stems with theme vowels, not roots with inflectional class features. *Probus* 25: 3–103. [CrossRef]

Bonet, Eulàlia, Maria-Rosa Lloret, and Joan Mascaró. 2007. Allomorph selection and lexical preferences: Two case studies. *Lingua* 117: 903–27. [CrossRef]

Bradley, Travis G. 2022. Judeo-Spanish. In *Manual of Romance Phonology and Phonetics*. Edited by Christoph Gabriel, Randall Gess and Trudel Meisenburg. Berlin and Boston: De Gruyter, pp. 808–38. [CrossRef]

Bradley, Travis G., and Jason Smith. 2011. The phonology-morphology interface in Judeo-Spanish diminutive formation: A lexical ordering and subcategorization approach. *Studies in Hispanic and Lusophone Linguistics* 4: 247–300. [CrossRef]

Bunis, David. 2003. Ottoman Judezmo diminutives and other hypocoristics. In *Linguistique des langues juives et linguistique générale*. Edited by Frank Alvarez-Péreyre and Jean Baumgarten. Paris: CNRS Éditions, pp. 193–246.

Bunis, David. 2008. The differential impact of Arabic on Ḥaketía and Turkish on Judezmo. In *El presente. Estudios sobre la cultura sefardí, vol. 2: La cultura Judeo-Española del Norte de Marruecos*. Edited by Tamar Alexander and Yaakov Bentolila. Beersheba: Universidad Ben-Gurion del Negev, Sentro Moshe David Gaon de Kultura Djudeo-Espanyola, pp. 177–207.

Bunis, David. 2011. Judezmo. The Jewish language of the Ottoman Sephardim. *European Judaism* 44: 22–35. [CrossRef]

Camus Bergareche, Bruno. 2018. Aspectos de la evolución histórica de la alomorfia en los diminutivos españoles. *Estudios de Lingüística del Español* 39: 107–24. [CrossRef]

Casali, Roderic F. 1997. Vowel elision in hiatus contexts: Which vowel goes? *Language* 73: 493–533. [CrossRef]

Casali, Roderic F. 2011. Hiatus resolution. In *The Blackwell Companion to Phonology, 5 vols*. Edited by Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume and Keren Rice. Malden: Wiley-Blackwell. [CrossRef]

Chomsky, Noam, and Morris Halle. 1968. *The Sound Pattern of English*. New York: Harper and Row. Available online: https://web.mit.edu/morrishalle/pubworks/papers/1968_Chomsky_Halle_The_Sound_Pattern_of_English.pdf (accessed on 12 November 2024).

Colina, Sonia. 2003. Diminutives in Spanish: A morpho-phonological account. *Southwest Journal of Linguistics* 22: 45–88.

Colina, Sonia. 2009. *Spanish Phonology: A Syllabic Perspective*. Washington, DC: Georgetown University Press.

Crosswhite, Katherine. 1998. Segmental vs. prosodic correspondence in Chamorro. *Phonology* 15: 281–316. [CrossRef]

Crowhurst, Megan. 1992. Diminutives and augmentatives in Mexican Spanish: A prosodic analysis. *Phonology* 9: 221–53. [CrossRef]

Davies, Mark. 2016–. Corpus del Español: Web/Dialects. Available online: https://www.corpusdelespanol.org/web-dial/ (accessed on 12 November 2024).

de Lacy, Paul. 2006. *Markedness: Reduction and Preservation in Phonology*. Cambridge: Cambridge University Press.

Eguren, Luis. 2001. Evaluative suffixation in Spanish and the syntax of derivational processes. In *Features and Interfaces in Romance: Essays in Honor of Heles Contreras*. Edited by Julia Herschensohn, Enrique Mallén and Karen Zagona. Amsterdam: John Benjamins, pp. 71–84. [CrossRef]

Fábregas, Antonio. 2013. Diminutives as heads or specifiers: The mapping between syntax and phonology. *Iberia: An International Journal of Theoretical Linguistics* 5: 1–44. Available online: https://idus.us.es/handle//11441/80052 (accessed on 12 December 2024).

Fontanella, María Beatriz. 1962. Algunas observaciones sobre el diminutivo en Bogotá. *Thesaurus* 17: 556–73.

Goldsmith, John. 1976. Autosegmental Phonology. Ph.D. thesis, MIT, Boston, MA, USA.

González Ollé, F. 1962. *Los Sufijos Diminutivos en Castellano Medieval*. Madrid: Consejo Superior de Investigaciones Científicas [*Revista de Filología Española* (75)].

Harris, James. 1994. The OCP, prosodic morphology and Sonoran Spanish diminutives: A reply to Crowhurst. *Phonology* 11: 179–90. [CrossRef]

Hayes, Bruce. 2009. *Introductory Phonology*. Malden: Wiley-Blackwell.

Ito, Junko, and Armin Mester. 1998. Markedness and word structure: OCP effects in Japanese. Unpublished manuscript, Available on Rutgers Optimality Archive, ROA-255.

Ito, Junko, and Armin Mester. 2003. On the sources of opacity in OT: Coda processes in German. In *The Syllable in Optimality Theory*. Edited by Caroline Féry and Ruben van de Vijver. Cambridge: Cambridge University Press, pp. 271–303, Available on Rutgers Optimality Archive, ROA-347.

Jaeggli, Osvaldo. 1980. Spanish diminutives. In *Contemporary studies in Romance languages*. Edited by Frank H. Nuessel. Bloomington: Indiana University Linguistic Club, pp. 142–58.

Kager, René. 1999. *Optimality Theory*. Cambridge: Cambridge University Press.

Kany, Charles. 1960. *American-Spanish Semantics*. Berkeley and Los Angeles: University of California Press.

Lázaro Mora, Fernando. 1999. La derivación apreciativa. In *Gramática descriptiva de la lengua Española*. Edited by Ignacio Bosque and Violeta Demonte. Madrid: Espasa Calpe, pp. 4645–82.

Leben, William. 1973. Suprasegmental Phonology. Ph.D. thesis, MIT, Boston, MA, USA.

Lipski, John. 1999. El sufijo -ico y las palabras afroibéricas agüé/awe y aguora/ahuora: Rutas de evolución y entorno dialectológico. In *El Caribe Hispánico: Perspectivas Lingüísticas Actuales*. Edited by Ortiz López. Frankfurt: Vervuert, pp. 17–42.

Mascaró, Joan. 2007. External allomorphy and lexical representation. *Linguistic Inquiry* 38: 715–35. [CrossRef]

McCarthy, John. 1986. OCP effects: Gemination and antigemination. *Linguistic Inquiry* 17: 207–63.

McCarthy, John. 2003. OT Constraints are categorical. *Phonology* 2003: 75–138. [CrossRef]

McCarthy, John. 2008. *Doing Optimality Theory: Applying Theory to Data*. Malden: Blackwell Publishing.

McCarthy, John, and Alan Prince. 1999. Faithfulness and identity in Prosodic Morphology. In *The Prosody Morphology Interface*. Edited by René Kager, Harry van der Hulst and Wim Zonneveld. Cambridge: Cambridge University Press, pp. 218–309.

McCarvel, Miranda. 2016. Harmonic Serialism with Lexical Selection: Evidence from Jèrrias Allomorphy. Ph.D. thesis, University of Utah, Salt Lake City, UT, USA.

Miranda, Inés Miranda. 1999. An Optimality Theoretic Analysis of Nicaraguan Spanish Diminutivization: Results of a Field Survey. Ph.D. thesis, University of Washington, Seattle, WC, USA.

Nebrija, Antonio de. 1492. Gramática castellana. Available online: https://www.bne.es/es/Micrositios/Guias/12Octubre/Lenguas/Castellano/ (accessed on 12 December 2024).

Nehama, Joseph. 1977. *Dictionnaire du judéo-espagnol*. Madrid: Consejo superior de investigaciones científicas, Instituto Benito Arias Montano.

Nieuwenhuis, Paul. 1985. Diminutives. Ph.D. thesis, University of Edinburgh, Edinburgh, UK.

Normann-Vigil, Ingrid. 2012. Accounting for variation of diminutive formation in Porteño Spanish. *Mester* 41: 99–122. Available online: https://escholarship.org/uc/item/173487vs (accessed on 12 December 2024). [CrossRef]

Ohannesian, María. 1996. La formación del diminutivo en castellano. Master's thesis, Universitat Autònoma de Barcelona, Spain.

Ohannesian, María. 2020. Allomorphic variation. In *The Routledge Handbook of Spanish Phonology*. Edited by Sonia Colina and Ferndando Martínez-Gil. New York: Routledge, pp. 288–306. [CrossRef]

Ohannesian, María, and Clàudia Pons. 2009. Shattering paradigms: An attempt to formalize pressures within subparadigms. In *Selected Proceedings of the 6th Décembrettes*. Edited by Fabio Montermini, Gilles Boyé and Jesse Tseng. Somerville: Cascadilla Proceedings Project, pp. 76–94. Available online: https://www.lingref.com/cpp/decemb/6/abstract2237.html (accessed on 12 December 2024).

Penny, Ralph. 2014. *Gramática histórica del español: Edición actualizada*. Barcelona: Ariel Letras.

Prieto, Pilar. 1992. Morphophonology of the Spanish diminutive formation: A case for prosodic sensitivity. *Hispanic Linguistics* 5: 169–205.

Prieto, Victor Moisés. 2005. Spanish Evaluative Morphology: Pragmantic, Sociolinguistics, and Semantic Issues. Ph.D. thesis, University of Florida, Gaineseville, FL, USA.

Prince, Alan, and Paul Smolensky. 1993/2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Malden and Oxford: Blackwell, Revision of 1993 technical report, Rutgers University Center for Cognitive Science. Available on Rutgers Optimality Archive, ROA-537.

Quintana, Aldina. 2006. *Geografía lingüística del judeoespañol*. Bern: Lang.

Ramsammy, Michael. 2017. The phonology-phonetics interface in constraint-based grammar. In *The Routledge Handbook of Phonological Theory*. Edited by S. J. Hannahs and Anna Bosch. New York: Routledge, pp. 68–99. [CrossRef]

Shaul, Moshe. 1979. Es ke ay menester de una nueva ortografia para el djudeo-espaniol. *Aki Yerushalayim* 1: 3–4.

Smith, Jason. 2011. Subcategorization and Optimality Theory: The Case of Spanish Diminutives. Ph.D. thesis, University of Claifornia, Davis, CA, USA.

Steriade, Donca. 1999. Lexical conservatism in French adjectival liason. In *Formal Perspectives on Romance Linguistics*. Edited by J.-Marc Authier, Barbara Bullock and Lisa Reid. Amsterdam: John Benjamins, pp. 243–70. [CrossRef]

Steriade, Donca. 2000. Paradigm uniformity and the phonetics-phonology interface. In *Acquisition and the Lexicon*. (Papers in Laboratory Phonology 5). Cambridge: Cambridge University Press, pp. 313–34.

Tarazona, Edinson. 2021. Los sufijos aumentativos y la reduplicación en español: El caso de *-ón* y *-ote*. *Cuaderno de Lingüística Hispánica* 37: e12584. [CrossRef]

Vadella, Katherine. 2017. The Morphosyntax of Gender and Word Class in Spanish: Evidence from *-(c)ito/a* Diminutives. Ph.D. thesis, Georgetown University, Washington, DC, USA.

Zacarías, Ramón. 2006. Formación de diminutivos con el sufijo /-ít-/. Una propuesta desde la morfología natural. *Anuario De Letras. Lingüística Y Filología* 44: 77–103. Available online: https://revistas-filologicas.unam.mx/anuario-letras/index.php/al/article/view/1019 (accessed on 12 December 2024).

*languages*

MDPI

*Article*

# The Role of (Re)Syllabification on Coarticulatory Nasalization: Aerodynamic Evidence from Spanish

**Ander Beristain**

Department of Languages, Literatures and Cultures, Saint Louis University, St. Louis, MO 63108, USA; ander.beristain@slu.edu

**Abstract:** Tautosyllabic segment sequences exhibit greater gestural overlap than heterosyllabic ones. In Spanish, it is presumed that word-final consonants followed by a word-initial vowel undergo resyllabification, and generative phonology assumes that canonical **C**V.**C**V# and derived CV.**C**#V onsets are structurally identical. However, recent studies have not found evidence of this structural similarity in the acoustics. The current goal is to investigate anticipatory and carryover vowel nasalization patterns in tautosyllabic, heterosyllabic, and resyllabified segment sequences in Spanish. Nine native speakers of Peninsular Spanish participated in a read-aloud task. Nasal airflow data were extracted using pressure transducers connected to a vented mask. Each participant produced forty target tokens with CV.CV# (control), CVN# (tautosyllabic), CV.NV# (heterosyllabic), and CV.N#V (resyllabification) structures. Forty timepoints were obtained from each vowel to observe airflow dynamics, resulting in a total of 25,200 datapoints analyzed. Regarding anticipatory vowel nasalization, the C**V**N# sequence shows an earlier onset of nasalization, while C**V**.NV# and C**V**.N#V sequences illustrate parallel patterns among them. Carryover vowel nasalization exhibited greater nasal spreading than anticipatory nasalization, and vowels in CV.N**V**# and CV.N#**V** structures showed symmetrical nasalization patterns. These results imply that syllable structure affects nasal gestural overlap and that aerodynamic characteristics of vowels are unaffected across word boundaries.

**Keywords:** coarticulatory nasalization; resyllabification; aerodynamics; Spanish

## 1. Introduction

Traditionally, generative phonology assumes that, in connected speech in Spanish, word-final consonants undergo resyllabification before a word that starts with a vowel, e.g., *un amigo* [u.n#a.mi.ɣo] 'a friend' (Harris and Kaisse 1999) (henceforth, 'C': consonant, 'V': vowel, 'N': nasal consonant, '.': syllable boundary, '#': word boundary). The process is understood as complete and a part of the phonology of Spanish. Nevertheless, recent laboratory studies have questioned the status of 'obligatory' or 'complete' resyllabification of C#V sequences in Spanish by providing evidence for acoustic and phonetic differences between 'canonical' onsets, e.g., **C**V.**C**V#, and 'derived' (resyllabified) onsets, e.g., CV.**C**#V (Bradley et al. 2022; Hualde and Prieto 2014; Jiménez-Bravo and Lahoz-Bengoechea 2023; Strycharczuk and Kohlberger 2016).

To date, most of the studies investigating resyllabification have focused on the consonantal segments, which were, in most cases, oral fricative sounds, disregarding how neighboring vowels or other types of segments such as nasal structures might be affected by (re)syllabification rules. Furthermore, research has shown that syllabic structure plays a role in nasalization degree (Byrd et al. 2009; Cohn 1993); thus, tautosyllabic VN segments will show greater gestural overlap than heterosyllabic V.N segments (see Krakow 1989, 1993, 1999).

Remarkably, the effect of (re)syllabification of nasal consonants on neighboring vowels in Spanish has not been thoroughly investigated, especially via articulatory methods. The present study seeks to understand the aerodynamics (Beristain 2022, 2023a, 2023b; Cohn 1993; Huffman and Krakow 1993; Shosted 2009; Shosted et al. 2012; Solé 1992;

Stoakes et al. 2020) of the spread of anticipatory and carryover coarticulatory nasalization across various syllabic contexts within word boundaries and across word junctures. The study determines that tautosyllabic segments show greater articulatory overlap and that carryover nasalization is unaffected by resyllabification, which provides evidence for complete resyllabification.

The structure in the current article is as follows: first, the relevant literature on syllable structure, resyllabification, and nasalization is provided; second, research questions and hypotheses are presented; third, the methods are included; fourth, the results pertaining to anticipatory and carryover nasalization are summarized, followed by the discussion and conclusions.

## 2. Literature Review

### 2.1. Resyllabification Revisited

Resyllabification is a process by which a word-final coda delinks from its original syllable structure and attaches to the following vowel-initial word as an onset. It should be noted that this process is not universal across languages and/or dialects. For instance, English has ambisyllabicity (Kahn 1976) or glottal stop insertion (Bissiri et al. 2011), German has glottal stop insertion (Kohler 1994), and eastern Abruzzese dialects opt for voiced glottal fricative [ɦ] or velar fricative [ɣ] as an empty onset repair mechanism (Passino et al. 2022, p. 93). On the other hand, resyllabification is found in Spanish (Harris 1983), French (Durand et al. 2011), and dialects of Occitan (Sauzet 2012), among others. Harris and Kaisse (1999) explain that for a segment to be resyllabified, three stages need to happen: (i) initial syllabification (no change); (ii) delinking (change initiates); and (iii) attach onset (resyllabification occurs, change is complete). Below is Figure 1, which illustrates the derivation of resyllabification for the word sequence *un año* 'a year' in Spanish (adapted from Harris and Kaisse 1999, p. 137).

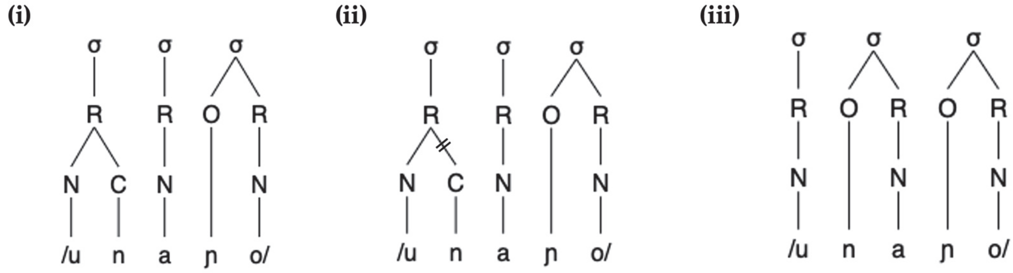



**Figure 1.** Derivation showing (re)syllabification in word sequence *un año* 'a year' in Spanish [where (**i**) initial syllabification; (**ii**) delinking; (**iii**) attach onset (resyllabification)].

In this view, resyllabified codas should behave like onsets. Remarkably, weakening phenomena that affect word-internal coda consonants may also affect word-final consonants, even if they are resyllabified as onsets preceding a vowel in the following word. These cases have been modeled using either rule ordering (Harris 1983) or constraint ranking within optimality theory (Colina 1995, 1997, 2009) approaches. In Spanish varieties where /s/ reduces to [h] in the coda (but not in word-internal position), word-final /s/ is aspirated before resyllabification occurs in Vs#V structures, suggesting that a sequence such as *los amigos* /los amigos/ 'the friends' would undergo [loh. a.mi.ɣoh → lo.ha.mi.ɣoh] versus VsV structures, e.g., *losa* /losa/ 'slab' → *[lo.ha] (Colina 2002; Harris 1983; Kaisse 1999). Nevertheless, as pointed out and illustrated by a reviewer, in more conservative varieties, the sibilant [s] is retained. Brown (2008) states that lower levels of word-final [s] retention correlate with high-frequency words, unless word-strings are taken into consideration and the following word starts with a stressed vowel, in which case the relation is inversed. For instance, Brown (2008, pp. 200–1) indicates that, because of the higher frequency of the word combination *dos años* 'two years', this word-string is stored as a single unit in memory, i.e., /dosaɲos/. This results in word-final /s/ being reanalyzed as word-medial, and being less likely to undergo reduction. On the other hand, the author points out that a low-frequency word-string such as *dos asnos* 'two donkeys' is stored as two separate words, i.e., /dos asnos/. As such, the word-final /s/ is being processed as word-final, and it is more likely to be reduced (Brown 2008, pp. 201–2). As far as nasal segments are concerned, in velarizing dialects, word-final /n/ is velarized to [ŋ] in Vn#V structures, e.g., *van a casa* /ban a kasa/ 'they go home' we find [baŋ. a. ka.sa → ba.ŋa.ka.sa] versus in VnV structures, e.g., *vana* /bana/ 'vain' → *[ba.ŋa] (Harris 1983; Robinson 2012).

The generative literature assumes that the syllabic structure of the resyllabified **C**#V segments (also termed 'derived onsets') and 'canonical' **C**V onsets are identical. From the perspective of connected speech, such an assumption could be well received. However, the postulate of structure similarity has been questioned, as will be shown in the next paragraph. It should be noted that the number of studies that have explored word-internal and across-word syllabic structure implementing articulatory approaches is scarce (see Byrd et al. 2009; Krakow 1999, and references therein), even though those approaches provide a more reliable and less ambiguous description of linguistic patterns and speech gestures than acoustic studies.

In recent years, there has been an effort by phoneticians to identify the acoustic correlates of resyllabified onsets. The language under study in this paper is Spanish, which shows phonotactic restrictions as to which consonants can appear in the coda position, those being /n, l, ɾ, d, s, θ/ (Hualde 2014, p. 62). Most recent studies on the acoustic correlates of resyllabified consonantal segments have focused on the sibilant fricative /s/. As will be explained, word-final /s/ is shorter and more voiced in the coda position; as such, it is expected that resyllabified onsets will have longer durations. Hualde and Prieto (2014) analyzed spontaneous acoustic data of /s/ in Madrid Spanish and investigated uninterrupted voicing, voicing frames, and duration of that segment in word-initial, medial, and final positions. Their results showed that intervocalic word-final position (i.e., resyllabification) led to higher rates of fully voiced /s/ and shorter durations, indicating differences from 'canonical' conditions (see Torreira and Ernestus 2012 for similar results). Furthermore, Strycharczuk and Kohlberger (2016) analyzed the fricative /s/ in northern and central Peninsular Spanish and compared the effects of the position within the syllable and word on the acoustic properties of the segment. They found that word-final, derived onset consonants were shorter than canonical ones. Strycharczuk and Kohlberger (2016) is among the few studies that also provided an acoustic description of neighboring vowels. Unlike for the duration of /s/, the researchers found evidence for complete resyllabification in terms of vowel durations. Strycharczuk and Kohlberger (2016, p. 11) encountered that the duration of vowels was not significantly different between word-initial onsets and derived onsets (the latter were 0.09 ms longer) but it was significantly greater compared to word-medial codas (the latter were 17.28 ms shorter) and word-medial onsets (the latter were 11.43 ms shorter). Moreover, Jiménez-Bravo and Lahoz-Bengoechea (2023) conducted an acoustic study where they compared canonical /s, n, l/ onsets, e.g., *vende naves* '(s/he) sells ships' and derived ones *venden aves* '(they) sell birds'. Similar to previous studies, their results illustrate that derived (resyllabified) onsets were shorter than canonical ones in duration, respectively; /s/: 97.6 vs. 90.6 ms, /n/: 53.1 vs. 50.2 ms, and /l/: 63.5 vs. 58.7 ms, which provides additional evidence for incomplete resyllabification patterns in Spanish.

In the case of word-final /n/ in Spanish, an allophonic variation in certain varieties is its velarization, i.e., [ŋ] (see Bongiovanni 2021). The velarization of word final /n/ has been attested in the north-west and south of Spain, the Canary Islands, Caribbean varieties, and certain Spanish varieties in South America (Hualde 2014, p. 174). Robinson (2012) provided impressionistic data from Ecuadorian Spanish, specifically from Quito and Cuenca, where participants were asked to resyllabify words that contained a final /n/ in a velarizing context across word junctures, and all participants exhibited a significant pause between words. This would indicate that resyllabification rules might not be applicable universally in Spanish, as the ‘expected’ syllabic production of the word was misaligned in connected speech.

*2.2. Syllable Structure and Vocalic Nasalization*

Coarticulatory vowel nasalization is the gestural coordination between a vowel (V) next to a nasal consonant (N). The effects of coarticulation can spread from left to right (carryover), e.g., *no* /no/ [nõ] ‘no’ in Spanish, or right to left (anticipatory), e.g., *en* [ẽn] ‘in’ in Spanish. The outcome in both conditions is gestural overlap with partial phonetic nasalization of the vowel. Cross-linguistic differences have been found with regard to the degree of gestural overlap across carryover and anticipatory, and within each type of process (see Beristain 2023a, 2023b; Clumeck 1976; Goodin-Mayeda 2016; Martínez 2021).

Previous studies have analyzed the effect of syllable structure on the nasalization of the vowel (Beristain 2022; Byrd et al. 2009), and the general resolution that they have found is that nasalization develops earlier in vowels in tautosyllabic VN sequences than in heterosyllabic V.N ones because of gestural timing differences (Krakow 1989, 1999). For instance, Diakoumakou (2005) analyzed coarticulatory vowel nasalization in Modern Greek and compared it to Spanish, Chinese, Japanese, Thai, Ikalanga, French, Italian, American English, Hindi, and Bengali. As far as Greek is concerned, she provides acoustic evidence to point out that the differences found in the nasalization patterns of vowels in tautosyllabic or heterosyllabic environments is due to different articulatory patterns found in syllable-initial and syllable-final contexts, with the latter having a lower velum position than the former. Diakoumakou (2009) shows that only the final 17% of the vowel was nasalized in heterosyllabic V.N sequences, whereas 33% of the vowel was nasalized in tautosyllabic VN ones. Similarly, Cohn (1990) finds that the onset of nasalization in tautosyllabic VN sequences is earlier than in those of heterosyllabic V.N ones in French. Moreover, Krakow (1999, p. 27) analyzed the effect of syllable structure on nasalization patterns in English, comparing word sets such as *see more* vs. *seam ore* and *pa made* vs. *palm aid* (i.e., CV#NV vs. CVN#V). She used the Velotrace (Horiguchi and Bell-Berti 1987), which collected velum raising and lowering movements with an LED attached to the lower lip with the Selspot System to capture labial aperture and closure during the bilabial nasal /m/ and contiguous vowels. Similarly to Cohn (1990), Krakow found that, in carefully read speech, tautosyllabic VN sequences show an earlier onset of nasalization than heterosyllabic V.N ones. Under similar contexts, Lahoz-Bengoechea and Jiménez-Bravo (forthcoming) compared the degree of nasalization in VN#V and V#NV sequences using Nasalization from Acoustic Features (NAF) measurements (see Carignan 2021 for a methodological overview). Their acoustic data came from 19 individuals from Spain whose first language was Spanish. The authors found no significant differences in the degree of nasal coarticulation between the two contexts: VN#V: 0.714; V#NV: 0.713. These results provide evidence in support of complete resyllabification in connected speech.

We may surmise that an earlier onset of nasalization in tautosyllabic segments is correlated with the development of nasal vowels in Romance languages. As Sampson (1999, p. 35) points out, “the process of vowel nasalization has often been significantly affected by whether the conditioning nasal is tautosyllabic or heterosyllabic, reflecting the importance of the syllable as a structural unit in the diachronic and synchronic phonological patterning of Romance.” Notice how vowels in VN tautosyllabic sequences developed and eventually became nasal vowels in some languages, e.g., BONU ‘good, masc.’ > [bon] > [bõn] > [bõ] (in Modern French), whereas vowels in heterosyllabic sequences showed more coarticulatory resistance, in some cases even leading to denasalization, e.g., BONA ‘good, fem.’ > [bona] in Northern Italian (Hajek 1997, p. 10).

Diakoumakou (2005) points out that languages with a tendency for open syllables show greater carryover nasalization, while languages with a preference for close syllables have more extensive anticipatory nasalization. She conducted an acoustic study investigating the temporal extent of vowel nasalization in Modern Greek. She obtained data from six native speakers and found that the temporal extent of nasalization was 27 ms long in the heterosyllabic anticipatory condition, 48 ms long for the tautosyllabic anticipatory condition, and 70 ms for the carryover condition. Diakoumakou discusses these results, comparing them to what has been found in other languages, and explains that the tendency for open syllables, such as in Greek, Spanish, or Italian, seems to be conducive for greater carryover nasalization. On the other hand, she mentions that languages that show a preference for closed syllables show greater anticipatory nasalization (see results for English in Krakow (1999)).

When comparing anticipatory versus carryover nasalization patterns, Cohn (1990) provides airflow contours for various contexts that include tautosyllabic VN and heterosyllabic V.NV and V#NV sequences within a word and across word junctures in French. Her results indicate that there is greater carryover nasalization (e.g., *nez* 'nose', p. 123) than anticipatory nasalization (e.g., *bonne* 'good', p. 97). Similar aerodynamic results were found by Delvaux et al. (2009). In Cohn (1990), airflow contours in derived onsets, e.g., *bonne ode* 'good ode' (p. 123) vs. canonical onsets, e.g., *beau nez* 'beautiful nose' (p. 101) do not exhibit significant differences in French. Both conditions show a cline-like rise that appears late in the vocalic segment and then a rapid drop after the offset of the nasal consonant, which is followed by a constant and smooth transition throughout the nasalized segment. This indicates that articulatory correlates are similar across heterosyllabic V#NV and resyllabification V.N#V structures.

## 3. Research Questions

A vast amount of research about resyllabification has focused on the properties of the consonant that changes its syllabic linking. The features associated with such consonants were usually [+oral][-nasal]. Evidence has been provided for sub-phonemic differences between canonical and derived onsets (see Bradley et al. 2022). The role that resyllabification may have on the vowels surrounding the resyllabified consonant has been less investigated, especially in the context of coarticulatory nasalization. Studies on gestural timing and nasalization point out that the syllable structure of contiguous segments plays a significant role, as tautosyllabic segments will show greater gestural overlap than heterosyllabic ones (Krakow 1989; Byrd et al. 2009). Moreover, Diakoumakou (2009) and Delvaux et al. (2009) state that languages that show a tendency towards open syllables usually exhibit greater carryover coarticulatory nasalization than anticipatory nasalization.

Based on the previous findings, the specific research questions (RQs) and hypotheses that the present study will investigate are as follows:

- *RQ 1:* Are there differences between nasal airflow contours in the degree of phonetic implementation of carryover and anticipatory coarticulatory nasalization in Spanish?

**Hypothesis 1:** *Yes. Considering that Spanish shows a tendency for open syllables (Hualde 2014, p. 59), according to Cohn (1990), Diakoumakou (2009), and Delvaux et al. (2009), it is expected that carryover vocalic nasalization should show greater nasal spreading than anticipatory nasalization.*

- *RQ 2:* Does syllable structure play a role in the degree of anticipatory coarticulatory nasality in Spanish?

**Hypothesis 2:** *Yes. Based on Krakow (1989, 1999) and Cohn (1990), it is expected that vowels in tautosyllabic CVN# segment sequences will show an earlier onset of nasalization than vowels in canonical CV.NV# and derived CV.N#V heterosyllabic segment sequences. Furthermore, vowels in heterosyllabic (canonical and derived) contexts should exhibit similar anticipatory nasalization patterns.*

- *RQ 3:* Do vowels across word junctures (resyllabification) show parallel carryover nasalization patterns as those within word boundaries?

**Hypothesis 3:** *Yes. According to derivational rules of Spanish phonology and studies such as Cohn (1990) and Lahoz-Bengoechea and Jiménez-Bravo (forthcoming), it is expected that vocalic nasalization patterns in resyllabification CV.N#V contexts should present symmetrical patterns to those in heterosyllabic CV.NV# sequences.*

## 4. Methodology

### *4.1. Participants*

Nine native speakers of Northern Peninsular Spanish (7F, 2M) participated in this experiment. Their average age was 26 (age range = 23–29). All the participants were graduate students at a US university when the experiment took place. They were originally from the Basque Country, spoke Spanish as a native language, and had different levels of proficiency in Basque. There is no reason to believe that the internal structure of Basque could interfere with Spanish nasalization patterns, because, firstly, both Basque and Spanish show resyllabification (Hualde and Ortiz de Urbina 2003; Hualde 2014, respectively); secondly, there are no phonologically nasal vowels in the Basque varieties spoken by the participants (Central and Western), nor is there direct contact with the French language (Zuazo 2014); and, thirdly, none of the participants were proficient in any language that included phonologically nasal vowels. On the other hand, strictly speaking, the evidence presented here describes the Spanish variety of the Basque Country. In principle, there could be dialectal differences in this respect.

### *4.2. Stimuli*

The target tokens were adapted from Beristain (2022, p. 50) and fell under the following four different conditions: (1) CV.CV, tautosyllabic oral C and oral V sequences (as oral control tokens); (2) CVN#, tautosyllabic nasalized V and contiguous coda N; (3) CV.NV#, heterosyllabic nasalized V and canonical onset N, as well as a proceeding nasalized V; and (4) CV.N#V, derived onset N, with preceding and proceeding nasalized Vs (resyllabification context). The first vowel of each word was each of the vowels in Spanish, /i, e, a, o, u/, and the second vowel in the nasal structures was always /a/. The wordlist can be found in Table 1.

**Table 1.** Target tokens.

| | CV.CV (Oral Control) | CVN# (Tautosyllabic) | CV.NV# (Heterosyllabic) | CVN#V (Resyllabification) |
|---|---|---|---|---|
| /i/ | *tita*<br>'aunt' | *patín*<br>'rollerblade' | *patina*<br>'s/he rollerblades' | *patín atado*<br>'tied rollerblade' |
| /e/ | *cateto*<br>'ignorant' | *ten*<br>'you have, IMP.' | *tena*<br>'timber' | *ten atado*<br>'have (it) tied' |
| /a/ | *tato*<br>'little brother' | *tan*<br>'so' | *gitana*<br>'gypsy' | *tan atado*<br>'so tied' |
| /o/ | *pitote*<br>'fuss' | *botón*<br>'button' | *botona*<br>'s/he buttons' | *botón atado*<br>'tied button' |
| /u/ | *batuta*<br>'baton' | *atún*<br>'tuna' | *gatuna*<br>'cat-like' | *atún atado*<br>'tied tuna' |

As can be noted, the stress patterns across words were uniform: the first vowel (V1) was always stressed, and the second vowel (V2), in the CV.N**V**# and CV.N#**V** contexts, was an unstressed /a/. This was undertaken to normalize the data because research has shown that differences in stress may lead to varying degrees of nasalization (Byrd et al. 2009; Cohn 1990; Krakow 1989, 1999).

The list was presented to participants before the experiment started, and the meaning of the words was explained in case they were not familiar with it, which did not happen in the current experiment. Target words were included in the carrier phrase Digo TARGET ligeramente 'I say TARGET softly'.

### *4.3. Equipment*

The aerodynamic data were collected via a vented Scicon OM-2 oral mask (Scicon R&D, Inc., Beverly Hills, CA, USA) that was connected to two TSD160A pressure transducers (operational pressure 72.5 cm $H_2O$; Biopac Systems, Goleta, CA, USA). Simultaneous acoustic data were obtained to facilitate the segmentation process. For that, the signal was preamplified using a Grace M101 microphone preamplifier (Grace Designs, Boulder, CO, USA) and digitized at 2 kHz, the highest allowed in the Biopac System (BIOPAC 2020). Participants wore an AKG C-520 head-mounted microphone (Harman International, Stamford, CT, USA) that was located approximately 3 cm (1 inch) away from their mouths. For a more detailed report about the equipment, see Beristain (2022, 2023a).

### *4.4. Data Collection and Procedure*

The author of the current study calibrated the equipment manually before every session by utilizing an AFTA6A Calibration syringe of 600 mL (Beristain 2023a, 2023b) and annotating the correction value of the signal for every word uttered and the surface area of the air volume expelled by the syringe. Those numbers were inserted in a calibration formula that was applied to the raw data of each participant. The conversion equation used is presented in (1) (Beristain 2023a, 2023b; Shosted 2009), as follows:

$$\mathrm{s}\prime = \mathit{filtfilt}\ (\mathrm{s}\ \times\ \frac{v}{\int s} + \mathrm{c}) \tag{1}$$

As cited in Beristain (2023b, p. 9): "$s'$ is the new, resulting signal; *filtfilt* is the MATLAB function; *s* is the original, unaltered signal; $\int$ means integration in order to calculate the area under the signal curve; *v* is the total volume of the syringe (= 600 mL), and *c* is the correction number (specific to every word)".

Data were collected in a sound-attenuating booth inside a phonetics laboratory of a university in the USA. In order to avoid as much linguistic co-activation as possible, the author of the current study (who is a native speaker of Northern Peninsular Spanish) communicated with the participants in Spanish at all times. Participants signed a written consent before starting the experiment.[1]

Participants were then informed about the experiment and shown how to operate the equipment. Before every experimental procedure, a trial session took place to ensure no air leakage was present. None of the current participants showed anomalous results in their Spanish trial productions. Once the experiment began, participants held the mask holding the handle attached to it. Recordings were stopped after every minute, in order to alleviate any possible discomfort.

The software that was utilized to obtain the data was AcqKnowledge (version 3.9.1). Three different and simultaneous signals were collected: (1) nasal airflow, (2) oral airflow, and (3) audio (acoustics). The current study will only present nasal airflow data. Future studies will provide a multidimensional analysis presenting a combination of several signal types. MATLAB (version 2020a; MATLAB 2020) was used to convert the AcqKnowledge files into .wav files that were readable in Praat (Boersma and Weenink 2020).

### *4.5. Data Annotation*

In Praat, data were segmented and annotated by inspecting the nasal airflow, oral airflow, and audio signals. The onset of the first vowel (V1) was located after the visible burst of air of /t/ decreased and a more periodic signal was present both in the oral airflow and audio signals, indicating vowel periodicity (see Beristain 2023a, 2023b; Delvaux et al. 2009). The amplitude of the spectrogram was also a clear cue for distinguishing the vowel from contiguous consonants, both oral and nasal. The cues used to segment oral second vowels (V2) were: the oral air burst of /t/, the more periodic signal apparent in oral and audio channels, and the greater amplitude. The cues for nasalized V2s were as follows: the onset of oral airflow following the decrease in nasal airflow (indicating the N-V transition) and the greater amplitude present in the vowel. The offset of nasalized V2s was located when the amplitude decreased significantly from the V-/C/ transition. Figure 2 presents two examples with segmentations that include a fully oral CVCV token, *batuta* 'baton', and a CV.N#V (resyllabification) one, *atún atado* 'tied tuna'. Notice how the signal in Channel 1 (nasal airflow) is virtually flat, showing essentially zero nasal airflow in *batuta* (N: nasal; O: oral; A: audio/acoustic).

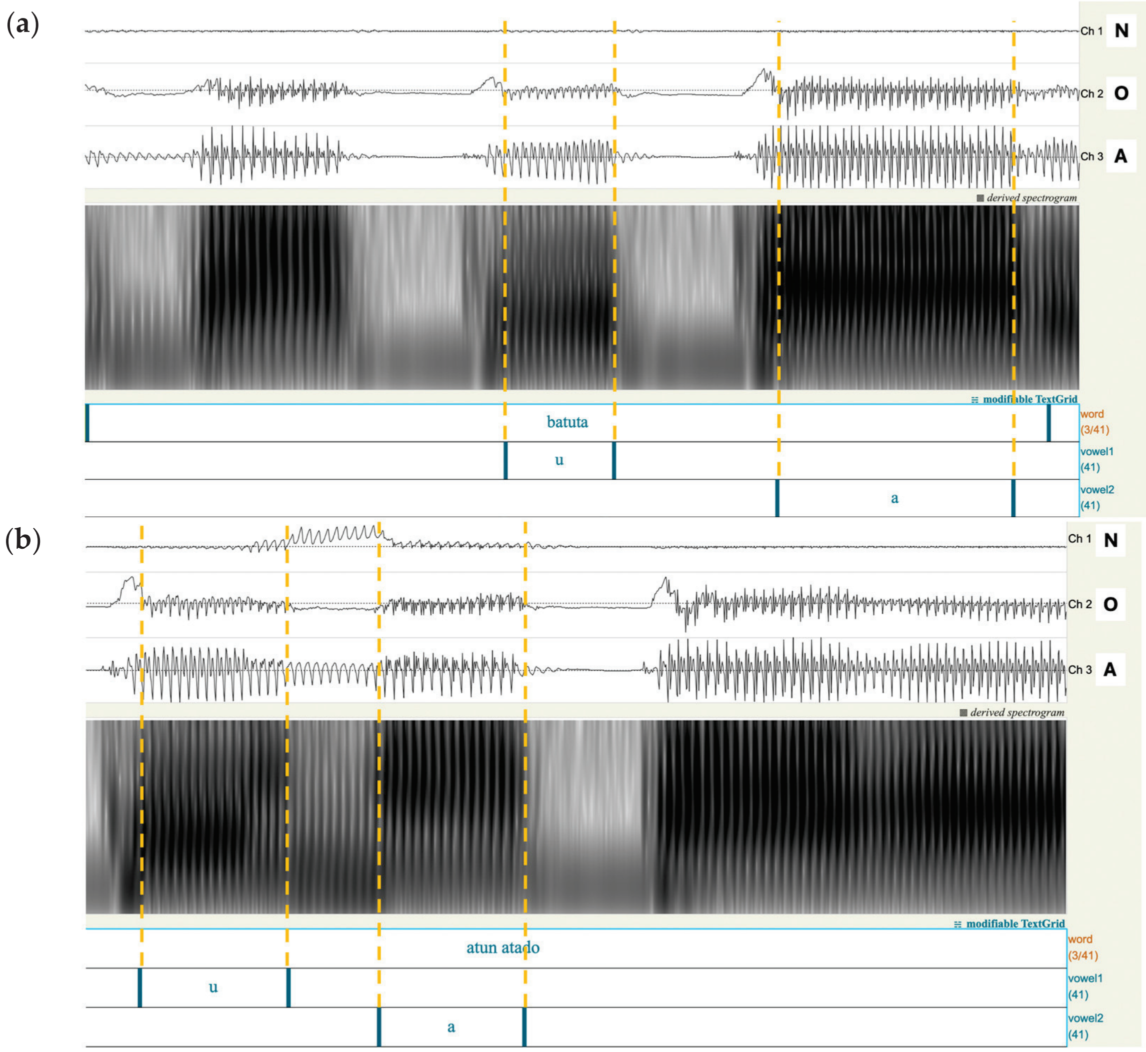



**Figure 2.** Sample segmentation of CVCV and CV.N#V sequences: (**a**) *batuta* 'baton'; (**b**) *atún atado* 'tied tuna'; (Speaker #21, Rep. #1).

### 4.6. Data Analysis

Taking into consideration the goal of this experiment, the time dynamics of nasal airflow in V1 and V2 were analyzed. Each segment was further divided into 40 equidistant timepoints. The total number of datapoints gathered and submitted for the statistical analysis is presented in Table 2.

**Table 2.** Number of tokens analyzed.

| | Number |
|---|---|
| Anticipatory nasalization (V1) | 9 speakers × 5 vowel conditions [i, e, a, o, u] × 40 time-points × 4 syllable conditions [CVCV; CVN#; CV.NV#; CV.N#V] × 2 repetitions × = 14,400 |
| Carryover nasalization (V2) | 9 speakers × 5 vowel conditions [i, e, a, o, u] × 40 time-points × 3 syllable conditions [CVCV; CV.NV#; CV.N#V] × 2 repetitions = 10,800 |
| | Total: 25,200 datapoints |

To statistically analyze the time dynamics of nasal airflow, generalized additive mixed models (GAMMs) were run in R (R Core Team 2020) and RStudio (R Studio Team 2020) using the *mgcv*, v. 1.8.31 package (Wood 2011). Data visualization was conducted using the *itsadug* package, v. 2.3 (Van Rij et al. 2017). The optimal statistical models for V1 and V2 were chosen by inspecting Akaike Information Criterion (AIC) scores (Akaike 1974). Such models are listed below (see Appendix A for further details).

- Optimal model for V1: `Context * Vowel + Sex + s(NormTime, by = interaction (Context, Vowel)) + s(NormTime, by = Sex) + s(NormTime, Speaker, by = interaction(Context, Vowel), bs = "fs", m = 1) + s(Word, bs = "re")`
- Optimal model for V2: `Context + Sex + s(NormTime, by = Context) + s(NormTime, by = Sex) + s(NormTime, Speaker, by = Context, bs ="fs", m = 1) + s(Word, bs = "re")`

As shown, the statistical model to analyze anticipatory vowel nasalization (V1) included an interaction between CONTEXT (C**V**CV; C**V**N#; C**V**.NV#; C**V**.N#V) and VOWEL (i; e; a; o; u) as well as the fixed factor of SEX (male; female). A smooth function was included through NORMTIME (from 0 to 1 with 0.025 increments). The statistical model to analyze carryover nasalization (V2) was similar to that of V1 without VOWEL in it. This is because V2 was always /a/ in CV.N**V**# and CV.N#**V** tokens.

## 5. Results

This section is divided into two parts: (1) V1 (anticipatory vowel nasalization) and (2) V2 (carryover vowel nasalization). As is customary with GAMM reports, figures will present contours and difference curves. The $x$-axis of each figure indicates normalized time (NormTime), which analyzes nasal airflow (in L/s) across 40 equidistant timepoints from the onset to the offset of the vocalic segment. The darker line within the contour indicates the mean values, while the lighter color indicates confidence intervals, i.e., variability in the data. The nasalized target and oral control contexts are presented. Note that even the oral control context shows a certain degree of nasal airflow, as air will inevitably be exhaled through both oral and nasal apertures when producing speech. Most importantly, the quantity and shape of nasal airflow in the oral control context is significantly lesser and stable, as opposed to rising contours in sequences that contain nasal structures.

The way in which statistical significances are provided and onset and offset of nasalization are calculated is by means of statistical pair-wise comparisons, retrieved via the plot_diff() function under the *itsadug* package[2]. Oral control segments and nasalized target segments are compared, and the difference curves of their nasal airflow are analyzed, thus pointing out and locating the onset and offset of the nasal gesture in the time window of the whole segment (Beristain 2023a). The area in which the two airflow curves are significantly different is the "time region of significance", and that will be delimited by red dotted lines in the difference curve figures. Those areas are what will be reported in the results section. The full statistical model output for V1 can be found in Appendix B, while that for V2 can be found in Appendix C.

### 5.1. Anticipatory Vowel Nasalization

The results for V1 nasal airflow curves are illustrated in Figure 3.

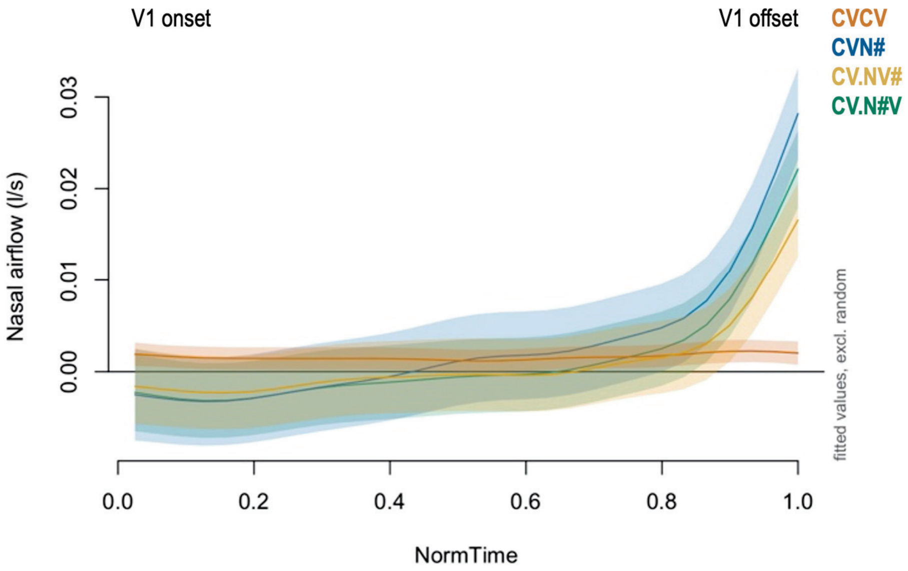



**Figure 3.** Nasal airflow GAMM curves of V1 in C**V**CV (oral control), C**V**N# (tautosyllabic), C**V**.NV# (heterosyllabic), and C**V**.N#V (resyllabification)[3].

As can be seen, the nasal airflow contour in the oral control (C**V**CV) context is stable and virtually non-existent. The tautosyllabic (C**V**N#) sequence shows the greatest amount of nasal airflow for V1, followed by resyllabification (C**V**.N#V), and then the heterosyllabic (C**V**.NV#) one. The pair-wise comparisons between each nasalized context and the oral control exhibit differences. Such differences are shown and described in Figures 4a,b, 5a,b and 6a,b.[4]

Figure 4a compares anticipatory nasalization contours in the oral control C**V**CV structure and the tautosyllabic C**V**N# one and illustrates a minimal and stable development of nasal airflow contour for C**V**CV, while a rise is seen towards to end of the segment for C**V**N#. This is to be expected, as the segment to follow V1 in C**V**N# is a nasal consonant, and as such, due to anticipatory coarticulatory nasalization, the onset of the nasal gesture that applies to the nasal consonant will begin before the offset of the previous vowel, i.e., V1. As shown in Figure 4b, the time region of significance starts at 0.86. Aerodynamic data for Spanish has shown delayed onsets of nasalization, as opposed to other languages like English (Beristain 2023a).

(**a**)

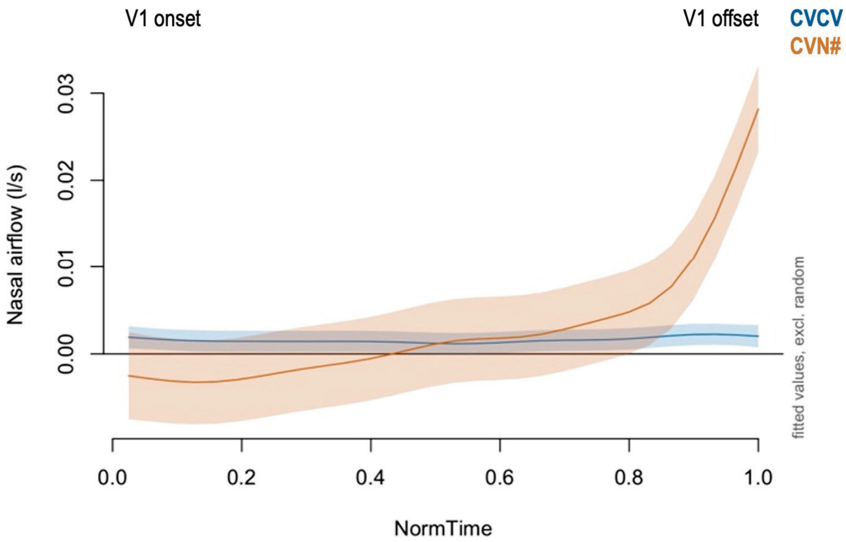



**Figure 4.** *Cont.*

(**b**)

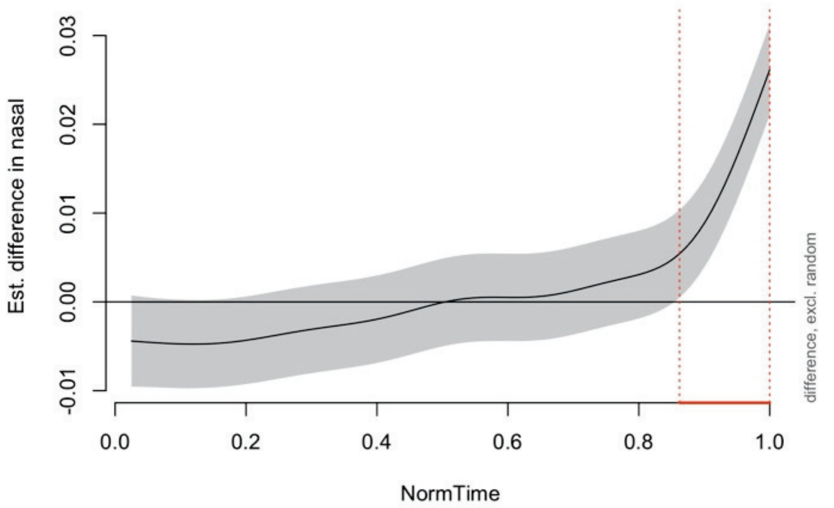
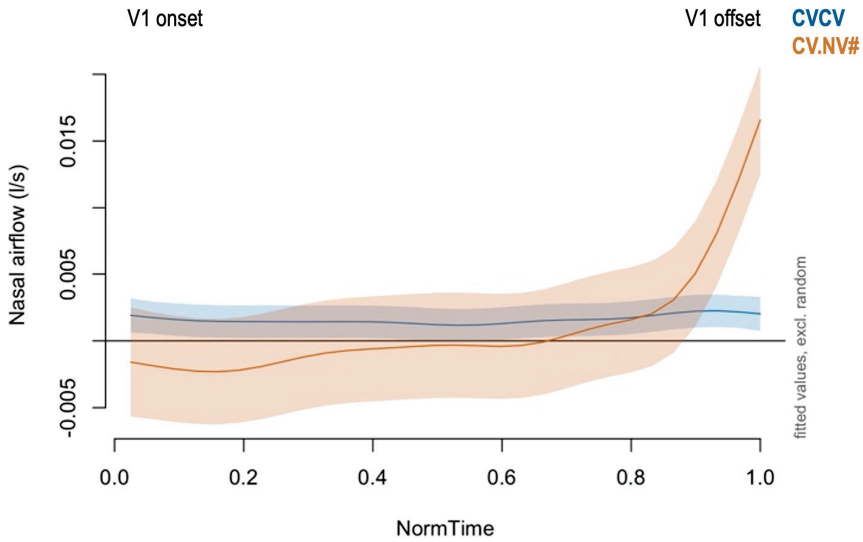



**Figure 4.** (**a**) Nasal airflow GAMM curves of V1 in C**V**CV and C**V**N# syllable structures; (**b**) difference curve of C**V**N# and C**V**CV (in V1).

(**a**)




(**b**)

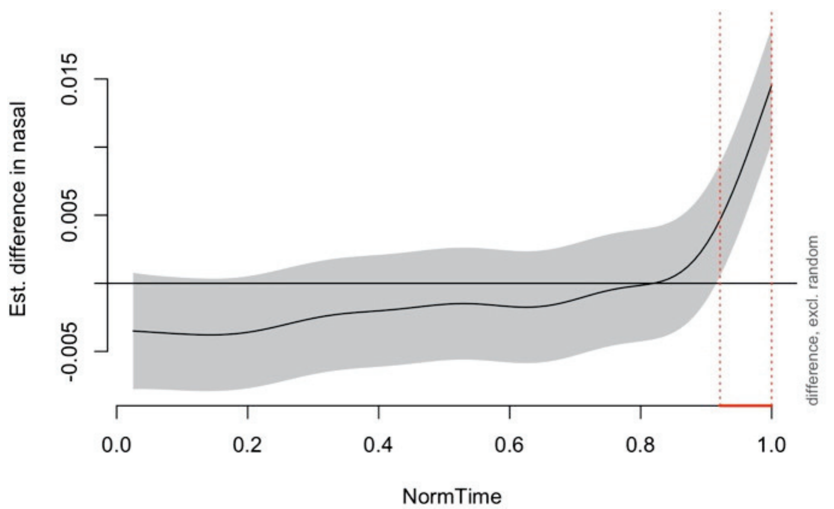
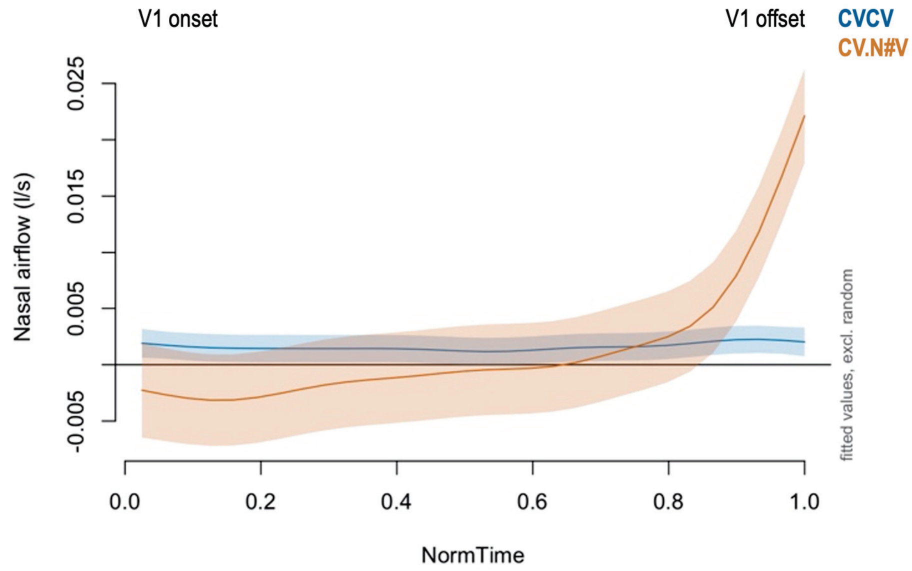



**Figure 5.** (**a**) Nasal airflow GAMM curves of V1 in C**V**CV and C**V**.NV# syllable structures; (**b**) difference curve of C**V**.NV# and C**V**CV (in V1).

(**a**)




(**b**)

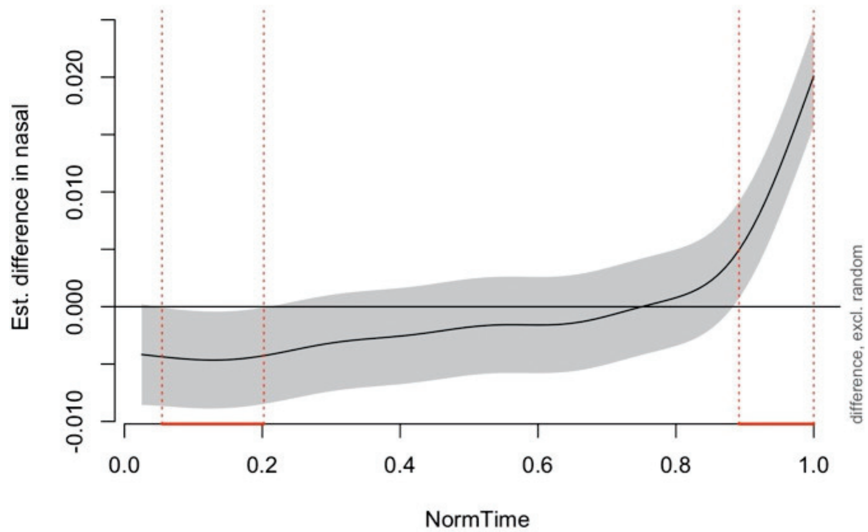



**Figure 6.** (**a**) Nasal airflow GAMM curves of V1 in C**V**CV and C**V**.N#V syllable structures; (**b**) difference curve of C**V**.N#V and C**V**CV (in V1).

Figure 5a compares anticipatory nasalization contours in C**V**CV and the heterosyllabic C**V**.NV# sequences, and it shows similar patterns to those in Figure 4a. The main difference is that the nasal airflow contour for C**V**.NV# exhibits lower amplitude, and confidence intervals are wider. Figure 5b reveals that the time region of significance starts at 0.92, which is later than for C**V**N#.

In Figure 6a, the C**V**CV sequence is compared to the C**V**.N#V resyllabification context. The patterns found are more similar to those in C**V**N#. The first time region of significance in Figure 6b (0.05 to 0.20) should not be considered an indicator of onset of nasalization because the downward directionality of the nasal airflow could pertain to air inhalation and not velum lowering. The second time region of significance starts at 0.89.

Below is Table 3, which summarizes the onset of significant differences between nasalized and oral V1s, which we consider a cue for onset of nasalization.

**Table 3.** Onset of nasalization (0–1) of V1.

| | Onset of Nasalization |
|---|---|
| C**V**N# (tautosyllabic) | 0.86 |
| C**V**.NV# (heterosyllabic) | 0.92 |
| C**V**.N#V (resyllabification) | 0.89 |

Regarding inferential statistics, the only significant difference between the nasal airflow contours in nasalized vowels (C**V**N#, C**V**.NV#, and C**V**.N#V) was retrieved between C**V**N# and C**V**.NV#, as illustrated in Figure 7a,b. The tautosyllabic C**V**N# sequence exhibited a greater amount of nasal airflow than the resyllabification C**V**.N#V context. The time region of significance started at 0.91.

(**a**)

(**b**)

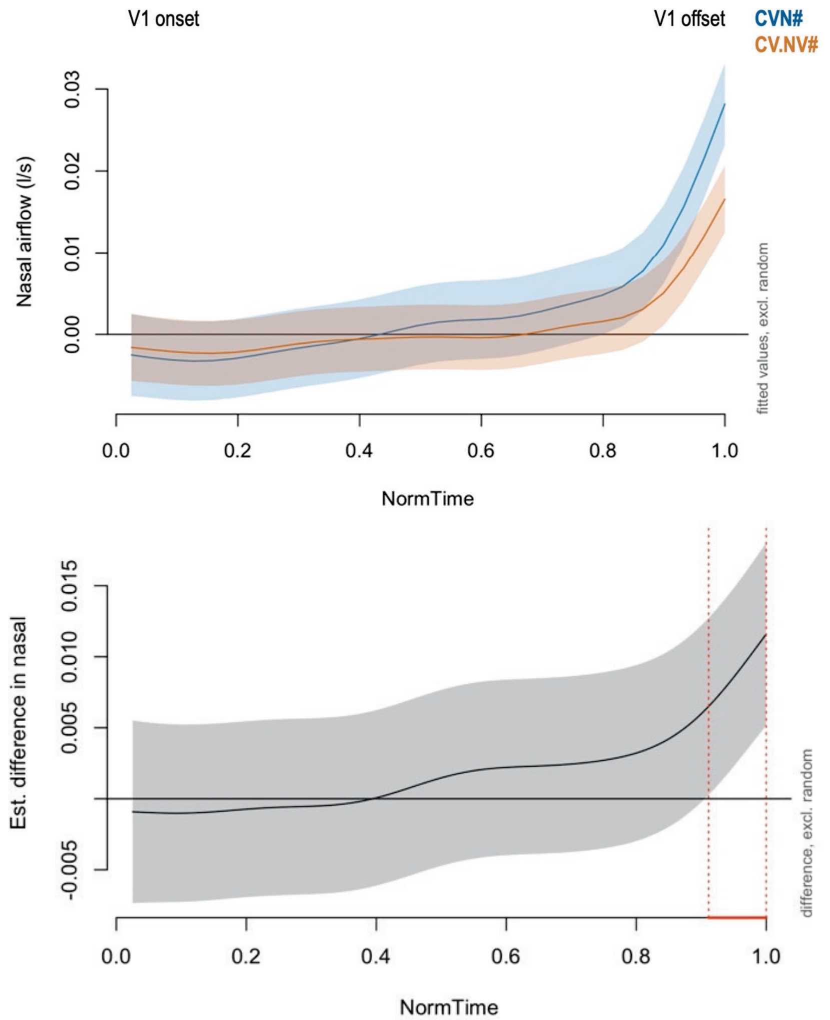
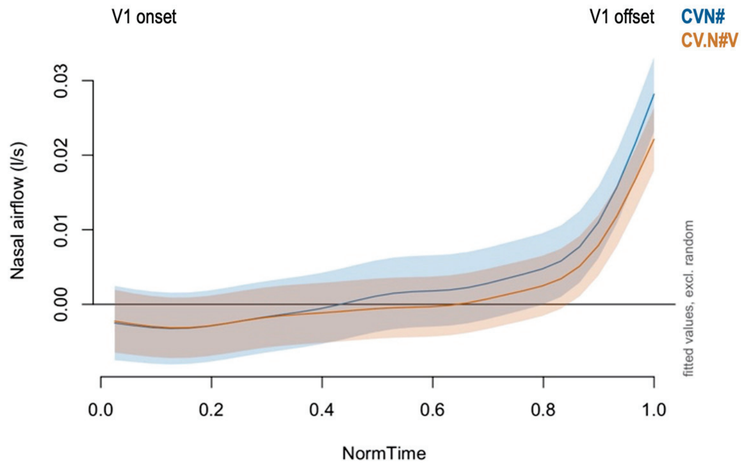



**Figure 7.** (**a**) Nasal airflow GAMM curves of V1 in C**V**N# and C**V**.NV# syllable structures; (**b**) difference curve of C**V**N# and C**V**.NV# (in V1).

Regarding the contour differences between C**V**N# and C**V**.N#V, the former showed a higher degree of nasal airflow starting relatively close to the midpoint of the vowel (see Figure 8a). However, no significant differences were found between the two contexts, as shown in Figure 8b.

(**a**)




(**b**)

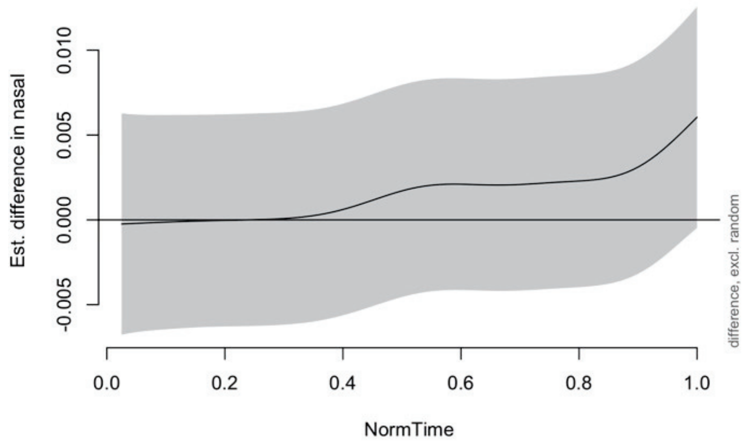



**Figure 8.** (**a**) Nasal airflow GAMM curves of V1 in C**V**N# and C**V**.N#V syllable structures; (**b**) difference curve of C**V**N# and C**V**.N#V (in V1).

## *5.2. Carryover Vowel Nasalization*

The results in this section pertain to nasal airflow in the second vowel (V2), i.e., CVC**V** (oral control), CV.N**V**# (heterosyllabic), and CV.N#**V** (resyllabification). Figure styles are similar to those in Section 5.1. The *x*-axis represents normalized time (NormTime) of 40 equidistant points, and *y*-axis indicates the amount of nasal airflow produced (in L/s). The left side of figures are the onset of the vowel, while the right side are its offset. Notice that the CVN# (tautosyllabic) syllable structure will not be included in this section because there is no V2. The offset of carryover nasalization was calculated by contrasting the V2 nasal airflow contours in CVC**V** (oral control) versus CV.N**V**# and CV.N#**V** (nasalized).

Figure 9 illustrates nasal airflow in CVC**V**, CV.N**V**#, and CV.N#**V**. As can be seen, CVC**V** shows a stable and virtually non-existent nasal airflow, which is to be expected as no nasal segment is present in that sequence. With regard to the nasalized contexts and carryover nasalization patterns, CV.N**V**# and CV.N#**V** show parallel patterns: a sharp decline is present at the beginning of V2, induced by the presence of the preceding nasal segment /n/. After this, nasal airflow stabilizes throughout the remainder of the vowel, always showing higher nasal airflow values than the oral control context CVC**V**.

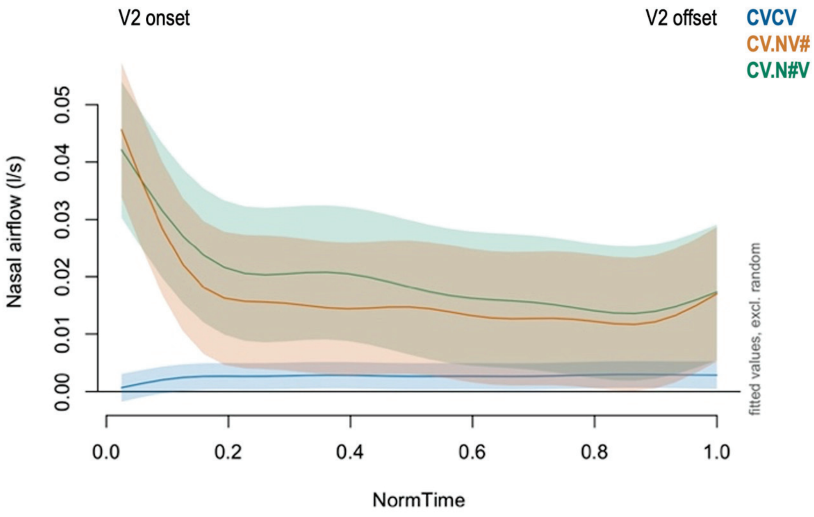



**Figure 9.** Nasal airflow GAMM curves of V2 in CVC**V**, CV.N**V**#, and CV.N#**V**.

Pair-wise comparisons are presented in Figures 10a,b, 11a,b, and 12a,b.

(**a**)

(**b**)

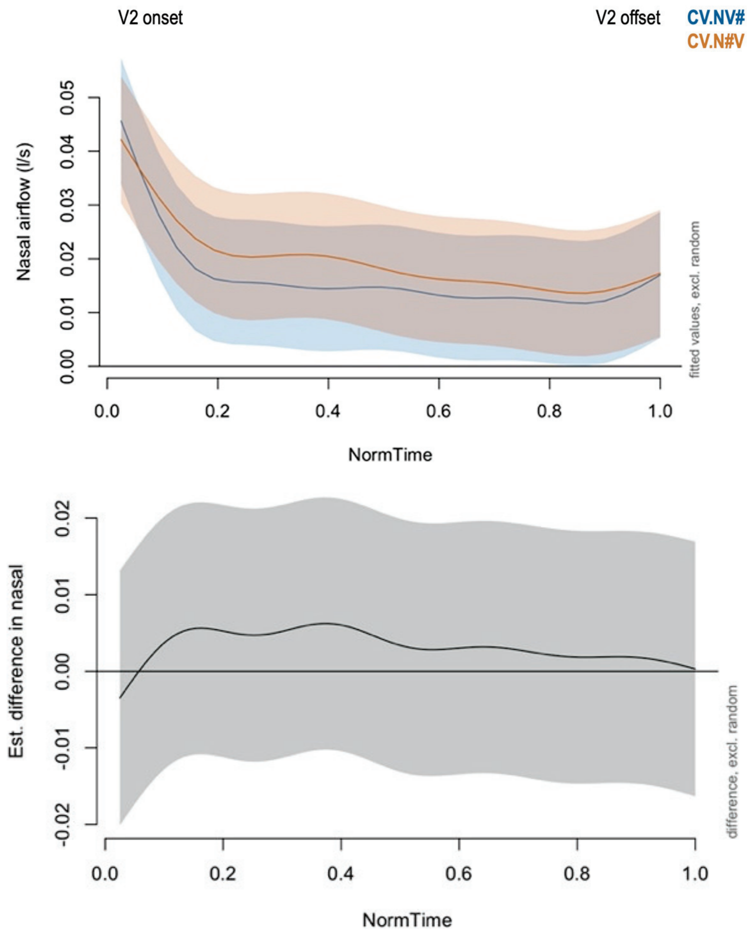



**Figure 10.** (**a**) Nasal airflow GAMM curves of V2 in CV.N**V**# and CV.N#**V** syllable structures; (**b**) difference curve of CV.N**V**# and CV.N#**V** (in V2).

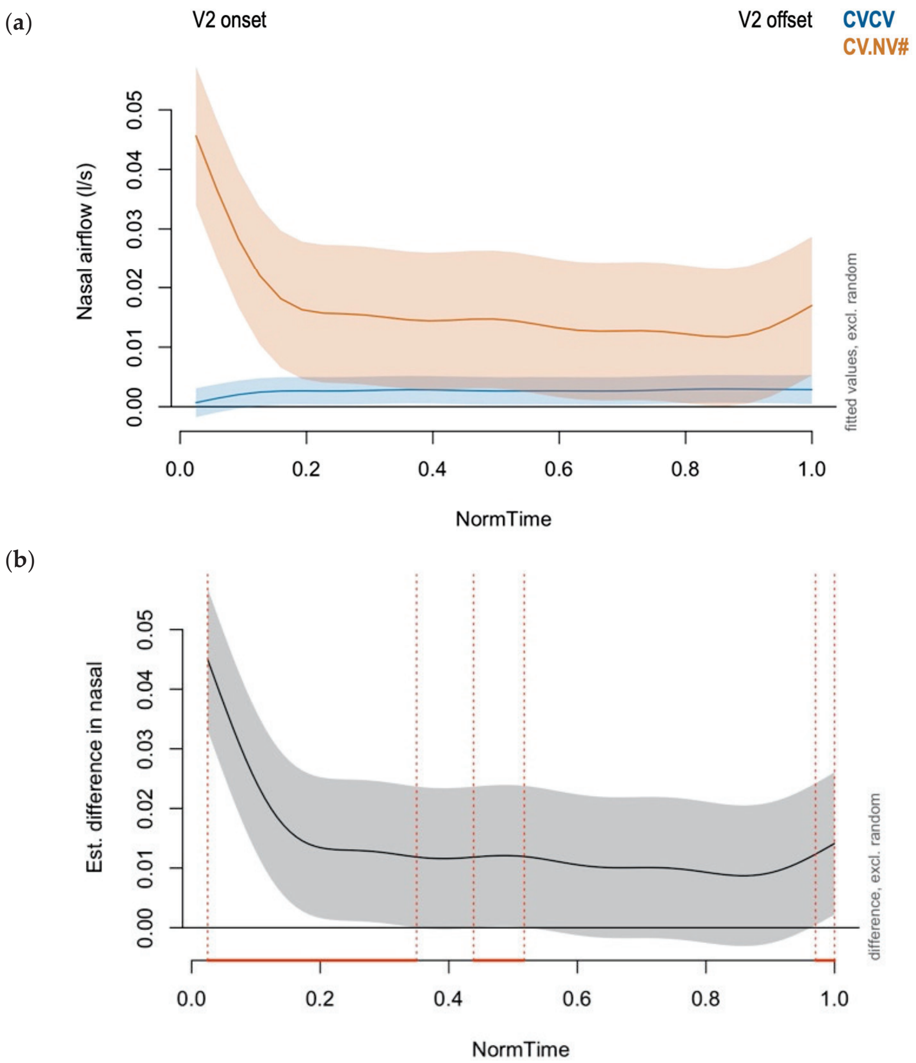
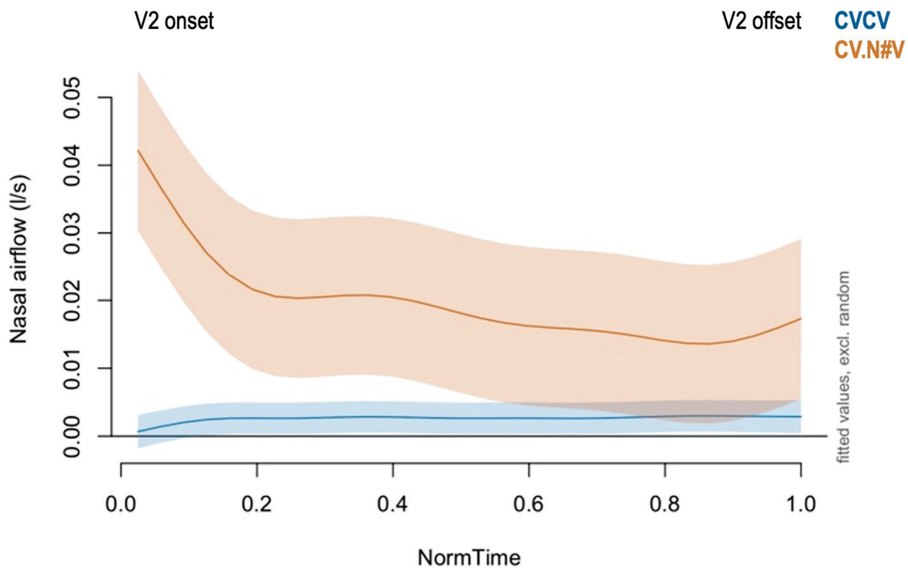



**Figure 11.** (**a**) Nasal airflow GAMM curves of V2 in CV.N**V**# and CVC**V** syllable structures; (**b**) difference curve of CV.N**V**# and CVC**V** (in V2).

As shown in Figure 10a, the V2 nasal airflow contours in nasalized contexts show similar patterns between CV.N**V**# and CV.N#**V**. As explained in Figure 9, a rapid decline during the initial portion of the vowel is followed by a more stable nasal airflow contour that does not decrease to zero. When contrasting both nasal airflow contours, no significant differences are found (see Figure 10b).

Figure 11a compares the nasal airflow contours of CVC**V** and CV.N**V**#. The oral context shows a stable and low degree of nasalization, to be expected from its structure, and the heterosyllabic CV.N**V**# sequence shows a rapid decline, with some overlap between the 0.2–1 time region. The time regions of significance between contours arise in 0–0.35, 0.43–0.51, and 0.97–1 (Figure 11b). One could argue the offset may be when the first time region of significance culminates, but we should be cautious considering the various time regions of significance and the physiological nature of the data. If we observe the data, we will see that the reason for the lack of statistical difference between the two time regions of significance is due to the downward directionality of the nasal airflow in CV.N**V**#. Since the nasal airflow remains relatively stable in that portion of the vowel, there is no reason to believe that the offset of carryover vocalic nasalization is that first instance, and we should thus set it at the second one at 0.51. The third time region of significance is not an indicator or cue of onset of nasalization, as the following consonant is a /l/.

(**a**)




(**b**)

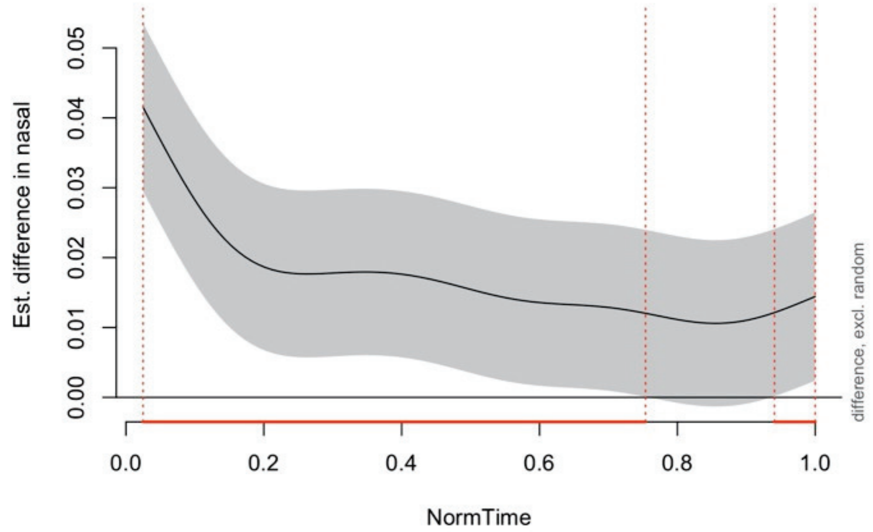



**Figure 12.** (**a**) Nasal airflow GAMM curves of V2 in CV.N#**V** and CVC**V** syllable structures; (**b**) difference curve of CV.N#**V** and CVC**V** (in V2).

In Figure 12a, the CV.N#**V** nasal airflow contour is compared to that of CVC**V**. As can be seen, the distance between the two curves is greater and the overlap is more delayed than in Figure 11a, which applies to CV.N**V**#. Approximately, the overlap in Figure 12a begins close to 0.6, and extends until the end of V2. By contrasting both contours, time regions of significance arise at 0–0.75 and 0.94–1 (see Figure 12b). Notice that the second significance window frame arises in the nasal airflow due to the rise of the CV.N#**V** curve in the final portion of V2.

Table 4 provides a summary of the offsets of nasalization for V2.

**Table 4.** Offset of nasalization (0–1) of V2.

| | **Offset of Nasalization** |
|---|---|
| CV.N**V**# (heterosyllabic) | 0.51 |
| CV.N#**V** (resyllabification) | 0.75 |

## 6. Discussion

Let us revisit the goals and hypotheses of the current study: first, to document the differences between anticipatory and carryover coarticulatory nasalization in Spanish. Based on previous studies such as Diakoumakou (2009) and Delvaux et al. (2009), we hypothesized that carryover nasalization would present to a greater extent than anticipatory nasalization. Second, to elucidate whether the extent of coarticulatory anticipatory and carryover nasalization in Spanish is affected by tautosyllabicity and heterosillabicity of VN sequences. We hypothesized that segments within the same syllable would show greater gestural overlap (Byrd et al. 2009) and an earlier onset of nasalization than those in canonical and derived (resyllabification) heterosyllabic sequences. Third, to test the role of resyllabification on anticipatory and carryover coarticulatory nasalization. Previous literature (Bradley et al. 2022; Hualde and Prieto 2014; Jiménez-Bravo and Lahoz-Bengoechea 2023; Strycharczuk and Kohlberger 2016) has shown differences between canonical and derived onset consonants. However, articulatory data show that the characteristics of neighboring vowels in those contexts are similar (Cohn 1990; Krakow 1989, 1999). Moreover, Lahoz-Bengoechea and Jiménez-Bravo (forthcoming) find the degree of anticipatory nasalization in **V**#NV and **V**N#V sequences to be identical.

This study showed that the degree of coarticulatory carryover nasalization is greater than that of anticipatory nasalization in Spanish (RQ1, Hypothesis 1). Anticipatory nasalization was implemented in the final 15% of the segment, while carryover nasalization extended from the beginning to 51–75% of the vocalic segment. Previous literature had diverging proposals as to which one of the two shows greater nasal spreading. An argument that is generally used in favor of this view is how current phonologically nasal vowels in Romance languages are a development of contiguous VN segments and not NV (Sampson 1999). Henderson (1984), by means of data obtained with a fiberoptic endoscope, argues that it is anticipatory nasalization which shows greater extent, and provides evidence for the velum reaching lower positions in C**V**N# sequences than in N**V**C#. However, it should be noted that, in Henderson's study, the coda consonant in the NVC# sequence is a /t/, which could induce an earlier rise of the velum in the gestural timing and transition. Other studies such as Diakoumakou (2009) have pointed out that the extent of coarticulatory nasalization is language-specific. Her investigation focuses on Modern Greek, but also considers other languages, and proposes that languages that favor open syllables show a tendency for greater carryover nasalization, while those languages that favor closed syllables show greater extent of anticipatory vowel nasalization. Spanish is an example of a language that favors the CVCV type of open syllables, and greater carryover nasalization is indeed the pattern that is encountered. In recent studies comparing coarticulatory nasalization in CVN sequences in English and Spanish, Beristain (2023a, 2023b) found that the onset of nasalization was significantly earlier in English, and that the ratio of nasal airflow proportion (to total airflow produced) was significantly higher in English, too. Whether this can simply be explained by the preference of English for closed syllables or linguistically internal reasons goes beyond the scope of the current study.

Furthermore, evidence for an earlier onset of vocalic nasalization in tautosyllabic C**V**N# sequences as opposed to heterosyllabic canonical C**V**.NV# and derived C**V**.N#V sequences (RQ2, Hypothesis 2) was found. This finding is in alignment with previous literature that considers (a) articulatory gestures to be an important part of the physiological part of the syllable (Byrd et al. 2009; Krakow 1999), and (b) that the timing among gestures of segments within the same syllable possess greater articulatory overlap and coordination than those in different syllables. In the C**V**N# sequence, the velum will have started to reach a spatially lower position during the vocalic segment by the time we have the onset of N. On the other hand, for a vowel that is contiguous to a nasal consonant in a heterosyllabic condition, the gestural coordination will be different. As a word-medial onset, the articulatory effort is not equal to a word-initial onset. Krakow (1999) points out that syllable-initial consonants are 'stronger' because "it is generally associated with tighter articulatory constrictions and with greater stability" (p. 25). The significant difference found in anticipatory nasalization between C**V**N# and C**V**.NV# could be induced because of syllable structure. However, the lack of significant differences between C**V**N# and C**V**.N#V complicates that argument. As far as structural similarities are concerned in the generative literature, one would have expected differences to arise between C**V**N# vs. C**V**.NV#, and C**V**N# vs. C**V**.N#V, not solely in the former[5]. Lahoz-Bengoechea and Jiménez-Bravo (forthcoming) found similar results when comparing VN sequences in **V**N# and **V**#N structures. The authors point out that as coarticulatory nasalization is a purely phonetic and a mechanical byproduct in Spanish, "[its effect arises in close contact to the nasal consonant, independently from its phonological structure]" (p. 3, translated by the author of this manuscript). Previous literature on the resyllabification of non-nasal segments has deduced that derived onsets exhibit distinct sub-phonemic characteristics from canonical onsets (Hualde and Prieto 2014; Strycharczuk and Kohlberger 2016). Although those studies have focused on oral fricatives mostly, it could be hypothesized that some additional cues might be affected in nasal consonants in resyllabification contexts as well. A future study will provide a multidimensional view of aerodynamic and acoustic characteristics of the nasal segments that were examined in the current study.

Third, in alignment with results presented in Cohn (1990), the current data demonstrated that carryover nasalization patterns were parallel between: (i) within word, heterosyllabic CV.N**V**#, and (ii) across word boundaries, resyllabification CV.N#**V** sequences (RQ3, Hypothesis 3). Most of the previous literature had focused on the way in which resyllabification affected the consonant, and those studies found differences in terms of the acoustic cues such as duration and voicing frames of oral fricatives. Little had been mentioned about how coarticulatory nasalization could spread across word junctures and across syllabic reorganization processes. This study showed that no significant differences were found regarding anticipatory or carryover nasalization among CV.N**V**# and CV.N#**V** sequences, which correlates with what Strycharczuk and Kohlberger (2016) found; that is, that vowels are less malleable in their phonetic production as opposed to the consonant for which the syllabic linking changes. As pointed out in the previous paragraph, one of the reasons for this may be that coarticulatory nasalization is a mechanical byproduct in Spanish, and the generative literature considers CV.N**V**# and CV.N#**V** sequences structurally identical. An important difference between previous studies and the current one is the nature of the data. While most of the previous studies investigating the effect of resyllabification explore acoustic correlates, the current one delves into the physiological characteristics of speech; in other words, it analyzes the way in which air is exhaled during speech production. This avenue had not been explored in the context of syllable structure realignment. As far as resyllabification is concerned, generative phonology considers its structure identical to any other heterosyllabic structure, yet previously mentioned laboratory studies provided evidence for acoustic differences between canonical and derived onsets. The results obtained in this study may elucidate a novel perspective on connected speech and its articulatory features. Here, the extent of phonetic characteristics such as coarticulation can be studied more accurately and without the need to conclude differences from fine-grained details in the acoustic signal.

Lastly, the limitations of the current study cannot go unnoticed. A sample of nine participants and a database of 630 vocalic tokens were used to draw conclusions. Furthermore, Spanish data were collected in an English-speaking environment. As such, we need to consider that participants' English could have influenced their Spanish productions. The researcher, a native Spanish speaker, tried to induce participants into a "monolingual"-mode to the best of his abilities where only Spanish was used during the experimental setting. Moreover, while none of the target tokens were uncommon words, their lexical frequency was not meticulously controlled (cf. Brown 2008). Doing so could have contributed to a more thorough understanding and analysis of the topic under study. Considering that aerodynamic experiments have reported small(er) sample sizes (Kochetov 2020), the current study is a significant contribution to the literature on syllable structure, Spanish phonology, and articulatory phonetics.

## 7. Conclusions

The two aims of the current study were to investigate aerodynamic differences between anticipatory and carryover coarticulatory nasalization in Spanish and to observe whether nasalization can be used as a cue to resyllabification.

The contours of aerodynamic nasal airflow data illustrate that carryover nasalization exhibited greater nasal spreading than the anticipatory context. This was in alignment with previous literature, which points out that such a pattern can be found in languages that favor open syllables. In that regard, the degree of anticipatory nasalization was greater in C**V**N# tautosyllabic structures compared to in heterosyllabic C**V**.NV# and resyllabification C**V**.N#V ones. Regarding carryover nasalization, vowels in the heterosyllabic CV.N**V**# and resyllabification CV.N#**V** contexts showed parallel patterns.

These findings indicate that the position and structure of the syllable are crucial elements to take into consideration and that they affect inter-articulatory gestural overlap. Furthermore, the current study demonstrates that vowels in heterosyllabic C**V**.N**V**# and resyllabification C**V**.N#**V** structures are similar in terms of their aerodynamic properties, thus providing evidence in favor of derivational rules in Spanish phonology as an exemplar of complete resyllabification in connected speech.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of the University of Illinois Urbana-Champaign (protocol code #20071 and 16 August 2019) for studies involving humans.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Not applicable due to the IRB protocol does not allow the author to share the data openly for privacy and confidentiality restrictions.

**Acknowledgments:** I would like to thank the audience at the 51st *Linguistic Symposium on Romance Languages*, the two anonymous reviewers and editor for their comments, José Ignacio Hualde for his input on a previous version of this manuscript, and Jennifer Zhang for stylistic feedback.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Appendix A

Model selection and AIC scores

| Model | AIC Score | Deviance Explained |
|---|---|---|
| **Coarticulatory anticipatory nasalization (V1):** | | |
| `Context * Vowel + Sex + s(NormTime, by = interaction(Context, Vowel)) + s(NormTime, by = Sex) + s(NormTime, Speaker, by = interaction(Context, Vowel), bs = "fs", m = 1) + s(Word, bs = "re")` | −127,117.7 | **72.9%** |
| `Context * Vowel + s(NormTime, by = interaction(Context, Vowel)) + s(NormTime, Speaker, by = interaction(Context, Vowel), bs = "fs", m = 1) + s(Word, bs = "re")` | −114,413.8 | 67% |
| **Coarticulatory carryover nasalization (V2):** | | |
| `Context + Sex + s(NormTime, by = Context) + s(NormTime, by = Sex) + s(NormTime, Speaker, by = Context, bs = "fs", m = 1) + s(Word, bs = "re")` | −86,151.6 | **80.7%** |
| `Context + s(NormTime, by = Context) + s(NormTime, Speaker, by = Context, bs = "fs", m = 1) + s(Word, bs = "re")` | −77,586.6 | 78.5% |

## Appendix B

Model output for anticipatory vowel nasalization (V1)
Formula: `nasal ~ Context * Vowel + Sex + s(NormTime, by = interaction(Context, Vowel)) + s(NormTime, by = Sex) + s(NormTime, Speaker, by = interaction (Context, Vowel), bs = ‘‘fs’’, m = 1) + s(Word, bs = ‘‘re’’)`

| Parametric Coefficients: | Estimate | SE | *t*-Value | *p*-Value |
|---|---|---|---|---|
| (Intercept) | 0.0009 | 0.002 | 0.502 | 0.61580 |
| Context CVN# | 0.002 | 0.003 | 0.735 | 0.46240 |
| Context CV.N#V | 0.0004 | 0.002 | 0.174 | 0.86202 |
| Context CVCV | 0.0007 | 0.002 | 0.378 | 0.70522 |
| Vowel e | 0.00006 | 0.002 | 0.027 | 0.97817 |
| Vowel i | 0.0008 | 0.002 | 0.321 | 0.74846 |
| Vowel o | 0.00000 | 0.003 | 0.000 | 0.99962 |
| Vowel u | 0.002 | 0.003 | 0.494 | 0.62124 |
| Sex Male | −0.002 | 0.001 | −3.128 | <0.01 |
| Context CVN# x Vowel e | −0.002 | 0.003 | −0.667 | 0.50486 |
| Context CV.N#V x Vowel e | 0.0031 | 0.004 | 0.843 | 0.39919 |
| Context CVCV x Vowel e | 0.0003 | 0.002 | 0.121 | 0.90404 |
| Context CVN# x Vowel i | 0.0003 | 0.004 | 0.073 | 0.94213 |
| Context CV.N#V x Vowel i | 0.0001 | 0.003 | 0.034 | 0.97307 |
| Context CVCV x Vowel i | −0.00008 | 0.002 | −0.036 | 0.97096 |
| Context CVN# x Vowel o | −0.0024 | 0.004 | −0.564 | 0.57271 |
| Context CV.N#V x Vowel o | −0.0003 | 0.004 | −0.075 | 0.93988 |
| Context CVCV x Vowel o | 0.0007 | 0.003 | 0.227 | 0.82066 |
| Context CVN# x Vowel u | −0.0009 | 0.005 | −0.183 | 0.85497 |
| Context CV.N#V x Vowel u | 0.0008 | 0.004 | 0.191 | 0.84872 |
| Context CVCV x Vowel u | −0.0005 | 0.003 | −0.170 | 0.86534 |

Deviance explained = 72.9%.

## Appendix C

Model output for carryover vowel nasalization (V2)
Formula: `nasal ~ Context + Sex + s(NormTime, by = Context) + s(NormTime, by = Sex) + s(NormTime, Speaker, by = Context, bs = ‘‘fs’’, m = 1) + s(Word, bs = ‘‘re’’)`

| Parametric Coefficients: | Estimate | SE | *t*-Value | *p*-Value |
|---|---|---|---|---|
| (Intercept) | 0.017 | 0.005 | 3.148 | <0.01 |
| Context CVCV | −0.0149 | 0.005 | −2.781 | <0.01 |
| Context CVN#V | 0.0018 | 0.007 | 0.239 | 0.81 |
| Sex Male | −0.0008 | 0.001 | −0.563 | 0.57 |

Deviance explained = 80.7%.

## Notes

1 IRB Protocol Number: 20071, University of Illinois Urbana-Champaign.

2 It is worth mentioning that plot_diff() curve differences can only conduct pair-wise comparisons.

3 The colors in this palette are color vision deficiency-friendly in the original version in color.

4 A remark should be made about the visible “negative” airflow results. As Beristain (2023a) pointed out, the equipment’s built-in electric voltage and DC offset could explain such results, or there might be some phonetic cue involving negative airflow and vocalic nasalization in Spanish, as such results only appeared in Spanish but not in other languages that were originally analyzed by Beristain (2022). Considering that the negative values are generally a part of the confidence intervals, it should not pose any significant alterations in the current results.

5 Using vowels with different height and a relatively small corpus may have contributed to a somewhat large standard error.

## References


Akaike, Hirotsugu. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19: 716–23. [CrossRef]

Beristain, Ander. 2022. The Acquisition of Acoustic and Aerodynamic Patterns of Coarticulation in Second and Heritage Languages. Ph.D. thesis, University of Illinois Urbana-Champaign, Champaign, IL, USA.

Beristain, Ander. 2023a. Gestural timing patterns of nasality in highly proficient Spanish learners of English: Aerodynamic evidence. *Language and Speech*. [CrossRef] [PubMed]

Beristain, Ander. 2023b. An airflow analysis of Spanish and English anticipatory vowel nasalization among heritage bilinguals. *Languages* 8: 205. [CrossRef]

BIOPAC Hardware Systems, Inc. 2020. *AcqKnowledge Data Acquisition and Analysis Software*. version 3.9.1. Goleta: BIOPAC Hardware Systems, Inc.

Bissiri, Maria P., Maria-Luisa García-Lecumberri, Martin Cooke, and Jan Volín. 2011. The role of word-initial glottal stops in recognizing English words. Paper presented at Interspeech 2011, Florence, Italy, August 27–31.

Boersma, Paul, and David Weenink. 2020. Praat: Doing Phonetics by Computer [Computer Program], Version 6.1.10. Available online: http://www.praat.org/ (accessed on 25 March 2020).

Bongiovanni, Silvina. 2021. Acoustic investigation of anticipatory vowel nasalization in a Caribbean and a non-Caribbean dialect of Spanish. *Linguistics Vanguard* 7: 1–11. [CrossRef]

Bradley, Travis G., Kimberly Morris, and Lillian Jones. 2022. Incomplete resyllabification and ambisyllabic gestural coupling in Spanish. *Loquens* 9: e094. [CrossRef]

Brown, Earl K. 2008. A Usage-Based Account of Syllable- and Word-Final /s/ Reduction in Four Dialects of Spanish. Ph.D. thesis, The University of New Mexico, Albuquerque, NM, USA.

Byrd, Dani, Stephen Tobin, Erik Bresch, and Shrikanth Narayanan. 2009. Timing effects of syllable structure and stress on nasals: A real-time MRI examination. *Journal of Phonetics* 37: 97–110. [CrossRef]

Carignan, Christopher. 2021. A practical method of estimating the time-varying degree of vowel nasalization from acoustic features. *Journal of the Acoustical Society of America* 149: 911–22. [CrossRef] [PubMed]

Clumeck, Harold. 1976. Patterns of soft palate movements in six languages. *Journal of Phonetics* 4: 337–51. [CrossRef]

Cohn, Abigail C. 1990. Phonetic and phonological rules of nasalization. *UCLA Working Papers* 76. Available online: https://escholarship.org/uc/item/1xq3d5hr (accessed on 20 December 2023).

Cohn, Abigail C. 1993. Nasalisation in English: Phonology or phonetics. *Phonology* 10: 43–81. [CrossRef]

Colina, Sonia. 1995. A Constraint-Based Analysis of Syllabification in Spanish, Catalan, and Galician. Ph.D. dissertation, University of Illinois at Urbana-Champaign, Champaign, IL, USA.

Colina, Sonia. 1997. Identity constraints and Spanish resyllabification. *Lingua* 103: 1–23. [CrossRef]

Colina, Sonia. 2002. Interdialectal variation in Spanish /s/ aspiration: The role of prosodic structure and output-to-output constraints. In *Structure, Meaning, and Acquisition in Spanish: Papers from the 4th Hispanic Linguistics Symposium*. Edited by James F. Lee, Kimberly L. Geeslin and J. Clancy Clements. Somervilla: Cascadilla Press, pp. 230–43.

Colina, Sonia. 2009. *Spanish Phonology: A Syllabic Perspective*. Washington, DC: Georgetown University Press.

Delvaux, Véronique, Didier Demolin, Bernard Harmegnies, and Alain Soquet. 2009. The aerodynamics of nasalization in French. *Journal of Phonetics* 36: 578–606. [CrossRef]

Diakoumakou, Evanthia. 2005. Coarticulatory nasalization in modern Greek: Evidence for a link between coarticulation and syllable structure. *Journal of the Acoustical Society of America* 117: 2459. [CrossRef]

Diakoumakou, Evanthia. 2009. *Coarticulatory Vowel Nasalization: Evidence from Modern Greek*. Riga: VDM Verlag.

Durand, Jacques, Bernard Laks, Basilio Calderone, and Atanas Tchobanov. 2011. Que savons-nous de la liaison aujourd'hui? [What do we know today about liaison]? *Langue Française* 169: 103–35. [CrossRef]

Goodin-Mayeda, C. Elizabeth. 2016. *Nasals and Nasalization in Spanish and Portuguese: Perception, Phonetics, and Phonology*. Amsterdam: John Benjamins. [CrossRef]

Hajek, John. 1997. *Universals of Sound Change in Nasalization*. Hoboken: Wiley-Blackwell.

Harris, James. 1983. *Syllable Structure and Stress in Spanish: A Nonlinear Analysis*. Cambridge, MA: MIT Press.

Harris, James, and Ellen M. Kaisse. 1999. Palatal vowels, glides and obstruents in Argentinian Spanish. *Phonology* 19: 117–90. [CrossRef]

Henderson, Janette B. 1984. Velopharyngeal Function in Oral and Nasal Vowels. Ph.D. thesis, University of Connecticut, Storrs, CT, USA.

Horiguchi, Satoshi, and Fredericka Bell-Berti. 1987. The Velotrace: A device for monitoring velar position. *Cleft Palate Journal* 24: 104–11. [PubMed]

Hualde, José I. 2014. *Los Sonidos del Español [The Sounds of Spanish]*. Cambridge: Cambridge University Press. [CrossRef]

Hualde, José I., and Jon Ortiz de Urbina. 2003. *A Grammar of Basque*. Berlin: De Gruyter Mouton. [CrossRef]

Hualde, José I., and Pilar Prieto. 2014. Lenition of sibilants in Catalan and Spanish. *Phonetica* 71: 109–27. [CrossRef] [PubMed]

Huffman, Marie K., and Rena A. Krakow, eds. 1993. *Nasals, Nasalization, and the Velum (Phonetics and Phonology V)*. San Diego: Academic Press. [CrossRef]

Jiménez-Bravo, Miguel, and José María Lahoz-Bengoechea. 2023. Durational cues to resyllabification in Spanish. *Loquens* 10: e099. [CrossRef]

Kahn, Daniel. 1976. Syllable-Based Generalizations in English Phonology. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA, USA.

Kaisse, Ellen M. 1999. Resyllabification precedes all segmental rules. In *Formal Perspectives on Romance Linguistics: Selected Papers from the 28th Linguistic Symposium on Romance Languages*. Edited by Jean-Marc Authier, Barbara E. Bullock and Lisa A. Reed. Amsterdam: John Benjamins, pp. 197–210.

Kochetov, Alexei. 2020. Research methods in articulatory phonetics II: Studying other gestures and recent trends. *Language and Linguistics Compass* 14: e12371. [CrossRef]

Kohler, Klaus J. 1994. Glottal stops and glottalization in German. *Phonetica* 51: 38–51. [CrossRef]

Krakow, Rena A. 1989. The Articulatory Organization of Syllables: A Kinematic Analysis of Labial and Velar Gestures. Ph.D. thesis, Yale University, New Haven, CT, USA.

Krakow, Rena A. 1993. Nonsegmental influences on velum movement patterns: Syllables, sentences, stress, and speaking rate. In *Nasals, Nasalization, and the Velum*. Edited by Marie K. Huffman and Rena A. Krakow. San Diego: Academic Press, pp. 87–116.

Krakow, Rena A. 1999. Articulatory organization of syllables: A review. *Journal of Phonetics* 27: 23–54. [CrossRef]

Lahoz-Bengoechea, José María, and Miguel Jiménez-Bravo. Forthcoming. Mismo grado de nasalización en VN# y V#N en español [Same degree of nasalization in VN# and V#N in Spanish]. In *Proceedings from the 9th International Conference on Experimental Phonetics (CIFE)*. Vigo: Universidad de Vigo, *in press*.

Martínez, Ruth M. 2021. The Bidirectional Acquisition of Oral and Nasal Gestural Timing in L2 Brazilian Portuguese and Rioplatense Spanish Nasal Structures. Ph.D. thesis, University of Toronto, Toronto, ON, USA.

MATLAB. 2020. *9.8.0.1323502. R2020a*. Natick: The MathWorks Inc.

Passino, Diana, Joaquim Brandão de Carvalho, and Tobias Scheer. 2022. Syllable structure and (re)syllabification. In *Manual of Romance Phonetics and Phonology*. Edited by Christoph Gabriel, Randall Gess and Trudel Meisenburg. Berlin and Boston: De Gruyter, pp. 89–126. [CrossRef]

R Core Team. 2020. R: A Language for Statistical Computing. Available online: https://www.R-project.org/ (accessed on 15 January 2024).

R Studio Team. 2020. RStudio: Integrated Development for R. Available online: http://www.rstudio.com/ (accessed on 15 January 2024).

Robinson, Kimball. 2012. The dialectology of syllabification: A review of variation in the Ecuadorian Highlands. *Romance Philology* 66: 115–45. [CrossRef]

Sampson, Rodney. 1999. *Nasal Vowel Evolution in Romance*. Oxford: Oxford University Press.

Sauzet, Patrick. 2012. Occitan plurals. A case for a morpheme-based morphology. In *Inflection and Word Formation in Romance Languages*. Edited by Sascha Gaglia and Marc-Olivier Hinzelin. Amsterdam and Philadelphia: Benjamins, pp. 179–200.

Shosted, Ryan K. 2009. *The Aeroacoustics of Nasalized Fricatives: An Instrumental Study in Phonetic Typology*. Saarbrucken: VDM Verlag.

Shosted, Ryan K., Christopher Carignan, and Panying Rong. 2012. Managing the distinctiveness of phonemic nasal vowels: Articulatory evidence from Hindi. *The Journal of Acoustical Society of America* 131: 455–65. [CrossRef] [PubMed]

Solé, Maria-Josep. 1992. Phonetic and phonological processes: The case of nasalization. *Language and Speech* 35: 29–43. [CrossRef]

Stoakes, Hywel M., Janet M. Fletcher, and Andrew R. Butcher. 2020. Nasal coarticulation in Bininj Kunwok: An aerodynamic analysis. *Journal of the International Phonetic Association* 50: 305–32. [CrossRef]

Strycharczuk, Patricia, and Martin Kohlberger. 2016. Resyllabification reconsidered: On the durational properties of word-final /s/ in Spanish. *Journal of the Association for Laboratory Phonology* 7: 3. [CrossRef]

Torreira, Francisco, and Mirjam Ernestus. 2012. Weakening of intervocalic /s/ in the Nijmegen Corpus of Casual Spanish. *Phonetica* 69: 124–48. [CrossRef]

Van Rij, Jacolien, Martijn Wieling, R. Harald Baayen, and Hedderik Van Rij. 2017. Itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs (R Package Version 2.3). Available online: https://cran.r-project.org/web/packages/itsadug/index.html (accessed on 15 January 2024).

Wood, Simon N. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)* 73: 3–36. [CrossRef]

Zuazo, Koldo. 2014. *Euskalkiak [The Basque Dialects]*. Donostia: Elkar.

*languages*

MDPI

*Article*

# Rhotic Variation in Brazilian Portuguese

**Michael Ramsammy [1,*] and Beatriz Raposo de Medeiros [2]**

[1] School of Philosophy, Psychology and Language Sciences, University of Edinburgh, Edinburgh EH8 9AD, UK
[2] Departamento de Linguística, Universidade de São Paulo, São Paulo CEP 05508-010, Brazil; biarm@usp.br
* Correspondence: m.ramsammy@ed.ac.uk

**Abstract:** We present acoustic and articulatory data from an experiment designed to test the phonetic variability of rhotics in Brazilian Portuguese, focusing on the São Paulo variety. Ultrasound tongue imaging was used to examine the realisation of rhotics in a range of phonological environments. Our analysis reveals that word-initial and intervocalic fricatives are acoustically and articulatorily distinct for most speakers. We attribute a tendency for utterance-initial fricatives to display longer duration, less voicing, and greater tongue-dorsum displacement than word-medial intervocalic counterparts to phonetic enhancement at the site of a major prosodic boundary. Similarly, rhotic taps in utterance-final position show a tendency for devoicing and frication (aspiration or assibilation) speaker-dependently. By comparison, word-medial pre-consonantal and intervocalic taps are characterised by shorter durations and greater voicing: hence, a pattern of phonetic reduction in prosodically weaker environments. We relate our findings to theoretical debates around the phonological status of rhotics in Portuguese. Whilst not providing conclusive proof in favour of any one particular approach, our results highlight the need to recognise the reality of prosodically driven strengthening in developing a full account of rhotic variation in the variety.

**Keywords:** Brazilian Portuguese; São Paulo; rhotics; variation; ultrasound; acoustics

## 1. Introduction

This article presents an analysis of rhotic variation in Brazilian Portuguese (BP), and specifically the variety spoken in São Paulo (i.e., the *paulista* variety). Our study draws on acoustic and articulatory data from an experiment using ultrasound tongue imaging (UTI). The contribution of this element of the study is descriptive in nature. We also make a contribution to theoretical debate around the theme of context-dependent articulatory strengthening and weakening as well as considering the broader impact of our results for phonological analysis of rhotic allophony in the language.

In this section, we begin by outlining some key facts about rhotic variation in Portuguese, both from phonetic and phonological perspectives. We then discuss previous work of relevance on the phonetic characteristics of rhotic consonants cross-linguistically.

### *1.1. Portuguese Rhotic Variation*

Portuguese as a language displays extensive cross-dialectal variation in the realisation of rhotic consonants. Brazilian dialects, more specifically, show patterns of variation that may depend on regional, social, and stylistic factors as well as phonological contextual factors. The data in Table 1 below illustrate the major patterns (see Câmara 1972; Cristófaro Silva 2003; Mateus and d'Andrade 2000; Rennicke 2015).

In environments (a–b), a tap occurs in almost all varieties of Portuguese, and taps occur most commonly in (c). Thus, dialectal variation principally affects environments (d–f) and (g–h). Intervocalically, orthographical <rr>, as in *carro*, is typically a fricative. Glottal realisations, which can include both [h] and a voiced variant, [ɦ], predominate in the São Paulo variety under investigation here. Velar [x], which may also occur with voicing (hence, [ɣ]), is characteristic of Rio de Janeiro, i.e., the *Carioca* variety. Uvular [ʁ] is the most wide-spread variant in urban dialects of European Portuguese (EP), whereas the alveolar trill occurs in a more restricted, typically rural, dialect space in the present day.

**Table 1.** Portuguese rhotic variation by phonological environment.

| Environment | Variant(s) | Examples |
|---|---|---|
| a. Intervocalic | [ɾ] | *caro* 'expensive' |
| b. $C_2$ in tautosyllabic cluster | [ɾ] | *prato* 'plate' |
| c. Word-final pre-vocalic | [ɾ] | *mar alto* 'high sea' |
| d. Intervocalic | [h], [x], [ʁ], [r] | *carro* 'car' |
| e. $C_2$ in heterosyllabic cluster | [h], [x], [ʁ], [r] | *Israel* 'Israel' |
| f. Word-initial pre-vocalic | [h], [x], [ʁ], [r] | *rato* 'rat' |
| g. Word-medial pre-consonantal | [h], [x], [ɹ], [ɻ], [ɾ] | *carta* 'letter' |
| h. Word-final pre-pausal | [h], [x], [ɹ], [ɻ], [ɾ] | *bar* 'bar' |

The same patterns of phonetic variation that affect intervocalic <rr> are also observed when rhotics occur as the second member of a heterosyllabic cluster. Words exhibiting such sequences are not numerous in the language, such that (e) in Table 1 can be considered a marginal pattern. Descriptions of Portuguese generally state that the same dialect-specific variants that occur intervocalically in environment (d) also occur word-initially in environment (f), at least with regard to place and manner of articulation. Voicing is more variable, with greater incidence of, e.g., [ɦ] and [ɣ], in intervocalic contexts (further details below).

Environments (g–h) in Table 1 admit variants that may or may not also occur in other environments. In EP, rhotics in canonical coda position are realised as taps. However, BP dialects display variation between fricative realisations like [h] or [x] that are, again, characteristic of Rio de Janeiro. Particular to São Paulo is the variable use of rhotic approximants (i.e., [ɹ] or [ɻ]) in pre-consonantal and pre-pausal coda contexts.

### *1.2. Descriptive and Theoretical Research on Portuguese Rhotics*

Regarding the phonetics of this variation, the most extensive study on rhotics in BP to date comes from Rennicke (2015). Rennicke studied rhotic productions in 14 speakers of BP from Lavras in the state of Minas Gerais using a combination of semi-structured interviews and a sentence-completion task. Due to its relative proximity to the state of São Paulo, dialects spoken in the south-west of Minas Gerais display certain similarities to the paulista variety: importantly, this includes the occurrence of rhotic approximants in coda position. Rennicke (2015, pp. 88–89) observes that speakers residing north of an isogloss line that bisects Minas Gerais favour fricative realisations of coda rhotics instead of approximants.

Following the taxonomy laid out in Cristófaro Silva (2003: further comments below), Rennicke employs a categorisation system that includes *strong-R* and *weak-R*—i.e., fricative realisations vs tap-like realisations, respectively. This relies on an impressionistic identification method to identify rhotic variants, which includes observation of voicing bars in the spectrogram to distinguish voiced from voiceless realisations (cf. Sebregts 2014 for Dutch). Fricative productions with observable high-energy noise were categorised as velar or uvular, whereas low-energy fricatives were categorised as glottal.

In accordance with these criteria, participants in Rennicke's study seemed to overwhelmingly favour [h] and [ɦ] in *strong-R* contexts, which comprise the word-initial and intervocalic environments in (d) and (f) in Table 1. Marginal occurrence of other variants, including [ʁ], [χ] and deletion, was also noted. Realisations of *weak-R* were more highly variable. In VRV-sequences, nearly all realisations are tap-like; however, Rennicke notes that a high proportion of taps are partly weakened. Both full [ɾ] and weakened [ɾ̞] also occur with high frequency in cluster contexts. Devoicing of taps is relatively common, particularly in post-tonic pre-pausal environments and following voiceless consonants. Other variants, such as approximants and rhotic deletion, occurred sporadically.

Rennicke treats coda rhotics separately from *strong*- and *weak-R*. The analysis reveals extensive variation in both word-medial and word-final coda contexts. In word-medial (pre-consonantal) codas, rhotic approximants are the dominant realisation. These include [ɹ], retroflected [ɻ], and an R-coloured vowel transcribed as [ɚ]. Acoustically, Rennicke distinguishes [ɹ] and [ɻ] through observation of F3: rhotic realisations displaying a shallow rise in the F3 trajectory are transcribed as [ɹ], whereas [ɻ]-productions are identified through a falling F3 contour over the VR sequence. Unlike [ɻ], [ɚ]-realisations are characterised by a flatter formant structure throughout the duration of the vowel, where the frequency of F3 is low and close to F2. Deletion is also fairly common in VRC contexts.

Word-finally, the situation is somewhat similar. Retroflex approximants are the most common variant across all contexts examined, particularly in pre-consonantal and prepausal contexts. Pre-vocalically, word-final rhotic realisations include [ɻ] and various tap-like productions (i.e., [ɾ], [ɾ̥], and [ɾ̞]). Word-final rhotic deletion is reasonably common, whereas other realisations occur with much lower frequency (e.g., glottal and aspirated variants).

Summing up the findings, Rennicke proposes that BP rhotic realisations span a gradient lenition trajectory comprising [ɾ] > [ɾ̞] > [ɹ] > [ɻ] > [ɚ]. A multitude of factors are argued to be responsible for the relative reduction of a given rhotic including phonological positional factors, as discussed above, as well as interactions with vowel quality, lexically specific variation, and stylistic factors, such as emphatic or controlled speech patterns. Rennicke also makes the point that a fuller understanding of some of the patterns of variation would only be possible with further work of an articulatory nature: it is to this need that the current paper aims to make an initial response.

In addition to Rennicke's work, rhotic variation has been the subject of variationist sociolinguistic research. Oushiro and Mendes (2013) studied rhotic productions in a 1.5-million-word spoken corpus of paulistano Portuguese (see also Ricardo and Schwindt 2023 on rhotics in Porto Alegre BP). Focusing on coda rhotics, the authors first draw attention to geographical patterns, with coda [h, ɦ] having associations with Belo Horizonte (state capital of Minas Gerais)[1] and [x, ɣ] strongly indexing Carioca speech, as already mentioned. For São Paulo, a distinction is made between coda [ɾ], which is characteristic of normative, educated city speech, and use of coda [ɻ], which is expected to occur in more rural areas and in the speech of less highly educated individuals. Relatedly, Rennicke (2015) and Oushiro (2021), inter alia, draw attention to the relative stigmatised status of [ɻ].

As well as studies with an empirical focus, Portuguese rhotics have been the subject of theoretical analysis. Mateus and d'Andrade (2000) approach rhotic variation as a case of positional allophony. They make a fundamental claim in this regard, namely, that all surface rhotics derive synchronically from a single phoneme, /ɾ/. Accordingly, intervocalic dorsal fricatives (as in *carro*) are analysed as deriving from an underlying geminate (i.e., /ɾɾ/). In this case, a two-stage derivation first syllabifies instances of /ɾɾ/ heterosyllabically and then assigns [DORSAL] place to the onset: hence, /ɾɾ/→/ɾ.ʁ/ (referring to EP). Secondly, the coda /ɾ/ deletes, thereby generating [ʁ]. On the surface, this [ʁ] is therefore representationally indistinguishable from word-initial [ʁ], which is derived by a simpler [DORSAL]-default rule that applies syllable-initially, i.e., [$_{\sigma}$ɾ→[ʁ].[2]

Cristófaro Silva (2003) takes a different approach that aims to account for rhotic patterns in a range of dialects on the basis of a set of shared representations. Unlike Mateus and d'Andrade's monophonemic approach,[3] Cristófaro Silva assumes a three-way lexical contrast, namely, between a phonemic tap and two further rhotic phonemes notated as /R̄/ and /R/, respectively (i.e., the *weak*, *strong*, and *coda Rs* referred to by Rennicke 2015). This approach is perhaps best suited to dialects exhibiting a three-way surface pattern, such as the São Paulo dialect. Assuming the occurrence of intervocalic alveolar trills in this variety (rather than glottal fricatives), Cristófaro Silva posits the rules listed in Table 2.

**Table 2.** Rhotic allophony; adapted from Cristófaro Silva (2003, pp. 142–43).

| | | | | | |
|---|---|---|---|---|---|
| a. | /R̄/ | *'Strong-R'* | → | [r] | *Word-initial, intervocalic, and post-consonantal contexts* |
| b. | /R/ | *'Coda-R'* | → | [ɹ] | *Coda contexts* |
| c. | /ɾ/ | *'Weak-R'* | → | [ɾ] | *Onset clusters and intervocalically* |

On the one hand, a possible objection to this characterisation of the facts is its lack of predictive power. That is to say that deriving the surface forms from a three-way underlying contrast, the distribution of surface variants relies exclusively on the asymmetric patterning of /ɾ, R̄, R/ at an underlying level. On the other hand, a potential advantage of this analysis is that it does not rely on simplification of putative lexical /ɾɾ/-sequences to generate intervocalic fricatives.

### *1.3. Articulatory Studies on Rhotic Consonants*

Phonetic studies—and particularly articulatory studies—on Portuguese have generally focused on phenomena other than rhotic allophony: e.g., the laterals, obstruents and oral/ nasal vowel contrasts (Barlaz et al. 2018; Charles and Lulich 2018; Howson et al. 2022; Martins et al. 2008, inter alia). However, one previous UTI study on BP rhotics by Howson and Kochetov (2018) is particularly relevant in regard of the aims of the current research. Howson and Kochetov (2018) studied rhotic fricatives and taps in four vowel contexts in the speech of six speakers of BP from São Paulo. They found evidence that both types of rhotic ([ʁ] and [ɾ] in their transcriptions) are sensitive to coarticulation with surrounding vowels, with particularly strong effects observed for /i/. Phonological context also exerted an effect on the articulation of the rhotic fricatives, with greater variability in tongue position in intervocalic <rr> as compared to word-initial fricatives in corresponding vowel environments. The authors therefore suggest that, in comparison to word-initial [ʁ], the intervocalic fricative lacks an oral place of articulation.

Furthermore, research on other languages provides a relevant point of comparison for investigation of rhotic variation in BP. For example, Recasens and Pallarès (1999) studied coarticulatory patterns involving /r/ and /ɾ/ in Catalan using a combination of electropalatography (EPG) and acoustic methods. Across both vowel contexts studied (/i_i/ and /a_a/), they found a significant difference in F2, which reached a higher frequency with the tap than the trill. This is attributed to increased dorsal activity in the latter. Durational differences were also noted, with rhotic articulations in the /i_i/ environment tending to be longer than in /a_a/. Regarding articulation, both trills and taps were articulated at a more anterior location on the palate in /i_i/ than in /a_a/. The authors further argue that achieving a specific dorsal target for the production of the trill makes it more resistant to coarticulation with surrounding vowels than the tap, particularly in /i/ contexts.

In a follow-up EPG study on Majorcan and Valencian Catalan, Recasens and Espinosa (2007) examined trill and tap realisations in a broader range of phonological contexts and vowel environments. Evidence for somewhat stronger trill productions is reported for pre-vocalic as opposed to intervocalic contexts (i.e., larger and longer context periods in word-initial /r/ than in word-medial intervocalic /r/: see Baltazani and Nicolaidis 2013 for a similar result from an EPG study on Greek). Varying qualities of the pre- and post-rhotic vowels produce complex effects on the duration and contact profiles of /r/ and /ɾ/. Dialectal differences in the degree of dorsum retraction also occur in the data, for example, the tendency for coda tap-*reinforcement* in Eastern Catalan. Valencian speakers, on the other hand, tended to have weaker articulations, particularly in syllable-final environments. Context-dependently, rhotic devoicing and the occurrence of epenthetic vowels following tap-realisations are also reported, although these phenomena are strongly speaker-specific.

In addition to EPG, UTI has been applied to studying rhotic articulations. Proctor (2011) presents a comparative analysis of rhotics in Russian and Spanish. For Spanish, articulatory tracking reveals fine-grained articulatory differentiation of [ɾ] and [r]. Proctor (2011) highlights that both rhotics in Spanish have a dorsal target (see also Boyce et al. 2016). For the tap, this is described as being similar to a schwa-like mid-central vowel. The trill, by contrast, is characterised by a retracted dorsum and an [o]-like tongue-body configuration. Comparisons of Russian plain and palatalised trills similarly confirm that [r] has a dorsal target that resembles a mid-central vowel (hence, a similarity to Spanish [ɾ]). [rʲ] has a more fronted articulation than the plain trill due to the influence of the secondary palatalisation gesture.

English rhotics have been studied extensively using articulatory methods. For example, Campbell et al. (2010) studied the production of Canadian English /ɹ/ using lip-movement tracking and ultrasound. They found that the timing of lip, tongue-body, and tongue-root gestures varied according to the syllabic position in which the /ɹ/ occurred. In word-initial /ɹ/ contexts, labial and tongue-body movements were produced earlier and with greater overall magnitude than the radical gesture. Although with less consistency, this pattern was reversed when /ɹ/ occurred in word-final contexts (pre-vocalic and pre-consonantal).

Zhou et al. (2008) examined North American English /ɹ/ using a combination of MRI and acoustic methods. They note that /ɹ/ is produced with two main tongue configurations, either with bunching of the tongue body or retroflexion of the tongue tip (i.e., tip-down vs tip-up realisations). Acoustically, similar to what Rennicke (2015) and Sebregts (2014) describe for Portuguese and Dutch, respectively, both types of /ɹ/ are characterised by a lowering of F3, such that the frequency of F3 comes very close to F2. Computer modelling of the vocal tract based on MRI images from bunched and retroflex realisations reveals a more obvious acoustic differentiation at higher frequencies, specifically F4 and F5.

Acquisitional studies of North American /ɹ/ have provided further insights into the bunched/retroflex distinction. Klein et al. (2013) show that speech therapeutic interventions designed to correct misarticulated /ɹ/ can result in accurate productions that may be either bunched or retroflex (cf. Tiede et al. 2010). They suggest that a preference for one tongue shape or the other may partially reflect the morphology of the vocal tract in children. They also point out that consonantal /ɹ/ and vocalic /ɹ/ (i.e., realisations that are more /ɚ/-like) are acquired differently, and that adult populations with typically developed speech show an array of tongue shapes in /ɹ/ that sometimes do not neatly fit into the bunched or retroflex categories (Delattre and Freeman 1968). Similarly, on the basis of a UTI-study involving four children acquiring Canadian English, Magloughlin (2016) argues that children go through an exploratory period with bunched and retroflex /ɹ/ before settling on a dominant adult pattern.

Further investigation into the articulatory properties of American English rhotics has added to this picture. In a UTI study involving 27 speakers of American English, Mielke et al. (2016) report that some speakers use only a single rhotic variant, whereas others show what is described as "idiosyncratic allophony", i.e., a pattern of alternation between bunched /ɹ/ and retroflex /ɻ/ that is dependent on contextual factors. The variable speakers produced retroflex rhotics most commonly in word-initial pre-vocalic position and in the context of /a/ and /o/. With /i/ and in word-final postvocalic position, /ɹ/ tended to be bunched. The fact that some patterns of alternation between the two rhotic types are clearly speaker-specific rather than general is attributed to perceptual factors: the difficulty in distinguishing perceptually between /ɹ/ and /ɻ/ leads to an inconsistency in the use of either variant in specific environments for some speakers.

Dediu and Moisik (2019) further discuss the hypothesis that vocal-tract anatomy may contribute to variation in the use of bunched vs retroflex rhotics, as mentioned in Klein et al. (2013). Using MRI, the authors studied rhotic productions among 80 participants, the majority of whom were L2 speakers of American English. The results support the generalisation that whereas the bunched and retroflex variants have highly similar, if not perceptually indistinguishable, acoustic profiles, the two configurations may show different coarticulatory behaviour. With respect to their main research question, Dediu and Moisik (2019) also argue that anatomical differences between participants principally affect tongue bracing. This is relevant for retroflex articulations, which require a greater degree of radical-pharyngeal bracing than bunched /ɹ/. Thus, speakers with narrower vocal tracts may favour bunched realisations over retroflex ones.

Rhotic variation in English varieties and languages other than North American English has also been studied. Heyne et al. (2020) collected UTI data from 62 speakers with the aim of ascertaining whether similar variation between bunched and retroflex articulations occurs in New Zealand English. They showed that some speakers do show a pattern of categorical alternation between /ɹ/ and /ɻ/. Similar to the results of Mielke et al. (2016), these speakers favoured use of the retroflex /ɻ/ in word-initial position and in back-vowel contexts. Low central vowels also patterned with tip-up articulations in initial and intervocalic positions. Word-finally, and in the context of alveolar and velar consonants, tip-down articulations were more commonly observed. There were also non-alternating speakers, 25 of whom consistently used bunched articulations and 12 of whom consistently used a retroflected tongue configuration. In general, the authors concluded that the same patterns of variation reported for North American English occur in New Zealand, but that bunched tip-down articulations are generally less common.

Additionally, rhotic productions produced by English speakers from a variety of locations in England were studied using UTI by King and Ferragne (2020). Like New Zealand English, the authors note that tip-up retroflex articulations are much more common in English than American speech. Of the 29 speakers recorded, the majority favoured retroflex rhotics, whereas 7 speakers produced a higher proportion of bunched articulations, and 3 used a mixture of the two types of /ɹ/. Uniquely among studies looking at English /ɹ/, King and Ferragne (2020) also examined patterns of lip configuration in addition to tongue shape. They confirm that bunched /ɹ/ is produced with greater lip protrusion than retroflex /ɻ/; and interestingly, lip protrusion in /ɹ/ does not correlate directly with rounding on flanking vowels. The authors attribute this to a compensation strategy that arises due to bunched /ɹ/ having a smaller front cavity than retroflex /ɻ/.

Regarding Scottish English, Lawson et al. (2011) discuss variation in approximant realisations that include tip-up, front-up, and bunched articulations. These articulations form a socially stratified continuum that is not dissimilar to the variation described for the paulista rhotics. In Scottish English, middle-class speakers favour realisations in which the post-dorsum has visible inward bunching: a similar feature is observed for Canadian French rhoticised vowels in Mielke (2015). The working-class participants in Lawson et al. (2011) more regularly produced tip-/front-up articulations lacking observable bunching. Lawson et al. (2014) further note that blending of approximant-R with a preceding vowel, resulting in a rhoticised vowel, is a regular feature of middle-class speech. Acoustically, these realisations are characterised by a progressively falling F3, as noted by Rennicke (2015) for some BP realisations. Lawson et al. (2013) also find evidence for some working-class speakers producing rhotics with delayed tongue-tip raising, which results in the gestural maximum occurring within a period of voicelessness. These realisations also lack the obvious F3 drop that characterises approximant and [ɚ]-realisations.

In a study on Dutch, Strycharczuk and Sebregts (2018) examined allophonic alternation of rhotic consonants in three phonological environments, namely, word-initially, word-finally, and in a "fake-geminate" /-r#r-/ external-sandhi context. Similar to BP, Standard Dutch exhibits a uvular rhotic in onset position and a coda approximant, the latter of which can be produced with tongue bunching or retroflexion. These patterns were confirmed by ultrasound data in Strycharczuk and Sebregts (2018), in which evidence of coda-R reduction was also found. In the sandhi context, some speakers favoured [ʁ]-like articulations that strongly resemble the onset allophone. Other speakers, and in particular those who showed a tendency to reduce coda-R, produced a fake-geminate R that has an intermediate configuration between [ʁ] and the coda approximant realisation.

Articulatory research on non-European languages is less common. However, the results of two studies are particularly relevant. Dravidian as a language family is well-known for displaying a high number of phonological contrasts in place of articulation, in sonorants as well as in obstruents. Scobbie et al. (2013) report UTI data from a single speaker of

Malayalam, which includes realisations of [ɾ], [r], [l], and two retroflex sonorants: [ɭ] and [ʐ]. [ɭ] is characterised by extensive raising and retroflexion of the tongue tip (cf. comparisons of geminate [tː] vs retroflex [ʈː] reported in Kochetov et al. 2014 for the related Dravidian language, Kannada). The same is visible to a certain degree in tongue traces from [ʐ]; however, the degree of retroflexion is observably smaller than in [ɭ], both in terms of bunching in the front body and arching of the tongue tip. Clear differences in the tongue configuration for [ɾ] and [r] are also noted, with the trill displaying retraction at the root and a flatter front-body position.

Tabain and Beare (2018) studied the alveolar∼retroflex contrast in Arrernte (Pama Nyungan, Australia). Regarding the rhotics, Arrernte has both an alveolar tap and a retroflex approximant. UTI data show that [ɾ] and [ɻ] are not consistently differentiated in spatial terms for all speakers examined. Three of the six speakers studied articulated [ɾ] with the tongue root in a more retracted position than [ɻ]. All speakers produced the retroflex with the front body of the tongue in a higher configuration (as measured at the temporal midpoint of each rhotic); however, the difference was more extreme for some speakers than others. Tabain and Beare (2018) draw attention to a small hollowing or bunching in the front-body area of the tongue that resembles the Malayalam [ʐ], and perhaps also its alveolar trill (cf. Scobbie et al. 2013, p. 110). This occurs in both [ɾ] and [ɻ], which may suggest that some Arrernte speakers retroflect the tap as well as the approximant to some extent.

Huang et al. (2024) examined tongue and lip configurations in Southern Mandarin rhoticised schwas. Comparisons focus on phonemic /ɚ/ (i.e., /ɚ/ lacking any kind of morphological status) and suffixal /ɚ/ (i.e., /ɚ/ that fulfilled a diminutive derivational function historically). Despite descriptions that describe it as a retroflex vowel, Huang et al. (2024) find no evidence for tip-up realisations in either type of /ɚ/. Instead, they describe two subtypes of tip-down bunched realisations based on EMA and UTI recordings, namely, dorsum-up and dorsum-down. Dorsum-up articulations show some degree of tongue retraction, whereas dorsum-down realisations are characterised by a convex tongue shape. Small articulatory differences between morphemic and non-morphemic /ɚ/ are observed, and the authors further note on the basis of acoustic measures that non-morphemic /ɚ/ has a tendency to show a higher F1 than its morphemic counterpart. They further suggest that suffixal /ɚ/ is produced with some degree of diphthongisation. Interestingly, there are also differences in lip configuration, where realisations of the suffix /ɚ/ do not show increased lip protrusion that differentiates them from plain (i.e., non-rhoticised) vowels. By contrast, non-morphemic /ɚ/ displays similar degrees of lip protrusion to the monophthong /o/.

UTI has also been applied to studying rhotic productions among Standard Mandarin speakers in an L2 context. Chen et al. (2024) examined the articulation of English /ɹ/ in a group of Mandarin–English bilinguals with varying levels of proficiency in English. In addition to patterns of acoustic variation, Chen et al. (2024) show that the most proficient English speakers in the sample varied categorically: they used a bunched tongue configuration in articulating Mandarin /ɹ/ and switched to a retroflex articulation when speaking English. The Mandarin bunched /ɹ/ interestingly displays clear similarities to the bunched configurations observed for middle-class Scottish English in the studies cited above. By contrast, the English /ɹ/ that highly proficient Mandarin–English bilinguals used in Chen et al. (2024) is generally characterised by a higher front-body position than bunched /ɹ/. Less proficient speakers showed some tendency to use a retroflected articulation when speaking English, although this occurred only variably. The least proficient English speakers used a bunched Mandarin-like articulation consistently, i.e., when producing speech in both Mandarin and English.

### *1.4. The Current Study*

In light of the complexities that BP rhotics present and the current state of knowledge about the phonetics of rhotics cross-linguistically, as summarised above, this study has two main aims. Firstly, we ask to what extent detailed phonetic study of Portuguese rhotics confirms the occurrence of contextual patterns previously reported on the basis of impressionistic observation. Secondly, we aim to identify and describe fine-grained patterns of articulatory variation, particularly in view of previous findings about contextual reduction in BP (and other languages like Catalan and Dutch). Additionally, we aim to relate the experimental findings to theoretical claims about the operation of allophony, as described in Section 1.2.

The rest of the paper is structured as follows. Section 2 below presents our methodology. The results are then presented in Section 3. Section 3.1 considers the acoustic properties of rhotics in Brazilian Portuguese. We move on to discuss the articulatory properties of these rhotics in Section 3.2. Section 4 discusses the impact of our findings for our specific research questions. Section 5 concludes the paper.

## 2. Methods

### *2.1. Participants*

Seventeen L1 speakers of BP from São Paulo participated in the study. Ten self-reported their gender as female, and seven as male. Ethical approval for the study was obtained from the Ethics Committee at the University of Edinburgh prior to speaker recruitment. Participants responded to an advertisement posted at the University of São Paulo to volunteer for the study. The minimum level of education for the participants was therefore undergraduate degree level. Full informed consent of participants was sought prior to data collection.

Recordings took place in a soundproof laboratory at the University of São Paulo. None of the participants reported any diagnosed hearing or speech pathologies prior to taking part. Due to poor ultrasound image quality, data from three speakers were excluded from the analysis. Our analysis is therefore based on data from 14 speakers. This represents a convenience sample of participants who volunteered for the study with a reasonable binary gender split (8 females, whom we refer to as BPF1–BPF8, and 6 males, referred to as BPM1–BPM6).

### *2.2. Recording Setup*

Our data were collected using an Echoblaster 128 ultrasound machine and the Articulate Assistant Advanced (AAA) software package (Wrench 2003–2023). The ultrasound data were recorded with an average rate of 76 frames per second and with simultaneous audio recording sampled at 22.1 kHz using a Sony ECM-55B microphone. The microphone was attached to an ultrasound headset during recording. Participants wore the headset for the duration of the experiment. The headset holds the ultrasound probe in position under the chin during recording and allows for adjustment of the probe angle and placement to suit the physiology of individual participants.[4] An ultrasound probe with a depth of 70 mm and a 123.8° field of view was used for the experiment.

### *2.3. Materials*

During the experiment, participants were asked to read a set of sentences from a computer screen. The stimuli were programmed into AAA prior to data collection and were presented to participants in a randomised order. Table 3 below lists the target words that form the basis of our analysis. These words were embedded into carrier phrases for the purpose of examining contextual variation in rhotic production. The carrier phrases are presented in Table 4.

All the words listed in Table 3 are real lexical words. They were chosen specifically due to the occurrence of /a/ immediately adjacent to the target rhotic in each environment. Furthermore, non-rhotic consonants occurring within the target words were limited to labials and labio-dentals (e.g., *ra**m**o*, ***f**arra*, ***b**ar**b**a*, etc.).[5] This set of consonants was chosen so as to minimise lingual coarticulation with the rhotic productions that are the focus of the study. All words except for the monosyllabic *bar*, *mar*, and *par*, and trisyllabic *apara* [ɐ.ˈpa.ɾɐ] are disyllables with trochaic stress.

**Table 3.** Experimental test words by context.

| | Test Contexts | | Target Words | Gloss |
|---|---|---|---|---|
| a. | Word-initial fricatives | ##R & V#RV | *rabo*<br>*ramo*<br>*rapa* | 'tail'<br>'branch'<br>'scrape.IMPER' |
| b. | Word-medial fricatives | VRRV | *barra*<br>*farra*<br>*marra* | 'bar'<br>'fun, a laugh'<br>'unwillingly' |
| c. | Word-medial onset taps | VRV | *apara*<br>*para*<br>*vara* | 'chip'<br>'stop.IMPER'<br>'stick' |
| d. | Word-medial coda taps/approximants | VRC | *arma*<br>*barba*<br>*harpa* | 'weapon'<br>'beard'<br>'harp' |
| e. | Word-final taps/approximants | R## & VR#V | *bar*<br>*mar*<br>*par* | 'bar'<br>'sea'<br>'even' |

**Table 4.** Carrier sentences used in the experiment.

| | Test Environments | Carrier Phrase | Gloss |
|---|---|---|---|
| a. | $[_{Utt}\,[_{Wrd}R–$ | _____ *foi a palavra que disse.* | 'The word that I said was _____.' |
| b. | $[_{Utt} \ldots [_{Wrd}–R–] \ldots ]$ | *Digo* _____ *outra vez.* | 'I say _____ once again.' |
| c. | $–R_{Wrd}]\,_{Utt}]$ | *Outra vez, digo* _____. | 'Once again, I say _____.' |

The items in (b–d) in Table 3 were embedded into carrier sentence (b) exclusively. The words in Table 3 (a) were placed into carrier sentence (a) for the purpose of testing utterance-initial realisations (##R), and into sentence (b) for testing utterance-medial word-initial realisations (V#RV). Similarly, the items in Table 3 (e) were placed into frame sentence (b) to test word-final pre-vocalic rhotic realisations (VR#V), and into sentence (c) for testing word-final pre-pausal realisations (R##). Participants read each word-phrase combination twice (=six tokens per speaker per test context). They also read other sentences in the same format that did not contain target words with rhotic consonants. The number of additional phrases exceeded the target ones at a greater than 2:1 ratio. Thus, the non-target phrases functioned as distractors and prevented participants from fixating on rhotic productions in the experiment. Allowing some time for calibration of the equipment for each speaker, recording took approximately 30 minutes per participant.

### *2.4. Measurements and Analysis*

Prior to analysing the UTI data, audio recordings were extracted from AAA and exported to Praat (Boersma and Weenink 1992–2023). A small number of tokens containing disfluencies or reading errors were discarded (see Table 5). The remaining tokens were hand-segmented and subjected to acoustic analysis: some examples are shown in Section 3.1.3. Fricative realisations were segmented according to the occurrence of frication in the waveform and spectrogram. The segment boundaries for rhotic taps were placed where we observed a visible reduction in amplitude that typically coincided with obvious discontinuities in the established waveform patterns, corresponding to flanking vowels. Rhotic approximants occurred in prepausal and pre-consonantal environments, speaker-dependently. The offset boundary for approximants was placed where there was a clear loss of acoustic energy (coinciding either with an utterance break or consonantal closure). The onset boundary was sometimes more difficult to place due to gradience in transitions from preceding vowels. In these cases, we relied on observation of changes, typically loss of amplitude, of the upper formants and established waveform patterns. Where necessary, we also examined tongue movements in the UTI recordings so as to make a best estimate of the onset segment boundary.

We calculated the duration of each rhotic on the basis of the acoustically segmented material by script. The script also extracted pitch and formant-frequency readings. We measured $f_0$ at 11 equidistant temporal points in the longer rhotic realisations: the fricatives, approximants, and fricated taps. This method was not practical for the simple taps due to their short duration. Accordingly, we extracted pitch measurements at 3 points in [ɾ]-tokens (corresponding to the acoustic onset, the midpoint, and the acoustic offset). We determined voicing to be present in the signal at any measurement point where pitch values were extractable.

Values in Hertz were recoded using binary values: 1 indicated the presence of voicing at a given measurement point and 0 indicated the absence of voicing. These values were used to calculate a *voice-ratio* measurement for each token (see Ramsammy and Strycharczuk (2016) on similar use of voice ratio in the analysis of fricatives). Voice ratio is calculated by summing the binary-coded values and dividing by 11 in the case of the fricative and approximant rhotic realisations, and dividing by 3 in the case of [ɾ]-tokens. This yields values in a continuous range between 0 and 1, in which 0 represents complete absence of voicing and 1 indicates full voicing throughout the duration of consonant. The duration and voice-ratio data were modelled with mixed-effects linear regression using the *lmer* and *lmerTest* (Kuznetsova et al. 2020) functions in *R* (R Core Team 2018). Post hoc comparisons were conducted using estimated marginal means with conservative Bonferroni corrections (*emmeans* package: Lenth 2023).

Formant-frequency measurements were extracted for a subset of the data that included only rhotic approximants. We aimed to track movement in formant contours over time in order to visualise the occurrence of expected acoustic correlates of rhotic retroflexion. We took F1, F2, and F3 readings at 11 equidistant points over the full acoustic duration of each approximant token, i.e., the same points at which pitch values were also calculated. As we were particularly interested in tracing F3 over the vowel–rhotic transitional phases, we included two additional measurements before the acoustic onset of the rhotic: these were equidistant to the 11 measurements taken during the rhotic realisation. The full range of formant-frequency measurements therefore span the range $-0.2$ to 1 in normalised time (where values extracted at $-0.2$ and $-0.1$ correspond to the final periods of the [a] preceding the rhotic, 0 corresponds to the rhotic onset, and 1 corresponds to the rhotic offset).

The script was programmed to calculate 5 formants below 5 kHz with a 25 ms window length: we found this to be the ideal range for accurately measuring rhotic approximants produced by both male and female speakers. We normalised the raw data on a speaker-by-speaker basis using *z*-transform. These data were then submitted analysis using GAMs (*mgcv* package: Wood 2023), in which phonological context and normalised time were fixed predictors (see Section 3.1.2). Plots were generated using *ggplot2* (Wickham et al. 2023).

Following acoustic analysis, the annotation files were re-imported into AAA for the purpose of analysing the UTI data. Initial traces of the tongue surface were generated for all ultrasound frames in the recordings using the batch process function in AAA. Automatically generated traces occurring within the boundaries of pre-annotated rhotic realisations were hand-corrected where necessary. We then selected individual ultrasound frames that coincided most closely with the temporal midpoint of each annotation and extracted *xy*-coordinate values from a maximum of 42 points along the tongue-surface contour in those frames. We determined the acoustic midpoint to be the most suitable measurement location as this allows for maximal comparability between short [ɾ] and longer [h]- and [ɹ]-productions.

To partially correct for variation in the positioning of the ultrasound probe during recording for each participant, *x*- and *y*-values were scaled and centred at zero. This increased the visual comparability between speakers in plots. We calculated fitted splines on a speaker-by-speaker basis using GAMs. As per usual practice in ultrasound studies using the SS-ANOVA method (e.g., Davidson 2006; Howson and Kochetov 2018; Kochetov et al. 2014; Strycharczuk and Sebregts 2018, etc.), the fitted splines presented in Section 3.2 are plotted with 95% confidence intervals with the tongue tip to the right.[6] Given that the rhotic productions that we included in the ultrasound analysis were recorded in a single vowel context (i.e., with flanking /a/), we did not encounter issues in fitting GAMs using Cartesian coordinates (cf. Mielke 2015).

## 3. Results

### 3.1. Acoustic Analysis

#### 3.1.1. Overview

In this section, we present the results of the acoustic analysis of rhotic realisations. First and foremost, the high degree of inter- and intra-speaker variation in rhotic productions in syllable-final test contexts must be highlighted. Table 5 lists realisations per context based on broad categorical labels (approximant, fricative, tap, and trill). Word-initially, all speakers produced fricative realisations both in (a) the utterance-initial environment and in (b) the word-initial intervocalic environment. All speakers also consistently produced fricatives in (c) the word-medial <rr> environment. Some variation in voicing was observed across contexts (a–c): we discuss this further in Section 3.1.2. As expected, we observed only taps in (d) the word-medial VRV environment.

**Table 5.** Variation in rhotic productions across test contexts.

| Test Context | Realisation | N | % Total |
|---|---|---|---|
| a. ##R | fricative | 82 | 100 |
| b. V#RV | fricative | 84 | 100 |
| c. VRRV | fricative | 84 | 100 |
| d. VRV | tap | 84 | 100 |
| e. R## | approximant | 28 | 33 |
| | fricative | 6 | 7 |
| | tap | 45 | 54 |
| | trill | 5 | 6 |
| f. VR#V | approximant | 9 | 11 |
| | fricative | 0 | 0 |
| | tap | 67 | 81 |
| | trill | 7 | 8 |
| g. VRC | approximant | 21 | 26 |
| | fricative | 3 | 4 |
| | tap | 51 | 64 |
| | trill | 5 | 6 |

In the coda contexts, tapped realisations also predominated. One speaker, BPM1, produced only [h] utterance-finally, but used a tap in the word-final intervocalic environment. BPM1 also produced three pre-consonantal tokens with [h] but otherwise used an approximant in this context. Another speaker, BPF1, favoured the use of alveolar trills in the majority of coda-R realisations, but also produced taps in the utterance-final context. BPF4 articulated a single alveolar trill in the utterance-final context, but otherwise used only taps in this environment. Broadly speaking, all other variation in coda rhotics constitutes a distinction between approximant and tap realisations.

With regard to individual speakers, speakers BPF2, BPF4, BPF6, BPF7, BPF8, BPM1, BPM3, and BPM6 are *binary* speakers, according to our initial categorisation. This means that they produced exclusively fricatives in environments (a–c) in Table 5 and exclusively taps in environments (d–g).[7] The remaining participants, i.e., BPF3, BPF5, BPM2, BPM4, and BPM5, are *ternary* speakers who produce exclusively fricatives across contexts (a–c), exclusively taps in (d), and a mixture of taps and approximants in (e–g). None of the participants in this study produced rhotic approximants without also producing taps in at least one of the coda test contexts. Speaker BPM5 is the only ternary speaker who showed no within-context variation. He produced only approximants in utterance-final and pre-consonantal coda contexts and only [ɾ] in the word-final pre-vocalic environment.

### 3.1.2. Fricative and Approximant Realisations

We now turn to consider the non-tapped rhotic realisations in more detail (the tap data are presented in Section 3.1.3). Figure 1 below illustrates the durational variation in fricative and approximant rhotic realisations by test context. Recall here that fricative realisations account for 100% of the data in ##R, V#RV, and VRRV environments, as per Table 5. By contrast, the approximants that occur in R##, VR#V and VRC account for only a subset of realisations in these environments.

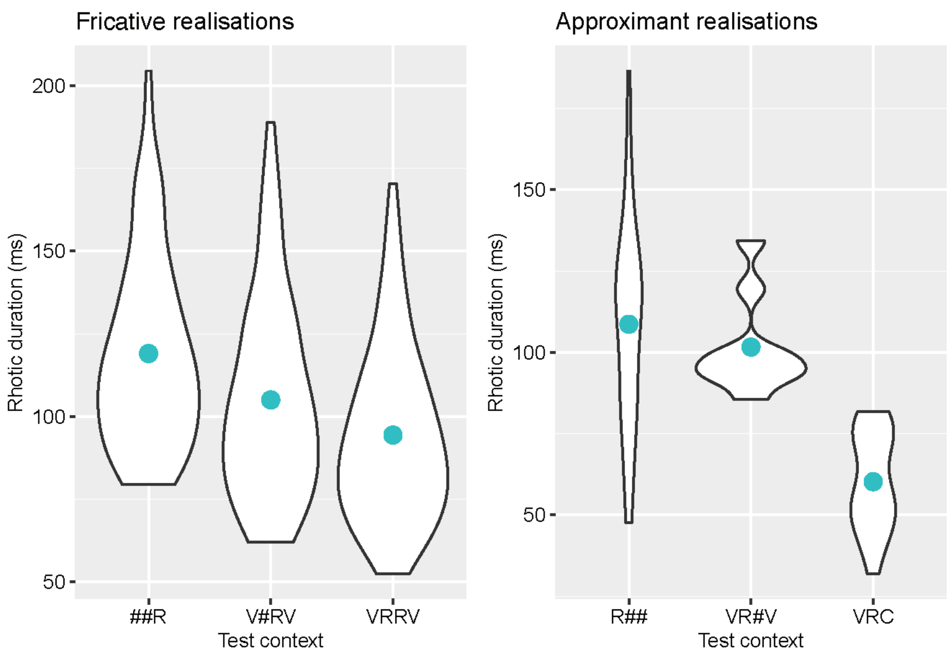



**Figure 1.** Duration of rhotic realisations in ms across test contexts (taps excluded). Dots represent mean values.

In the fricatives, we observe a progressive decline in duration across the three phonological contexts in which they occur. The longest fricatives occur utterance-initially (mean duration = 119 ms). In the word-initial utterance-medial environment, this drops to 105 ms. Word-medial intervocalic fricatives are the shortest, with a mean duration of 94.3 ms. Similarly, although there is considerable variation,[8] utterance-final realisations of [ɹ] have an average duration of 109 ms. The mean duration of utterance-medial word-final [ɹ] is 102 ms with the word-medial pre-consonantal approximants having a much shorter mean duration of 60.2 ms. Overall, these values indicate that both types of rhotic are produced with the greatest duration at the periphery of the utterance. Rhotics that occur immediately before or after a grammatical word boundary in an intervocalic context are somewhat shorter. Rhotics that occur word-medially are observably shorter, with the pre-consonantal codas displaying the shortest duration of all.

Statistical modelling confirms the significance of these findings. Given the phonetic differences between the fricative and approximant realisations, we modelled the two data subsets separately. In both cases, duration values in ms were the dependent variable. Phonological context and participant gender were included as fixed predictors along with a random intercept for participant. The contextual differences between the fricative realisations achieve significance after $p$-value adjustment: ##R∼V#RV $\beta = 13.9$, $t = 4.42$, $p < 0.001$; ##R∼VRRV $\beta = 24.5$, $t = 7.81$, $p < 0.001$; V#RV∼VRRV $\beta = 10.7$, $t = 3.41$, $p < 0.01$. With the approximants, word-final realisations are longer than pre-consonantal realisations: R##∼VRC $\beta = 53.6$, $t = 9.64$, $p < 0.001$; VR#V∼VRC $\beta = 52.8$, $t = 6.36$, $p < 0.001$. However, the durational difference between the utterance-final and utterance-medial approximants does not achieve significance: R##∼VR#V $\beta = 52.8$, $t = 0.11$, $p > 0.1$. No significant gender differences were observed in either model.

Regarding voice ratio, we observe an opposite tendency to duration. As shown in Figure 2, the highest levels of voicing are observed in word-medial contexts: that is, for fricative realisations in VRRV, in which the mean voice-ratio value is 0.747, and for approximants in VRC, in which it is 0.965. Again, although there is indication of a high degree of variation, utterance-initial R shows the lowest tendency for voicing (mean voice ratio = 0.2). Utterance-medial word-initial R is somewhat more prone to voicing, reaching a mean voice-ratio value of 0.559. With the approximants, utterance-final R## often displays devoicing over the final phase of the articulation. This accounts for the relative low voice-ratio mean of 0.5. Interestingly, this also occurs utterance-medially, where approximants in the VR#V are the least voiced of all test contexts (average voice ratio is 0.475 in this context).

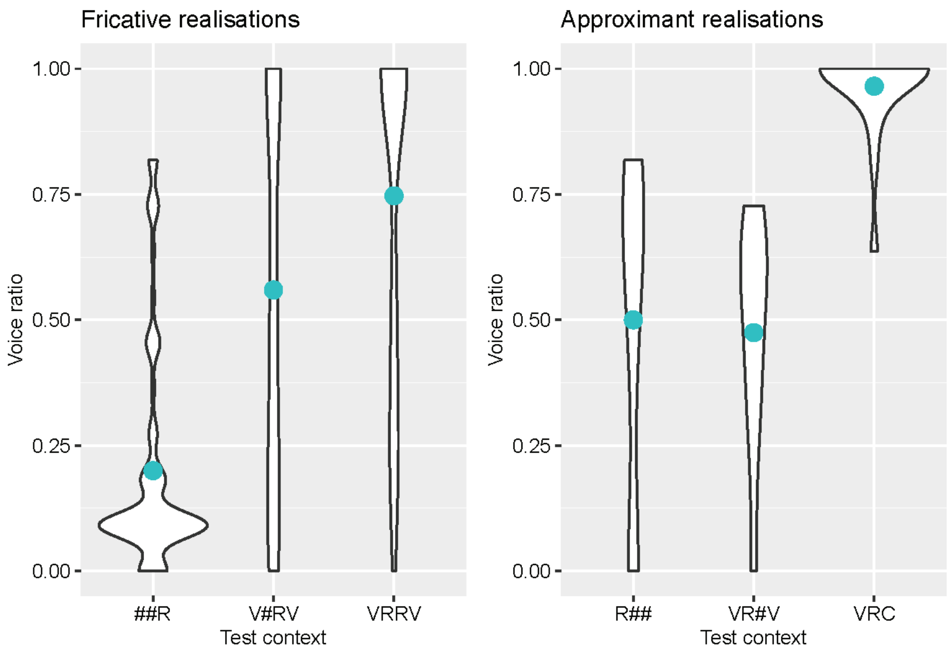



**Figure 2.** Voice ratio measurements (taps excluded) across test contexts. Dots represent mean values.

The voice-ratio data were modelled in the same way as the duration data reported above. In the fricative model, the cline in voice ratio across phonological contexts achieves significance ($p < 0.001$ in all comparisons): ##R∼V#RV $\beta = -0.365$, $t = -8.38$; ##R∼VRRV $\beta = -0.544$, $t = -12.8$; V#RV∼VRRV $\beta = -0.188$, $t = -4.45$. Mirroring the analysis of duration, voice ratio is significantly higher in VRC than in R## ($\beta = -0.422$, $t = -8.18$, $p < 0.001$) and VR#V ($\beta = -0.537$, $t = -6.98$, $p < 0.001$). However, the difference between approximants in R## and VR#V is non-significant: $\beta = 0.115$, $t = 1.583$, $p > 0.1$. Again, no significant gender differences were observed in either model.

In addition to duration and voice-ratio measurements, we also examined the formant-trace data extracted from approximant R-realisations. Based on existing descriptions, we were principally interested in visualising changes in F3 over vowel-to-rhotic transitions. Recall that Rennicke characterises [ɹ] as displaying a gradual narrowing of the formant space between F2 and F3 in pre-vocalic rhotics, whereas [ɹ] is expected to show a rise in F3 in the transitional periods between the vowel and the following rhotic.

In initial modelling, we found that fitting GAMs to the full set of data, covering 13 time points, led to inconsistencies: for example, where partial devoicing in some approximants caused interruption of established formant trajectories. The occurrence of glottal creak, particularly at word-final boundary points, sometimes also contributed to unreliable readings (see Figure 4 in Section 3 for an example of this in an assibilated [ɾˢ]-realisation). Consequently, we chose to work with a smaller temporal window for the analysis. This covers the two measurements points in the pre-rhotic vowel and all measurement values up to and including the acoustic midpoint of each approximant R (i.e., −0.2–0.5 in normalised time). Plots showing fitted F1, F2, and F3 contours over this timescale for the five speakers in the dataset who produced approximated rhotics are shown in Figure 3.

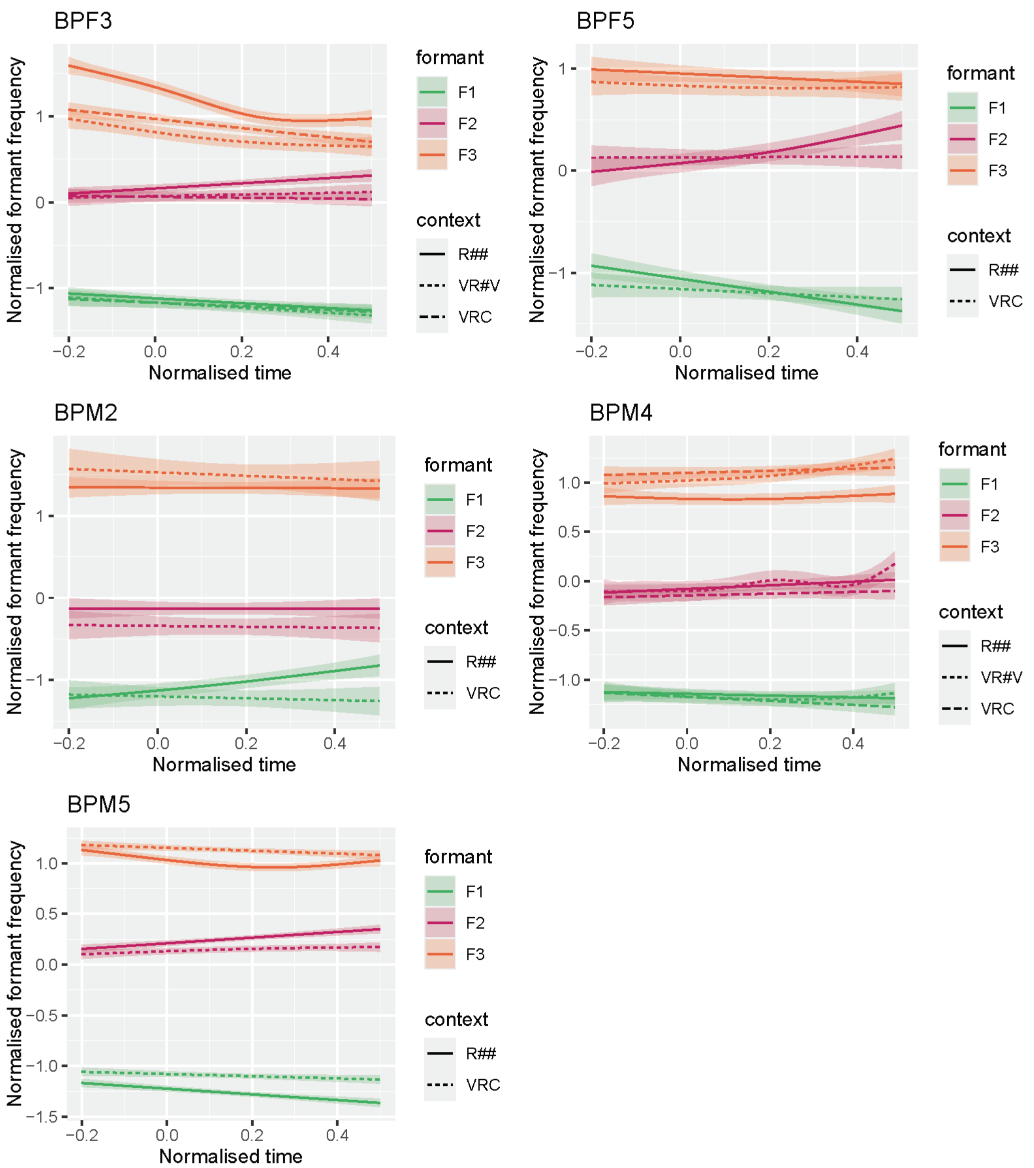



**Figure 3.** Fitted values for normalised F1, F2, and F3 in approximant R-realisations across normalised time (−0.2 to 0.5 timescale).

Approximants produced by speakers BPF3 and BPF5 display a progressive narrowing of the formant space between F2 and F3. As noted, this suggests a retroflected realisation of R, particularly in the utterance-final context. A similar, but visibly less extreme pattern occurs in utterance-final R with BPM5. The flatter F3 contours that we see for BPM2 and BPM4 may be indicative of a less retroflected approximant. The slight F3 rise for BPM4 suggests a more [ɹ]-like quality according to Rennicke's diagnostics; it is also possible that the flatter F3 trajectory, in general, indicates a more central alveolar approximant (with relatively little contextual variation for these speakers). In all cases, the patterns shown in Figure 3 are impressionistically somewhat less dynamic than the examples presented in Rennicke (2015). This may indicate that this set of speakers favour a more central type of approximant, context-dependently, or otherwise that the degree of retroflexion in their rhotic approximants is rather small on the whole. We discuss these patterns further with reference to the UTI data in Section 3.2.2 below.

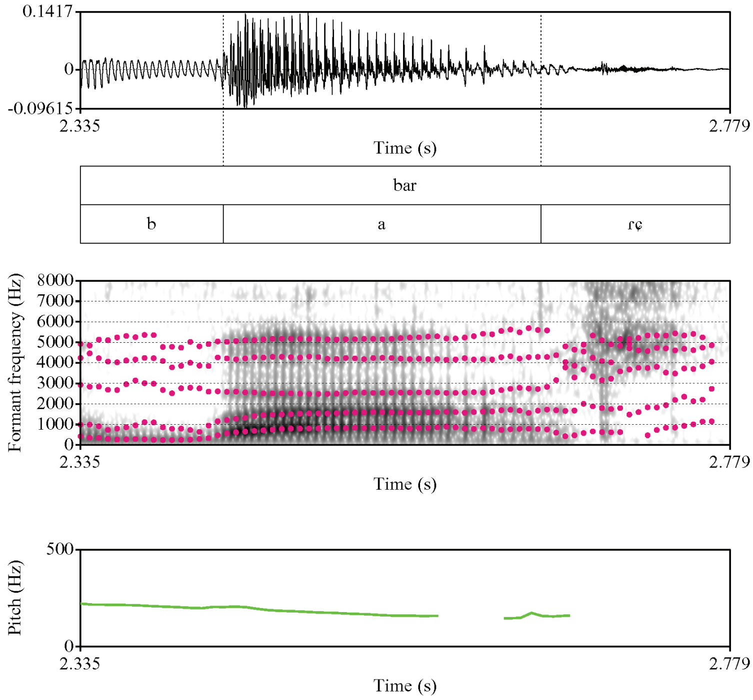



**Figure 4.** Assibilated [ɾʂ] in an utterance-final realisation of *bar*, produced by speaker BPF6.

### 3.1.3. Tapped Realisations

The rhotic realisations that we broadly categorised as tap-like present several complications for analysis. In addition to [ɾ] (the prototypical tap), we encountered assibilated and aspirated taps, i.e., [ɾᶳ] and [ɾʰ], respectively. Taps followed by a period of glottal frication occurred only in the utterance-final context: they were produced by BPF2, BPM3, and BPM6 exclusively. Assibilated taps, which closely resemble realisations reported for certain dialects of Spanish,[9] were observed in R## and VR#V. These were produced exclusively by female speakers (BPF6, BPF7, and BPF8). A typical assibilated realisation is shown in Figure 4. Observe here the occurrence of high-frequency noise following the production of [ɾ]. This coincides with a loss of periodic voicing following several periods of glottal creak that precede the articulation of [ɾ]. This contrasts with the utterance-final aspirated tap shown in Figure 5, where the frication is more diffuse, and with the devoiced trill, also shown in Figure 5, in which multiple short apico-alveolar contacts are observable in the spectrogram.

Additionally, rhotic taps in the VRC context were frequently produced with a short svarabhakti vowel excrescence between the tap and the following consonant.[10] This also occurred—although much more rarely—in utterance-medial VR#V. Examples of this are shown in Figure 6. Note here that the post-rhotic vocalic material is considerably shorter than the duration expected to be observed in a full vowel.

Regarding duration, we observe small differences in simple-tap realisations across test contexts. [ɾ]-realisations are longest utterance-finally, where the mean duration reaches 31.8 ms. [ɾ] displays the shortest duration in the word-medial intervocalic environment (mean of 19.7 ms), and mean values fall between these extremes in VRC (20.7 ms) and VR#V (23.6 ms). These values are within the expected range for taps: for instance, Recasens and Espinosa (2007) refer to 15–30 ms (exclusive of positional variation) as being the recorded durational range for [ɾ] cross-linguistically.

In the utterance-final context, where we observe the greatest differentiation of variants, trilled [r] and assibilated [ɾᶳ] have similar durations: i.e., average values of 147 ms and 148 ms, respectively (recall that the trill data represent realisations produced by a single speaker, BPF1, so the values reported here should not be taken as anything more than indicative). [ɾʰ]-realisations are considerably shorter, reaching a mean of 88 ms here. A further point to highlight is that the duration of [ɾˠ] appears to be somewhat more stable in VR#V-realisations, as compared to R##. Utterance-medially, the mean duration of [ɾˠ] is 98.2 ms, which is shorter that the duration of [r]-realisations produced by BFP1 in this context (i.e., 124 ms). BPF1's trills in the VRC context have a comparable mean duration of 119 ms.

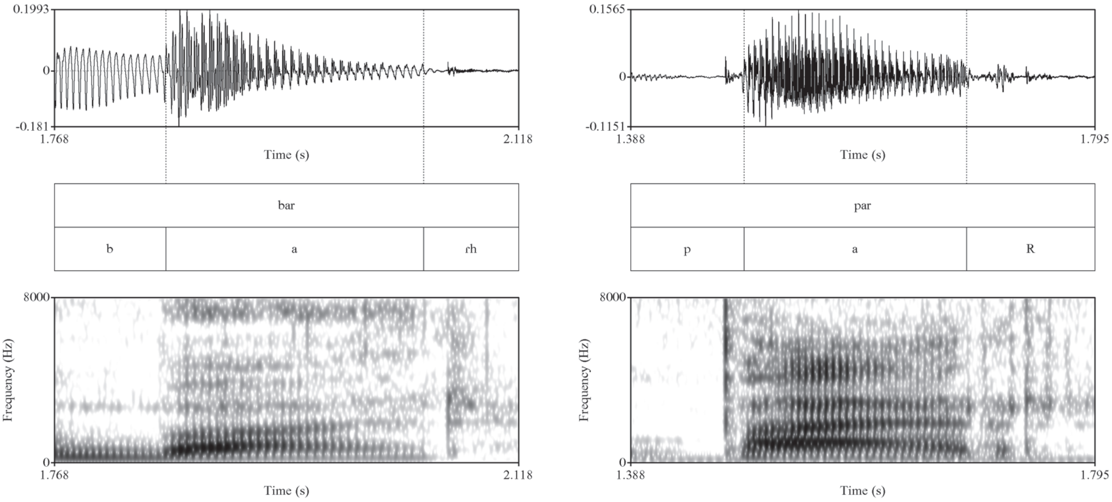



**Figure 5.** Aspirated [ɾh] in an utterance-final realisation of *bar* produced by speaker BPF2 (**left panel**) and an utterance-final trill in *par* produced by BPF1 (**right panel**).

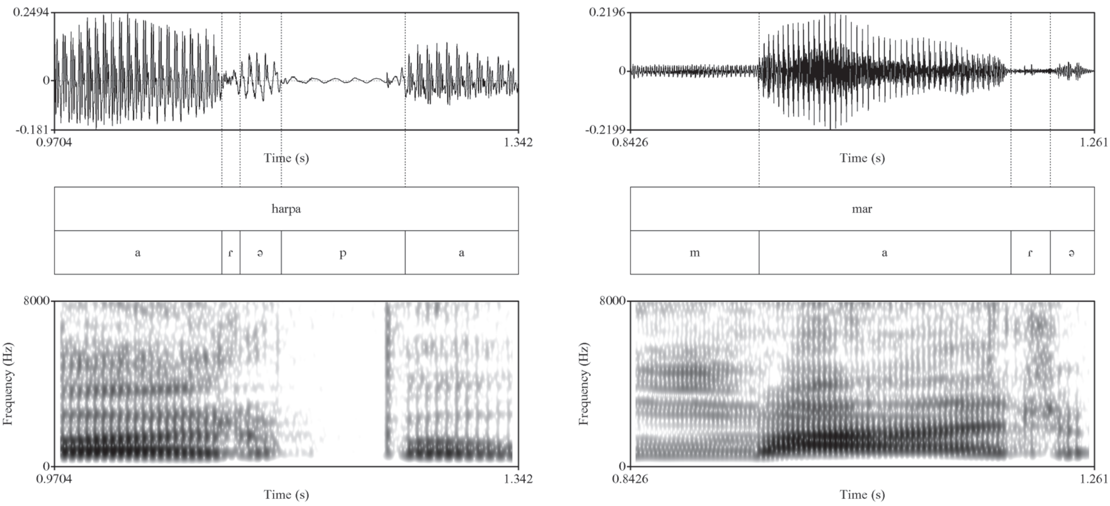



**Figure 6.** Svarabhakti vowel excrescence in VRC (**left panel**: *harpa* produced by BPM6) and VR#V (**right panel**: *mar* produced by BPF7).

Due to the extensive idiosyncratic variation, it was not possible to fit meaningful models to the duration data other than for the simple taps. As with the fricative and approximant-realisations, [ɾ]-durations were modelled with duration in ms as the dependent variable, context and gender as fixed effects, and a random intercept for speaker. Variation by speaker gender was insignificant ($\beta = 2.038$, $t = 1.219$, $p > 0.1$). Some comparisons against R## (in which [ɾ] displays longer duration) did achieve significance: R##∼VRV, $\beta = 11.61$, $t = 3.286$, $p > 0.01$; R##∼VRC, $\beta = 10.55$, $t = 2.944$, $p > 0.05$. The R##∼VR#V comparison is non-significant ($\beta = 7.68$, $t = 2.166$, $p > 0.1$). VRV realisations are shorter than VR#V ($\beta = 3.93$, $t = 3.238$, $p > 0.05$) but non-significantly shorter than word-medial taps in the VRC environment ($\beta = 1.06$, $t = 0.762$, $p > 0.1$).

Voice ratio is also observably affected by realisation and phonological context: we see that voicing is most consistent in word-medial VRV taps and least consistent in the utterance-final context. [ɾ]-realisations register a mean voice-ratio value of 0.25 utterance-finally, which rises progressively across the other test contexts: 0.758 in VR#V, 0.793 in VRC and a near-ceiling value of 0.914 in VRV. The lowest voice-ratio values are observed for [ɾʰ], which displays a mean value of 0.033. Utterance-final [ɾᶳ]-realisations are mainly unvoiced in all contexts: mean voice-ratio values are 0.15 in R##, 0.145 in VR#V, and 0.09 in VRC. The trills produced by BPF1 display more consistent voicing in VRC and VR#V, in which voice-ratio averages are 0.8 and 0.688, respectively. This drops to 0.145 in the utterance-final context.

As with duration, heavy skew in the voice-ratio data prevents full inferential analysis. Nevertheless, a model built on the data subset containing only the simple-tap realisations with the same structure as the duration model confirms the significance of some of the observed patterns. These bear a strong similarity to the results for duration. Comparisons with R##, in which voicing is least consistently observed, all achieve significance at least at the 0.05 level: R##∼VR#V, $\beta = -0.453$, $t = -3.03$, $p > 0.05$; R##∼VRC, $\beta = -0.517$, $t = -3.42$, $p > 0.01$; R##∼VRV, $\beta = -0.609$, $t = -4.032$, $p > 0.001$. Realisations of [ɾ] are only marginally more voiced in VR#V than in VRC ($\beta = -0.148$, $t = -2.888$, $p > 0.05$), whereas no significant difference between VRV and VRC emerges ($\beta = -0.085$, $t = -1.435$, $p < 0.1$).

In general, therefore, the analysis reveals a parallel patterning of duration and voicing: i.e., tap-realisations that are significantly differentiated in duration by phonological context also tend to display significant differentiation in voicing. By contrast, where context exerts only a minimal effect on [ɾ]-durations (e.g., the VRV∼VRC comparison), voicing is also not significantly affected.

### *3.2. Articulatory Analysis*

In this section, we present the results of the articulatory analysis. It was necessary to restrict this part of the analysis to five of the test contexts discussed above, namely, utterance-initial ##R, utterance-final R## and word-medial VRV, VRRV, and VRC. In VR#V-tokens, rhotic realisations were strongly coarticulated with the back vowel that occurs word-initially in carrier phrase (b) in Table 4. Likewise, coarticulation with the word-final back vowel in carrier phrase (c), used to test V#RV, meant that tokens produced in this context were not directly comparable to the word-medial environments in which we used an /a_a/ frame, as mentioned. Despite this, focusing the analysis on the three word-medial environments and the two utterance-peripheral environments allows for our main research questions to be addressed.

As an initial point of reference for the detailed discussion of contextual variation that follows, Figure 7 shows relative tongue positions for the five test contexts as produced by a representative binary-allophony speaker.

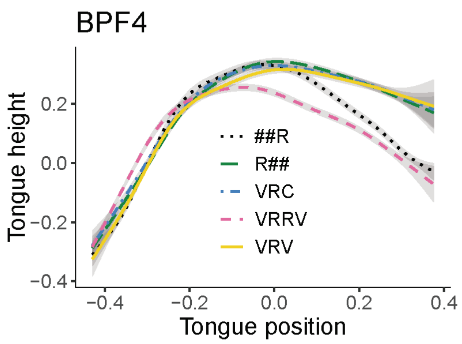



**Figure 7.** Fitted tongue splines for five test contexts: speaker BPF4.

There are two major distinctions to draw attention to here. Firstly, the tap realisations in VRV, VRC, and R## are mainly distinguished from the fricative realisations in VRRV and ##R by the position of the front body of the tongue and the tongue tip. As expected, [ɾ] is produced with a tip-up configuration, whereas the fricative realisations show a tip-down orientation. Secondly, although the tap splines show a high degree of overlap, the fricatives are not spatially identical: the dorsum and front body of the tongue show a significantly greater displacement in utterance-initial ##R than in VRRV. We comment further on these patterns with reference to the other participants in Sections 3.2.1–3.2.3 below.

3.2.1. Fricative Realisations

In the environments in which all speakers consistently produce fricative realisations of R, we observe two main patterns in the articulatory data. Figure 8 shows the rarer pattern: speakers BPF1 and BPM2 produce utterance-initial fricatives that do not differ in terms of tongue position from word-medial <rr>-fricatives. In both cases, the realisation is a period of glottal frication which, in articulatory terms, resembles the lingual configuration for the flanking /a/ vowels. There is a suggestion of an extremely minor amount of fronting in ##R for BPM2, but this is only differentiated from VRRV in a small section of the post-dorsum.

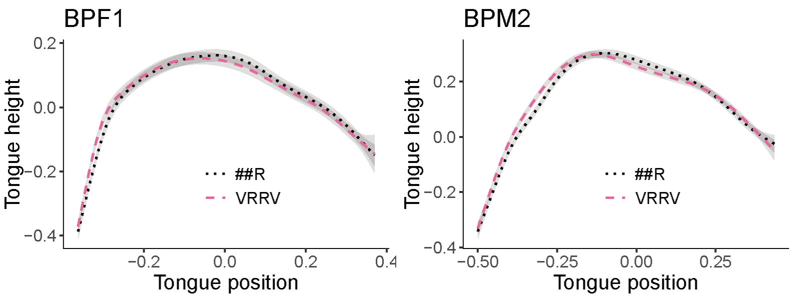



**Figure 8.** Tongue splines for fricative R-realisations in ##R and VRRV produced by BPF1 and BPM2.

Unlike BPF1 and BPM2, the other speakers in the sample all produce fricatives in the ##R environment that display some degree of tongue advancement relative to more retracted realisations word-medially in VRRV: see Figure 9. For a number of speakers, utterance-initial R is produced with the tongue dorsum raised towards the uvula. This is most clearly the case for BPF2, BPF3, BPF4, and BPM4, and to a less extreme extent BPF5 and BPM6. BPF2, BPF4, BPF5, and BPM4 achieve the more uvular utterance-initial target through a small amount of fronting of the tongue root as well as dorsum raising. The tongue tip, however, remains in a similar position to the more glottal realisations observed in the VRRV context (in which the tongue root is more retracted and the dorsum is much lower).

BPF8 and BPM3 also display root fronting in ##R. In comparison to the other speakers, both the dorsum and front body of the tongue have a more raised configuration in ##R. Realisations produced by BPM1, BPM5, and BPM6 in these test contexts trend in the same way; nevertheless, differentiation of VRRV and ##R is observably less extreme for these speakers, such that fricatives in both environments appear to have a more glottal than uvular place-of-articulation target. BPF6 and BPF7 show a lesser differentiation of VRRV and ##R fricatives. BPF6 does display the same tendency for utterance-initial root fronting as the other speakers; however, there is no evidence of dorsum or front-body raising. BPF7 shows almost no root fronting and only a very minor degree of front-body raising in ##R relative to a slightly lower configuration in VRRV.

Despite this, the generalisation applicable to the majority of speakers is that utterance-initial fricatives are produced with a larger displacement of the tongue, particularly at the root and dorsum, than word-medial ones. The degree of differentiation varies across speakers, but in general, the articulatory data match up with the acoustic data, in which longer durations and more devoicing were shown to occur utterance-initially than in word-medial VRRV.

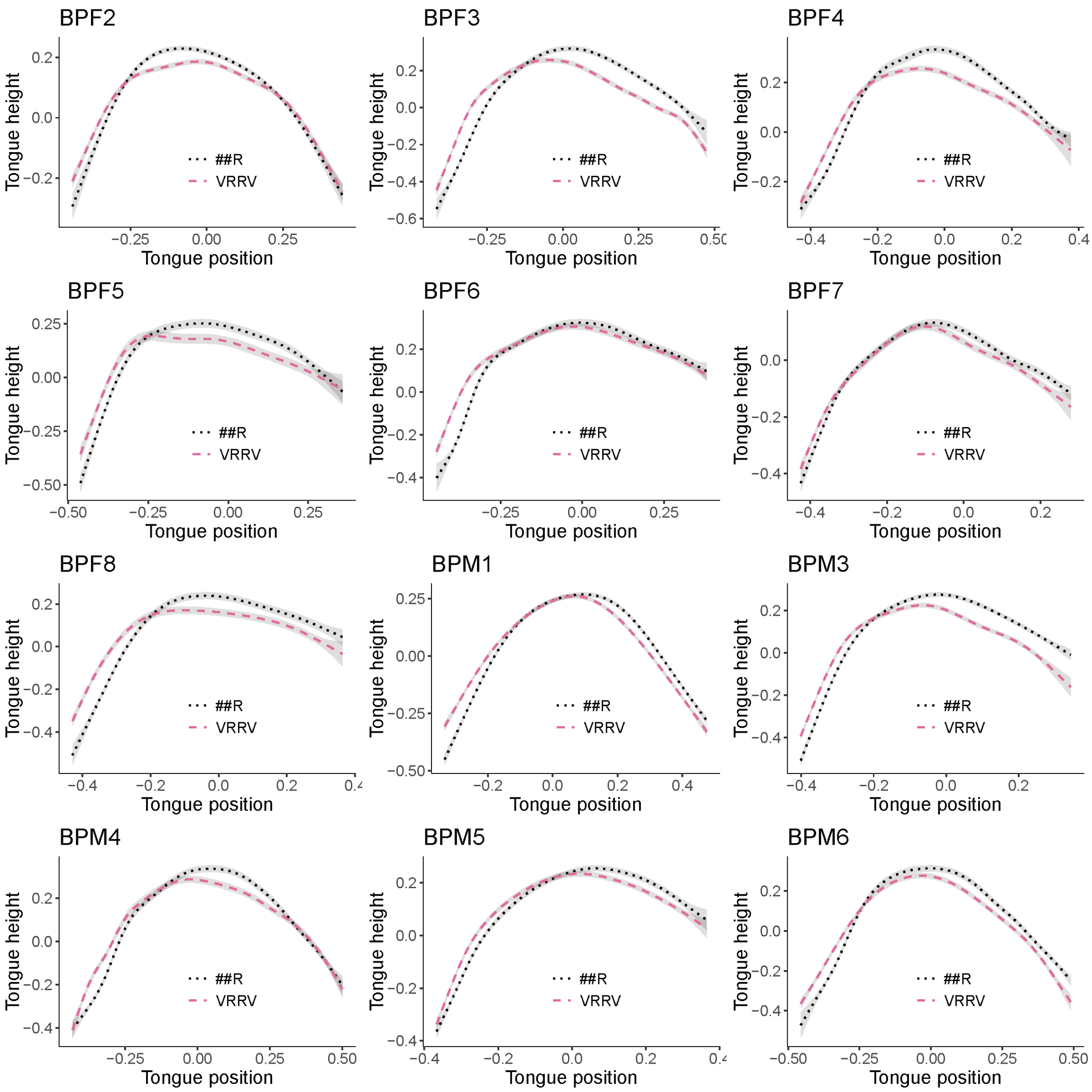



**Figure 9.** Tongue splines for fricative R-realisations in ##R and VRRV (all speakers except BPF1 and BPM2).

### 3.2.2. Approximant Realisations

Figure 10 below shows comparative plots for the speakers who produced rhotic approximants in the R## and VRC test contexts. Significant differences between the splines calculated for these environments are observed except with BPF3. This speaker's realisations differ observably from the other speakers in Figure 10 in that BPF3 produces a tip-down approximant in which bunching is observed in the dorsal region in both R## and VRC. These realisations bear some resemblance to the "dorsum-down" pattern reported by Huang et al. (2024) for Southwestern Mandarin. Here, the tongue tip shows a smaller displacement in pre-consonantal R than in utterance-final R. This is the area of the tongue in which positional variation is greatest (notice the large confidence intervals here). However, given that the tongue displays a near-identical configuration across the root, dorsum, and front-body sections, approximant realisations of R are not robustly differentiated by context for this speaker.

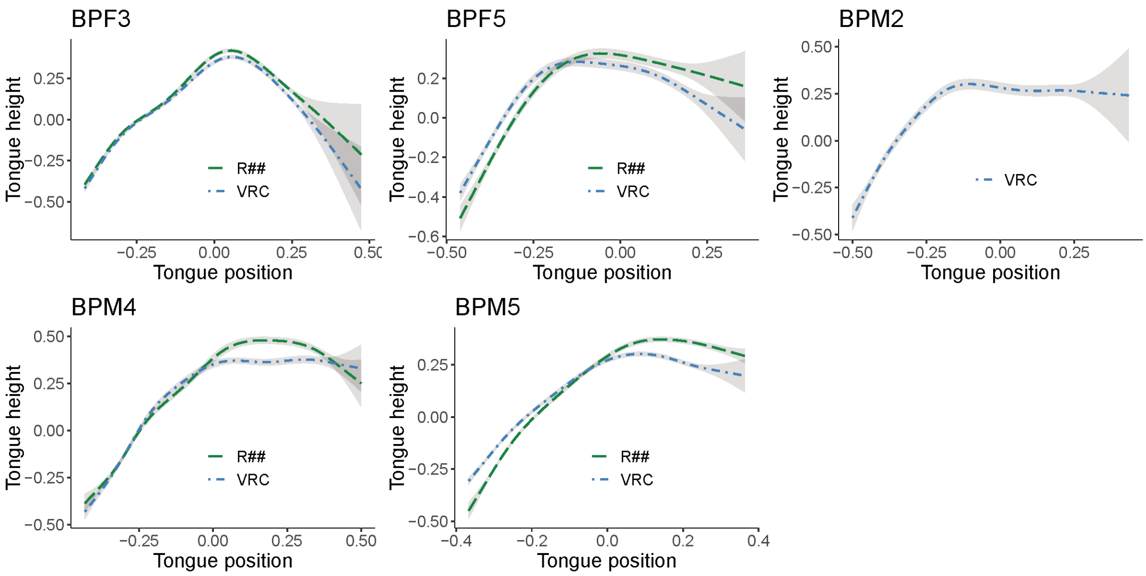



**Figure 10.** Tongue splines for approximant R-realisations in R## and VRC (BPF3, BPF5, BPM2, BPM4, BPM5).

Clear contextual differences are noted for the other speakers in Figure 10, all of whom produce tip-up approximants. Like BPF3, rhotic approximants produced by BPM4 in the two coda environments shown are not distinct in terms of the configuration of the tongue root. However, the tongue body is much higher in utterance-final realisations than in VRC. In the latter context, the tongue has a flat configuration across the dorsum and front body, possibly with some small pre-dorsal hollowing. This is also visible in Figure 11 as a small dip in the splines around the 0 mark on the *x*-axis for BPM2 (and note that the same occurs in BPM4's VRC-spline around 0.2). This kind of bunching can be a characteristic of rhotic retroflexion, as seen in Arrernte (Tabain and Beare 2018) and Malayalam (Scobbie et al. 2013). It contrasts with the more arched position that occurs, for example, in R##-realisations produced by BPM4.

Regarding BPF5 and BPM5, these speakers also differentiate utterance-final coda R from word-medial pre-consonantal R. In the former environment, the root is advanced and the front body of the tongue has a larger displacement than in VRC. A minor dip occurs in the VRC spline for BPM5 around 0.2 on the *x*-axis in Figure 10; however, nothing resembling this is visible for BPF5. Interestingly, these speakers' splines for the two contexts bear similarities to ultrasound data reported in Tabain and Beare (2018) for Arrernte, where a higher tongue position and more advanced root is characteristic of [ɹ] (as compared to [r], which is produced with a more retracted root and lower front body), at least for some speakers. This may therefore suggest that both BPF5 and BPM5 retroflect the tongue to a greater degree in utterance-final coda R than in the VRC environment.[11]

The UTI data from approximant realisations raise questions in light of the acoustic analysis presented above. In particular, the fact BPF3 appears to make a minimal articulatory distinction between R## and VRC is perhaps unexpected since this was the speaker for whom the formant traces extracted for these two environments were most distinct of all the approximant users (cf. Figure 3). Despite that, the higher front-tongue position observable in both BPF3 and BPF5's utterance-final approximants conceivably contributes to the more extensive narrowing of the F2/F3 space: i.e., tongue tip raising causing an observable increase in F2 across the V-to-R transitions in Figure 3.[12] The same pattern was noted for BPM5, who also articulates utterance-final R with the front body in a higher position. It is notable, likewise, that BPM4, for whom the V-to-R formant transitions did not display the characteristic drop in F3 expected of [ɹ], also produces utterance-final approximants with more of a central/palatal arch in the tongue. F3 also generally patterns lower in R## than VR#V and VRC for this speaker. The slight bunching that we observe in the male speakers' VRC splines in Figure 10 does not seem to have a strong acoustic correlate, other than perhaps a somewhat lower F2 frequency in the dynamic V-to-R measurements.

### 3.2.3. Tapped Realisations

We now turn to consider the articulatory data from the rhotic realisations that we broadly classified as taps. As noted above, this is the category that displays the most inter-speaker variation. It must also be highlighted that imaging tap movements with an ultrasound system that achieves ca. 75 frames per second means that measurements are not consistently extractable at or near the gestural maximum of [ɾ]. Whereas the UTI frame rate offers a significant improvement on, for example, EPG data, it must be acknowledged that our observations about the spatial variation in [ɾ] are to some extent limited by the capabilities of the equipment.

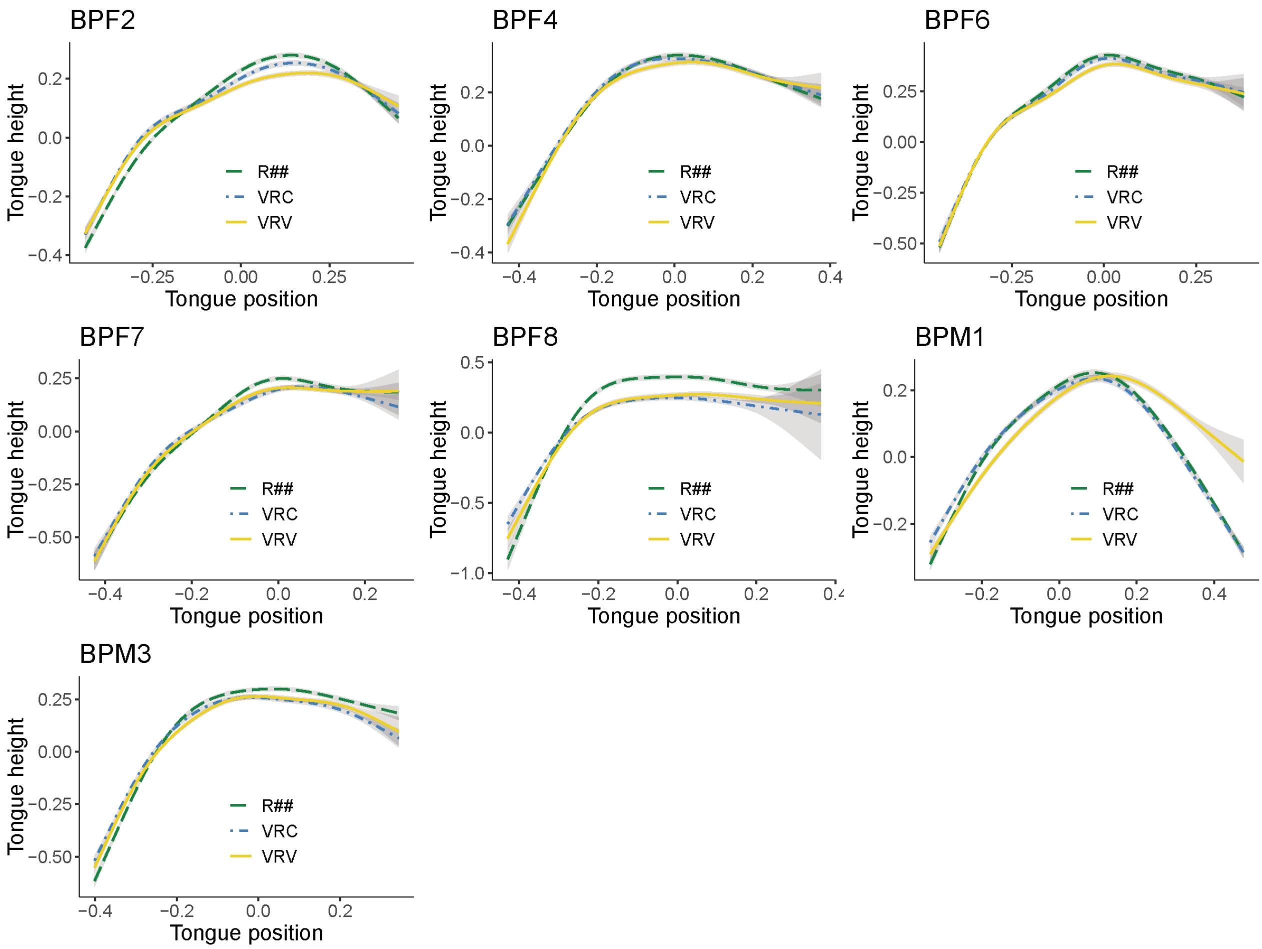



**Figure 11.** Tongue splines for [ɾ]-realisations in R##, VRC, and VRV test contexts (binary speakers).

With this caveat in mind, we begin by considering the *binary* speakers: i.e., those who exclusively produce fricatives word-initially and in VRRV, and tap-like realisations elsewhere. Figure 11 shows comparative plots for tap-like realisations in the R##, VRC, and VRV test contexts. We observe here that tongue splines for taps produced in the VRC and VRV environments are not significantly differentiated for the majority of speakers. Indeed, speakers BPF4, BPF6, BPF7, and BPM3 show little to no differentiation in tongue position across the three test contexts (but note that the tongue reaches a slightly higher displacement in R## for BPF7 and BPM3). Splines for BPF2 show greater differences across the three environments, with the front body of the tongue exhibiting a larger displacement in utterance-final tap realisations. In some cases (as with BPF2, BPF4, BPF6), the tongue has a marginally lower configuration in the word-medial intervocalic context; however, this only appears to be significant for BPF2. BPM1's [ɾ]-spline (i.e., in VRV) resembles taps produced by the other speakers. Recall that this participant produced [h]-realisations of coda R: thus, the tongue position in R## and VRC resembles the /a/-like configuration that we observe in Figure 9.

Speaker BPF8 presents a particularly interesting pattern. As previously noted, this participant is one of the female speakers who regularly produced assibilated rhotic realisations utterance-finally. In R## here, we see significant raising of the tongue body across the dorsal and front-body areas, extending into the tip. This possibly suggests active compression of the airsteam by channel narrowing for the generation of frication across the front area of the tongue. Nevertheless, it is remarkable that there is no evidence for this kind of gesture in the other consistent assibilating speakers (i.e., BPF6 and BPF7 here). In fact, tap productions produced by BPF6 and BPF7 look to be among the most homogeneous across test contexts, as noted. It seems to be the case, therefore, that assibilated taps produced by BPF6 and BPF7 are precisely that: [ɾ]-like articulations in which the tongue reaches its maximal displacement and remains in position for a short period. During that period, airflow is maintained but voicing tends to drop out, thus resulting in the generation of relatively weak apico-alveolar voiceless frication. BPF8 is clearly different in this regard: the tongue contour shown in Figure 11 looks to reflect a more obviously controlled sibilant production, possibly resembling something closer to [ɾ$^{ɕ}$] than [ɾ$^{ʂ}$].

Finally, we come to the word-medial rhotic realisations produced by the ternary speakers. Figure 12 shows splines for VRC and VRV in which these speakers tended to favour the use of taps (despite consistent use of approximants utterance-finally). BPF1 is included here for comparison: recall that this speaker articulated alveolar trills in VRC. In this case, the fitted splines show that the tongue orientation for [r] is not markedly different from that for [ɾ], at least when measured at the durational midpoint of both types of realisation. The dorsum reaches a slightly more retracted position in [r], as expected, but this difference is marginal. The other speakers in Figure 12 all produce approximants in VRC versus simple taps in VRV. For BPM2 and BPM5, the tongue configuration in the VRC-approximants is not robustly different from intervocalic [ɾ]. The root is slightly more advanced for the tap than for the approximant in both speakers, and the dorsum has a larger displacement in BPM5's approximant realisations. This suggests an articulation that is more [ɹ]-like than [ɻ]-like: as noted above, BPM5 produces approximants that look to be more retroflected in the R## environment.

Otherwise, BPF2 and BPM4 show approximant realisations in VRC in which the tongue has a higher configuration at the front body than in intervocalic [ɾ]. The characteristic bunching that is observable in BPM4's pre-consonantal approximant is also absent from the tap spline. There is also a small degree of bunching visible in BPF2's VRC spline, although in this case its locus aligns around –0.15 on the *x*-axis, i.e., in the dorsal region. Interestingly, BPF5's approximant spline shows a significantly lowered tongue tip relative to its position in [ɾ]. As with some of the utterance-final realisations discussed above, this may indicate a more palatal (or palatalised realisation) than [ɹ] or [ɻ].

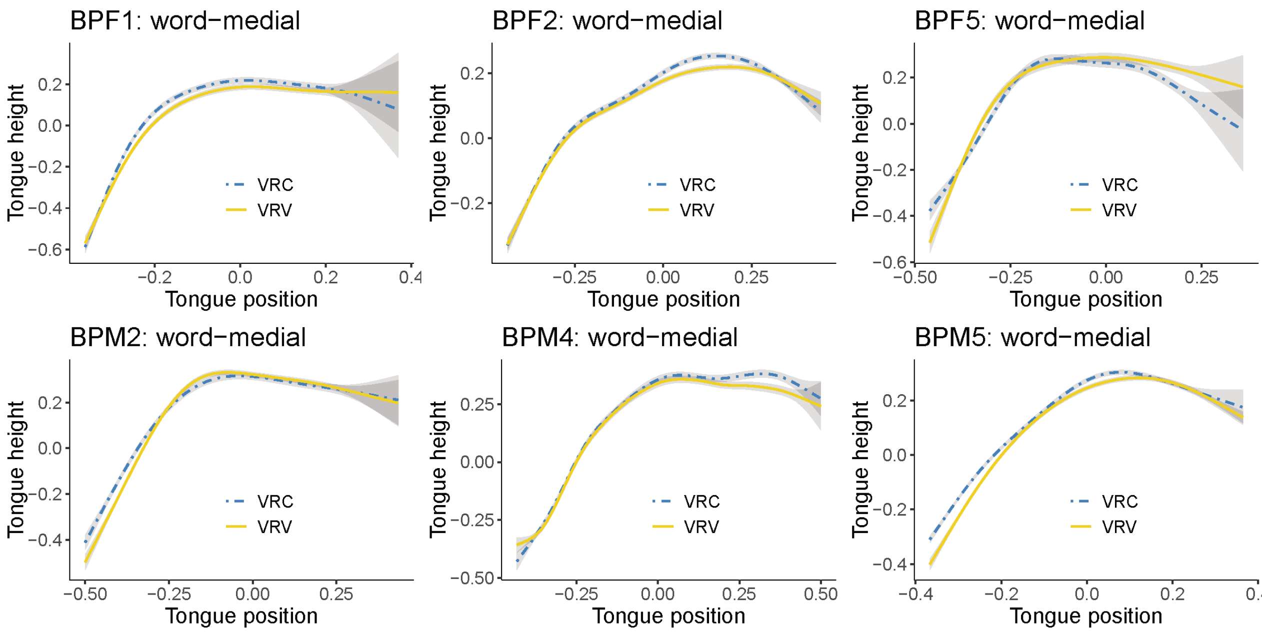



**Figure 12.** Tongue splines for word-medial rhotic realisations (ternary speakers).

### 3.3. Summary of Results

Our analysis confirms the existence of both binary and ternary allophony patterns among BP speakers in our participant sample. Binary speakers produce rhotic fricatives in ##R, V#RV, and VRRV contexts and taps elsewhere. Ternary speakers have the same patterning of rhotic fricatives, but taps are restricted to VRV; they instead articulate rhotic approximants in R## and VRC contexts. One speaker (BPF3) uses a bunched tip-down configuration in these environments, whereas the other speakers all produce tip-up approximants. We have also highlighted that tap-realisations are not uniform across test environments, specifically with regard to the occurrence of tap frication (of both glottal and assibilated qualities) and excrescent vowels context-dependently.

In addition to the categorical allophonic alternations, we have shown that there is variation in the acoustic and articulatory profiles of rhotic variants that is dependent upon prosodic context. Rhotic fricatives are longer, less voiced, and articulated with a larger displacement of the tongue in ##R than VRRV: i.e., [ʁ] shows evidence of strengthening when it occurs at the utterance edge. Similarly, with the exception of the bunched [ɹ]-realisations produced by BPF3, approximants are articulated with longer durations and larger tongue displacements in utterance-final environments than in word-medial VRC. Whereas similar spatial variation does not occur with the taps, [ɾ] is longer and less voiced in utterance-final position (for those speakers who use this variant in R##) than utterance-medially.

## 4. Discussion

The results presented above confirm a number of points about rhotic variation in BP. In agreement with previous work that has surveyed the phonological patterning of rhotic variants, our data reveal complex inter- and intra-speaker variation. And as traditional phonological accounts have outlined, we have encountered categorical allophonic patterns in the data. However, further than a broad distinction between, e.g., *strong-*, *weak-*, and *coda-R*, a finer categorisation of phonetic variability emerges.

One striking finding concerns the patterns of variability that seem to align with specific phonological boundaries. We have shown that rhotic fricatives in utterance-initial position are longer, less voiced, and often produced with larger displacement of the tongue dorsum than fricatives in the VRRV context. These phonetic characteristics are precisely what we would expect to observe under a scenario of edge strengthening: a controlled enhancement in the context of a high-level prosodic boundary.

This phenomenon has been documented for a number of languages on the basis of articulatory experimentation. As noted in Section 1.3, word-initial rhotics show a tendency for gestural enhancement in Catalan (Recasens and Espinosa 2007) and Greek (Baltazani and Nicolaidis 2013) relative to more reduced articulations in word-medial environments. Electromagnetic midsagittal articulometry studies on French (Tabain 2003) and English (Byrd et al. 2005) have also shown that articulatory gestures are bigger and are produced with faster velocities (and longer overall durations) when consonants occur adjacent to higher prosodic boundaries, like the utterance or IP, than when they occur domain-medially or adjacent to lower prosodic boundaries (e.g., phonological word or syllable). Similarly, research using EPG (Fougeron and Keating 1997; Keating et al. 2003, 1999) and UTI (Ramsammy and King 2023) confirms the existence of the same kind of prosodically-driven dynamic variation for other languages.

What we have observed for fricatives in ##R (and acoustically for V#RV) versus VRRV partially replicates one of the results from Howson and Kochetov (2018), who claim that BP fricatives are gesturally specified for a uvular/pharyngeal place-of-articulation target. Arguably, it is the occurrence of /ʁ/ at the left edge of a P-word boundary (utterance-medially in their data) that maximises the potential for realisation of this target. Word-medially in VRRV, by contrast, this target has a greater tendency to be undershot, resulting in what Howson and Kochetov (2018) refer to as a placeless contextual variant of /ʁ/. In addition to illustrating that word-medial fricatives are produced with smaller tongue displacements than in ##R, we interpret our acoustic findings as further evidence that VRRV admits gradient weakening—or lenition, in Howson and Kochetov's terms—due to its occurrence in a weak prosodic position. Conversely, the utterance boundary in ##R promotes a stronger, enhanced (or fortified) fricative realisation that affects both its articulatory and acoustic qualities.[13]

Like the fricative realisations, we have observed that taps are shorter and more voiced in word-medial contexts (VRV and VRC) than in utterance-final position. To some extent, loss of voicing and increased duration at the periphery of the phrase is suggestive of the same sort of strengthening as in utterance-initial fricatives. However, there is a complication in that it is also utterance-finally that we most commonly observe aspiration and assibilation of [ɾ]. It is reasonable to assume that these realisations reflect different articulatory timing strategies. In the former case, the tongue tip produces an [ɾ]-gesture and then rapidly lowers as vocal-fold vibration ceases, thereby generating a period of glottal frication. In [ɾᶳ], voicing also falls out, but the tongue tip must remain in a raised position as airflow is sustained. This points to both types of articulation being planned and controlled: the question that arises, then, is whether these can feasibly be conceived as edge-strengthening phenomena. On the one hand, it is possible that both are active edge-marking strategies that signal the end of an utterance, in which case tap frication could be considered an enhancement. On the other, the failure of voicing could be described in terms of partial undershoot or failed execution of a laryngeal target.

Indeed, Rennicke (2015) interprets the occurrence of devoiced and aspirated taps as weakening in line with the lenition trajectory that she proposes: [ɾ] > [ɾ̞] > [ɹ] > [ɻ] > [ɚ]. However, an alternative way of looking at this is that speakers actively manage the relative timing of laryngeal and lingual gestures in order to fortify rhotics at both utterance peripheries. Thus, what happens in a categorical way utterance-initially, namely, devoicing and frication, may have a more continuous and variable parallel utterance-finally. If this is the case, then any fortition–lenition trajectory that is proposed for a language like BP must have a sensitivity to prosodic environment built into it. Whereas a scale like [ɾ] > [ɾ̞] > [ɹ] > [ɻ] > [ɚ] may capture a general trend, it does not account for what our results indicate, namely, that strengthening and reduction are at least partially predicable on the basis of prosodic factors.

Regarding the approximants, it is worth considering to what extent edge strengthening may be responsible for the variability we have observed between VRC and R##. The tongue displacements shown in Figure 10 for R## certainly appear larger than the VRC counterparts; and furthermore, it is clear that at least some of the speakers produce approximants with different degrees of retroflexion in these contexts. In general, UTI is somewhat limited in recording very fine-grained differences in retraction and arching of the tongue tip. Other imaging techniques such as fMRI may be necessary to gain a fuller understanding of the contextual variation in retroflex rhotic articulations in BP (see recent work by Kochetov et al. 2023 on retroflex consonants in Kannada). Nevertheless, the current articulatory data coupled with the acoustic findings suggest that speakers may produce [ɹ]-like and [ɻ]-like approximants along a continuum, with the degree of retroflexion—along with variation in gestural duration and voicing—correlating with prosodic boundaries, at least in part.[14]

Recognising the reality of edge strengthening is also important in respect of the theoretic claims about the phonological status of rhotics in Portuguese. For example, recall that two of the speakers do not differentiate fricatives in spatial terms in the utterance-initial versus word-medial intervocalic environments, and that some other speakers show only relatively small differences here. These facts, perhaps, offer some support for the claim that rhotic fricatives in the two environments are representationally identical at the surface phonological level (as in Mateus and d'Andrade 2000, for example). Theoretical models of this type would nevertheless have to rely on the assumption of some sort of controlled, prosodically driven variability in phonetic production to account for the variation in the realisation of surface [ʁ] (as a derived variant of underlying /ɾ/) for the majority of speakers who show this pattern.

In fact, the same issue arises if one takes the opposite theoretical position and assumes that synchronic BP has a contrastive rhotic fricative, as in Cristófaro Silva (2003). Such a model still assumes a representational equivalence between fricatives in ##R and VRRV, such that controlled edge strengthening in the former context and domain-medial reduction in the latter is the most coherent way of explaining the variability we have described. However, the articulatory and acoustic differences in fricatives in these contexts could also be predicted by the third type of theoretical model, i.e., one in which only intervocalic fricatives are phonemic (cf. Câmara 1972). In this case, it could be concluded that the tendency for a more reduced fricative—or perhaps even a placeless one, as Howson and Kochetov (2018) propose—reflects the phonetic realisation of a separate phonological category. The realisations that are spatially and temporally different in word-initial contexts would thus reflect the realisation of a computationally derived fricative, e.g., by a /##ɾ/→[ʁ] rule, rather than a lexical one. In other words, observable non-isomorphisms between utterance-initial and word-medial fricatives might be taken as evidence that these are representationally distinct in the phonology in some way.

Regarding the putative derivation of intervocalic fricatives, there may be good diachronic reasons for being suspicious of analyses that rely on /ɾɾ/→[ʁ] mappings. Portuguese regularly reduced all other inherited geminate sonorants to singletons: as in Lat. *annus* > *ano* 'year', *olla* 'pot' vs *olaria* 'pottery'. It is possible that /ɾɾ/ may have been exceptional in this regard; but aside from orthographic conventions, it is equally possible that inherited rhotic geminates followed the same pattern of degemination as other sonorants. Crucially, as none of the theoretical approaches described above make very specific predictions about what fully derived versus partially derived allophony versus full lexical contrast might look like phonetically, these issues will be resolved here. That is to say that no phonetic data are capable of speaking decisively on the issue of which of the rhotic variants we have described may be contrastive and which contextually generated.

More concretely, there are sociophonetic implications to our findings. Given the reported stigmatised status of approximants in general in BP, we wonder whether any of the speakers we have categorised as using binary rhotic allophony actively suppressed use of approximants. A sentence-reading task involving UTI may well have caused some speakers to default to a more formal register; and it should also be reiterated that all participants in our sample were educated to university level. If some participants chose to favour the prestige variant—i.e., coda taps—then this patterns in the expected direction with regard to gender: only two out of eight female speakers produced approximants under experimental conditions, whereas three out of the six male participants did so. For those speakers who did produce them, social stigma could have been a relevant factor in the *degree* to which participants retroflected approximants. In this sense, there may be finely tuned socio-indexical associations that span the range of more "acceptable" [ɹ]-like productions to more highly stigmatised [ɻ]. This is something that merits further investigation, along the lines of what Lawson et al. (2014) have achieved for Scottish English.

Moreover, we believe that further research on the sociophonetic status of rhotic assibilation in BP is necessary. In our data, [ɾ$^{\text{ʂ}}$] was only produced by female participants. It is clear in the

Latin American dialects of Spanish in which assibilated rhotics have emerged, that the change is being led by women (Mazzaro and de Anda 2020). Whereas assibilation is a stigmatised feature in Spanish, we suspect that the situation is rather different in BP. In fact, we wonder whether [ɾ̝̊] represents an incipient prestige variant, and one that is perhaps a younger innovation in Portuguese than in Spanish. If this is the case, then a specific gender patterning, namely, greater frequency of occurrence in the speech of high-status women, would not be unexpected.

Before concluding, it is worth remarking on some of the limitations of this study. As already alluded to, the use of controlled speech stimuli recorded under laboratory conditions with UTI instrumentation may have prompted speakers to monitor their speech quite closely. It is therefore possible that our data are under-representative in certain aspects, particularly as regards the retroflex rhotic realisations that could be said to characterise paulista BP: i.e., in being well above the level of sociolingusitic awareness. Relatedly, we were not able to control for age or socio-economic status in the participant sample. The findings from previous sociolinguistic studies, such as Oushiro (2021) and Oushiro and Mendes (2013), lead us to believe that some of the variation we have described occurs due to complex interactions of phonological, phonetic, and sociolinguistic factors. Further research will certainly be necessary to disentangle these aspects of variation from one another.

## 5. Conclusions

This study has documented patterns of contextual variation in rhotics in the paulista variety of Brazilian Portuguese. In response to previous descriptive work, we applied acoustic and articulatory analysis to the complex rhotic allophony pattern that exists in the language. Beyond basic categorical alternations between taps, fricatives, and approximants that existing accounts describe, we have shown that there are much more fine-grained dimensions to the contextual variability of rhotics. Variation in duration and voicing as well as differentiation of tongue shape and position often aligns with prosodic structure. Thus, we have attributed some of the variation we have observed to gradient phonetic enhancement in the context of major prosodic boundaries (i.e., edge strengthening) and concomitant reduction or undershoot of articulatory targets in weaker prosodic environments. Regarding phonological analysis, we have argued that no single pre-existing theoretical approach to the alternations under discussion necessarily captures the data we have presented better than any other. It may well be the case that new theoretical thinking will need to be applied in the future to the issues that rhotic variation in BP presents. We have further highlighted sociophonetic patterns of interest in the data and outlined specific areas that would benefit from additional empirical research.

**Author Contributions:** Conceptualisation, M.R. and B.R.d.M.; methodology, M.R.; data collection, M.R. and B.R.d.M.; data curation, M.R.; data analysis, M.R.; writing—original draft preparation, M.R. and B.R.d.M.; writing—review and editing, M.R. and B.R.d.M.; project administration, M.R.; funding acquisition, M.R. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by a grant awarded by the School of Philosophy, Psychology and Language Sciences at the University of Edinburgh to the first author.

**Institutional Review Board Statement:** This study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee in the School of Philosophy, Psychology and Language Sciences at the University of Edinburgh (Reference 417-1819/1).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author (restrictions due to ethical considerations apply).

**Acknowledgments:** The authors gratefully acknowledge data-processing and annotation work undertaken by Gilly Marchini that has contributed to the research reported in this paper.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Notes

1 Based on linguistic atlas data, Duarte Sanches and Camarão (2020) also report high incidence of coda [h, ɦ] in Portuguese varieties spoken in Amapá in the north of Brazil.

2 See also Chabot (2019) who refers to a similar chain of rules in analysing Carioca Portuguese: i.e., /ɾɾ/ → /x.ɾ/ (coda /ɾ/-velarisation) → /x.ɦ/ (onset /ɾ/-debuccalisation) → [x] (assimilation/absorption).

3 Also departing from a monophonemic analysis, Bonet and Mascaró (1997) argue for a pan-Iberian contrast between two representationally distinct tap phonemes. It is not clear that this type of analysis can cope with dialects of BP exhibiting a three-way surface distinction in rhotics, such that the specific details of this theoretical approach will not be discussed further.

4 The headset is manufactured by Articulate Instruments Ltd. The same model has been used in a number of UTI studies: e.g., King and Ferragne (2020); Lawson et al. (2011, 2013, 2014); Ramsammy and King (2023); Strycharczuk and Sebregts (2018), among others.

5 Note that orthographic <h> is not phonetically realised in Portuguese. Additionally, where stress occurs on /a/, the realisation is [a]. Unstressed /a/ reduces to [ɐ] through a phonologically regular alternation: hence, a typical pronunciation of *harpa* 'harp' is [ˈaɾ.pɐ].

6 It should be noted that we did not have access to bite plates for the experiment reported here. Bite plates are used in some studies to provide an occlusal-plane measurement (e.g., King and Ferragne 2020; Lawson et al. 2011, 2013, 2014), which can be used to rotate splines for the purpose of increasing comparability between speakers. In the absence of these devices, the spline plots presented in Section 3 are based on unrotated data.

7 This basic categorisation overlooks the single utterance-final trill produced by BPF4, as noted.

8 As noted above, [Vɹ]-realisations present a particular challenge for segmentation: it is possible that could be responsible for some of the high variability we observe particularly in [ɹ]-durations utterance-finally. Nevertheless, the general picture that Figure 1 shows—i.e., increased duration in R## and VR#V relative to reduced duration in VRC—is reliable and consistent with our impressionistic observation of the data.

9 For example, varieties spoken in Costa Rica (Carranza 2006), Dominican Republic (Willis 2007), Ecuador (Bradley 1999, 2004) and Mexico (Bradley and Willis 2012; Mazzaro and de Anda 2020), among others.

10 For clarity, *svarabhakti* here is intended in the sense of a brief vowel-like phonetic event occurring after /ɾ/, i.e., a vowel *fragment*, as described by Sanskrit grammarians (Whitney 1896, §230). We do not consider this vocalic material to be a phonologically "visible" epenthetic vowel.

11 Utterance-final R-realisations produced by BPM2 were particularly variable: it was therefore not possible to produce a fitted spline for R## that reliably represents a typical realisation. Further to this, a reviewer queries whether a white shadow above the tongue tip could be used as a diagnostic of retroflexion in these realisations. Visual inspection of ultrasound splines provides no evidence of white shadows in any approximant realisation produced by BPF5, BPM2, BPM4 or BPM5. These shadows occur regularly in realisations of trilled [r] for those speakers who produce them. This may indicate that the BP rhotic approximants are produced with a smaller degree of retroflexion and/or a lesser narrowing of the articulatory space between the tongue tip and the palate than has been reported for some English varieties (cf. King and Ferragne 2020; Mielke et al. 2016). They instead show a greater resemblance to the "front up" realisations reported in Lawson et al. (2011) for some speakers of Scottish English.

12 Recall that the 0.5 normalised time point in Figure 3 corresponds to the measurement point at which the ultrasound tongue traces were extracted.

13 In accordance with a traditional understanding of the term in historic phonology, this assumes that both articulatory magnification and rhotic devoicing are indicative of *fortition*: i.e., a change affecting consonants "towards less vowel-like qualities" that involves "devoicing, occlusion, loss of sonorancy" (Cser 2014, p. 201).

14 As discussed in Section 1.3, similar patterns have been shown to occur in other languages, such as English, Dutch and Mandarin. A reviewer queries whether this might be said to be indicative of a cross-linguistic (possibly universal?) tendency for continuous [ɹ]-[ɻ] variation in languages that have [ɻ] as a speech sound. Whereas an implicational generalisation of the type "if [ɻ], then also [ɹ]" seems to be a reasonable characterisation of rhotic variability in some languages studied so far, our view is that further articulatory research on rhotics based on a richer sample of languages would be necessary to address this hypothesis in a meaningful way.

## References

Baltazani, Mary, and Katerina Nicolaidis. 2013. The many faces of /r/. In *Rhotics: New Data and Perspectives*. Edited by Lorenzo Spreafico and Alessandro Vietti. Bolzano: Bozen Bolzano University Press, pp. 125–44.

Barlaz, Marissa, Ryan Shosted, Maojing Fu, and Brad Sutton. 2018. Oropharyngeal articulation of phonemic and phonetic nasalization in Brazilian Portuguese. *Journal of Phonetics* 71: 81–97.

Boersma, Paul, and David Weenink. 1992–2023. Praat: Doing Phonetics by Computer. Computer Software, website: praat.org. Available online: https://www.fon.hum.uva.nl/praat/ (accessed on 10 November 2024).

Bonet, Eulàlia, and Joan Mascaró. 1997. On the representation of contrasting rhotics. In *Issues in the Phonology and Morphology of the Major Iberian Languages*. Edited by Fernando Martínez Gil and Alonso Morales-Front. Georgetown: Georgetown University Press, pp. 103–26.

Boyce, Suzanne E., Sarah M. Hamilton, and Ahmed Rivera-Campos. 2016. Acquiring rhoticity across languages: An ultrasound study of differentiating tongue movements. *Clinical Linguistics and Phonetics* 30: 174–201.

Bradley, Travis. 1999. Assibilation in Ecuadorian Spanish: A phonology-phonetics account. In *Formal Perspectives on Romance Linguistics: Selected papers from the 28th Linguistic Symposium on Romance Languages (LSRL XXVIII), University Park, 16–19 April 1998*. Edited by Jean-M. Authier, Barbara E. Bullock and Lisa A. Reed. Amsterdam: John Benjamins, pp. 57–71.
Bradley, Travis. 2004. Gestural timing and rhotic variation in Spanish. In *Laboratory Approaches to Spanish Phonology 7*. Edited by Timothy Face. Berlin: Mouton de Gruyter, pp. 197–224.
Bradley, Travis, and Erik W. Willis. 2012. Rhotic variation and contrast in Veracruz Mexican Spanish. *Estudios de fonéTica Experimental* 21: 43–74.
Byrd, Dani, Sungbok Lee, Daylen Riggs, and Jason Adams. 2005. Interacting effects of syllable and phrase position on consonant articulation. *Journal of the Acoustical Society of America* 118: 3860–73.
Campbell, Fiona, Bryan Gick, Ian Wilson, and Eric Vatikiotis-Bateson. 2010. Spatial and temporal properties of gestures in North American English /ɹ/. *Language and Speech* 51: 49–69.
Carranza, Luz Marina Vásquez. 2006. On the phonetic realization and distribution of Costa Rican rhotics. *Revista de Filología y Lingüística de la Universidad de Costa Rica* 32: 291–309.
Câmara, Joaquim Mattoso. 1972. *The Portuguese Language*. Chicago: University of Chicago Press.
Chabot, Alex. 2019. What's wrong with being a rhotic? *Glossa: A Journal of General Linguistics* 4: 38.
Charles, Sherman, and Steven M. Lulich. 2018. Case study of Brazilian Portuguese laterals using a novel articulatory-acoustic methodology with 3D/4D ultrasound. *Speech Communication* 103: 37–48.
Chen, Shuwen, D. H. Whalen, and Peggy Pik Ki Mok. 2024. Production of the English /ɹ/ by Mandarin-English bilingual speakers. *Language and Speech* 1–38, *online first*.
Cristófaro Silva, Thaïs. 2003. *Fonética e Fonologia do Português*. São Paulo: Editora Contexto.
Cser, András. 2014. Basic types of phonological change. In *The Oxford Handbook of Historical Phonology*. Edited by Patrick Honeybone and Joseph Salmons. Oxford: Oxford University Press, pp. 193–204.
Davidson, Lisa. 2006. Comparing tongue shapes from ultrasound imaging using smoothing splines analysis of variance. *Journal of the Acoustical Society of America* 120: 407–15.
Dediu, Dan, and Scott R. Moisik. 2019. Pushes and pulls from below: Anatomical variation, articulation and sound change. *Glossa* 4: 7.
Delattre, Pierre, and Donald C. Freeman. 1968. A dialect study of American r's by x-ray motion picture. *Linguistics* 6: 29–68.
Duarte Sanches, Romário, and Loerhana Geisielle Quintela Miranda Camarão. 2020. Perfil do fonema /R/ em coda silábica no falar amapaense. *Estudos Linguísticos e Literários* 67: 367–89.
Fougeron, Cécile, and Patricia Keating. 1997. Articulatory strengthening at the edges of prosodic domains. *Journal of the Acoustical Society of America* 101: 3728–40.
Heyne, Matthias, Xuan Wang, Donald Derrick, Kieran Dorreen, and Kevin Watson. 2020. The articulation of /ɹ/ in New Zealand English. *Journal of the International Phonetic Association* 50: 366–88.
Howson, Phil J., and Alexei Kochetov. 2018. Gestural lenition of rhotics captures variation in Brazilian Portuguese. Paper presented at Interspeech 2018, Hyderabad, India, September 2–6, pp. 182–86.
Howson, Phil J., Scott Moisik, and Marzena Żygis. 2022. Lateral vocalization in Brazilian Portuguese. *Journal of the Acoustical Society of America* 152: 281–94.
Huang, Jing, Feng-fan Hsieh, Yueh-chin Chang, and Mark Tiede. 2024. On the two rhotic schwas in Southwestern Mandarin: When homophony meets morphology in action. *Phonetica* 81: 43–80.
Keating, Patricia A., Taehong Cho, Cécile Fougeron, and Chai-Shune Hsu. 2003. Domain-initial articulatory strengthening in four languages. In *Papers in Laboratory Phonology 6*. Edited by John Local, Richard Ogden and Rosalind Temple. Amsterdam: John Benjamins, pp. 143–61.
Keating, Patricia A., Richard Wright, and Jie Zhang. 1999. Word-level asymmetries in consonant articulation. *UCLA Working Papers in Phonetics* 97: 157–73.
King, Hannah, and Emmanuel Ferragne. 2020. Loose lips and tongue tips: The central role of the /r/-typical labial gesture in Anglo-English. *Journal of Phonetics* 80: 1–19.
Klein, Harriet B., Tara McAllister Byun, Lisa Davidson, and Maria I. Grigos. 2013. A multidimensional investigation of children's /r/ productions: Perceptual, ultrasound and acoustic measures. *American Journal of Speech-Language Pathology* 22: 540–53.
Kochetov, Alexei, Christophe Savariaux, Laurent Lamalle, Camille Noûs, and Pierre Badin. 2023. An MRI-based articulatory analysis of the Kannada dental-retroflex contrast. *Journal of the International Phonetic Association* 54: 227–63.
Kochetov, Alexei, N. Sreedevi, Midula Kasim, and R. Manjula. 2014. Spatial and dynamic aspects of retroflex production: An ultrasound and ema study of Kannada geminate stops. *Journal of Phonetics* 46: 168–84.
Kuznetsova, Alexandra, Per Bruun Brockhoff, and Rune Haubo Bojesen Christensen. 2020. lmerTest. In *R-Package*. Ver. 3.1-3. Available online: https://cran.r-project.org/web/packages/lmerTest/index.html (accessed on 10 November 2024).
Lawson, Eleanor, James M. Scobbie, and Jane Stuart-Smith. 2011. The social stratification of tongue shape for postvocalic /r/ in Scottish English. *Journal of Sociolinguistics* 15: 256–68.
Lawson, Eleanor, James M. Scobbie, and Jane Stuart-Smith. 2013. Bunched /r/ promotes vowel merger to schwar: An ultrasound tongue imaging study of Scottish sociophonetic variation. *Journal of Phonetics* 41: 198–210.
Lawson, Eleanor, James M. Scobbie, and Jane Stuart-Smith. 2014. A socio-articulatory study of Scottish rhoticity. In *Sociolinguistics in Scotland*. Edited by Robert Lawson. Basingstoke: Palgrave Macmillan, pp. 53–78.

Lenth, Russell V. 2023. Emmeans. In *R-Package*. Ver. 1.8.9. Available online: https://cran.r-project.org/web/packages/emmeans/index.html (accessed on 10 November 2024).
Magloughlin, Lyra. 2016. Accounting for variability in North American English /ɹ/: Evidence from children's articulation. *Journal of Phonetics* 54: 51–67.
Martins, Paula, Inês Carbone, Augusto Silva, and António Teixeira. 2008. European Portuguese MRI based speech production studies. *Speech Communication* 50: 925–52.
Mateus, Maria Helena, and Ernesto d'Andrade. 2000. *The Phonology of Portuguese*. Oxford: Oxford University Press.
Mazzaro, Natalia, and Raquel González de Anda. 2020. Men finally got it! Rhotic assibilation in Mexican Spanish in Chihuahua. *Languages* 5: 38.
Mielke, Jeff. 2015. An ultrasound study on Canadian French rhotic vowels with polar smoothing spline comparisons. *Journal of the Acoustical Society of America* 137: 2858–69.
Mielke, Jeff, Adam Baker, and Diana Archangeli. 2016. Individual-level contact limits phonological complexity: Evidence from bunched and retroflex /ɹ/. *Language* 92: 101–40.
Oushiro, Livia. 2021. Avaliações e percepções sociolinguísticas. *Estudos Linguísticos* 50: 318–36.
Oushiro, Livia, and Ronal Beline Mendes. 2013. The nature of sociolinguistic perception. *Revista do GEL* 8: 66–95.
Proctor, Michael. 2011. Towards a gestural characterization of liquids: Evidence from Spanish and Russian. *Journal of Laboratory Phonology* 2: 451–85.
Ramsammy, Michael, and Matthew King. 2023. Edge strengthening and phonetic variability in Spanish /l/: An ultrasound study. *Phonetica* 80: 259–307.
Ramsammy, Michael, and Patrycja Strycharczuk. 2016. From phonetic enhancement to phonological underspecification: Hybrid voicing contrast in European Portuguese. *Papers in Historical Phonology* 1: 285–315.
R Core Team. 2018. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
Recasens, Daniel, and Aina Espinosa. 2007. Phonetic typology and positional allophones for alveolar rhotics in Catalan. *Phonetica* 64: 1–28.
Recasens, Daniel, and Maria Dolors Pallarès. 1999. A study on /ɾ/ and /r/ in the light of the "DAC" coarticulation model. *Journal of Phonetics* 27: 143–69.
Rennicke, Iiris. 2015. Variation and Change in the Rhotics of Brazilian Portuguese. Ph.D. thesis, University of Helsinki, Helsinki, Brazil.
Ricardo, Júlia, and Luiz Carlos da Silva Schwindt. 2023. O rótico retroflexo na Região Metropolitana de Porto Alegre. *Domínios de Linguagem Uberlândia* 17: 1–23.
Scobbie, James M., Reenu Punnoose, and Ghada Khattab. 2013. Articulating five liquids: A single speaker ultrasound study of Malayalam. In *Rhotics: New Data and Perspectives*. Edited by Lorenzo Spreafico and Alessandro Vietti. Bolzano: Bozen Bolzano University Press, pp. 99–124.
Sebregts, Koen. 2014. *The Sociophonetics and Phonology of Dutch r*. Amsterdam: LOT, Netherlands Graduate School of Linguistics.
Strycharczuk, Patrycja, and Koen Sebregts. 2018. Articulatory dynamics of (de)gemination in Dutch. *Journal of Phonetics* 68: 138–49.
Tabain, Marija. 2003. Effects of prosodic boundary on /aC/ sequnces: Articulatory results. *Journal of the Acoustical Society of America* 113: 2834–49.
Tabain, Marija, and Richard Beare. 2018. An ultrasound study of coronal places of articulation in Central Arrernte: Apicals, laminals and rhotics. *Journal of Phonetics* 66: 63–81.
Tiede, Mark K., Suzanne E. Boyce, Carol Y. Espy-Wilson, and Vincent L. Gracco. 2010. Variability in North American /r/ production in response to palatal perturbation. In *Speech Motor Control: New Developments in Basic and Applied Research*. Edited by Ben Maassen and Pascal van Lieshout. Oxford: Oxford University Press, pp. 53–68.
Whitney, William Dwight. 1896. *Sanskrit Grammar: Including Both the Classical Language, and the Older Dialects, of Veda and Brahmana*, 3rd ed. Leipzig: Breitkopf & Härtel.
Wickham, Hadley, Winston Chang, Lionel Henry, Thomas Lin Pedersen, Kohske Takahashi, Claus Wilke, Kara Woo, Hiroaki Yutani, and Dewey Dunnington. 2023. ggplot2. In *R-Package*. Ver. 3.4.4. Available online: https://cran.r-project.org/web/packages/ggplot2/index.html (accessed on 10 November 2024).
Willis, Erik W. 2007. An acoustic study of the 'pre-aspirated trill' in Cibaeño Dominican Spanish. *Journal of the International Phonetic Association* 37: 33–49.
Wood, Simon. 2023. mgcv. In *R-Package*. Ver. 1.9-1. Available online: https://cran.r-project.org/web/packages/mgcv/index.html (accessed on 10 November 2024).
Wrench, Alan. 2003–2023. Articulate Assistant Advannced, Articulate Instruments Ltd. Computer Software. Available online: https://www.articulateinstruments.com (accessed on 10 November 2024).
Zhou, Xinhui, Carol Y. Espy-Wilson, Suzanne Boyce, Mark Tiede, Christy Holland, and Ann Choe. 2008. A magnetic resonance imaging-based articulatory and acoustic study of "retroflex" and "bunched" American English /r/. *Journal of the Acoustical Society of America* 123: 4466–81.

*languages*

MDPI

*Article*

# The Effect of Manner of Articulation and Syllable Affiliation on Tongue Configuration for Catalan Stop–Liquid and Liquid–Stop Sequences: An Ultrasound Study

**Daniel Recasens**

Departament de Filologia Catalana, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain; daniel.recasens@uab.es

**Abstract:** The present study reports tongue configuration data recorded with ultrasound for two sets of consonant sequences uttered by five native Catalan speakers. Articulatory data for the onset cluster pairs [kl]-[ɣl] and [kɾ]-[ɣɾ], and also for [l#k]-[l#ɣ] and [r#k]-[r#ɣ], analyzed in the first part of the investigation revealed that, as a general rule, the (shorter) velar approximant is less constricted than the (longer) voiceless velar stop at the velar and palatal zones while exhibiting a more retracted tongue body at the pharynx. These manner of articulation-dependent differences may extend into the preceding liquid. Data for [k#l]-[kl] and [k#r]-[kɾ] dealt with in the second part of the study show that the velar is articulated with more tongue body retraction for [k#l] vs. [kl] and for [k#r] vs. [kɾ], and with a higher tongue dorsum for [k#l] vs. [kl] and the reverse for [k#r] vs. [kɾ]. Therefore, clusters are produced with a more extreme lingual configuration across a word boundary than in syllable-onset position, which at least in part may be predicted by segmental factors for the [k#r]-[kɾ] pair. These articulatory data are compared with duration data for all sequence pairs.

**Keywords:** stop + liquid and liquid + stop clusters; Catalan; ultrasound

## 1. Introduction

The present study is an ultrasound investigation of changes in tongue position in Catalan two-consonant sequences composed of an underlying stop and an alveolar lateral or rhotic as a function of differences in stop manner of articulation (voiced stop vs. voiced approximant) and syllable affiliation of the two consecutive consonants (heterosyllabic vs. tautosyllabic).

Regarding the first major research topic, articulatory differences as a function of the manner of articulation of the underlying stop consonant will be investigated through a comparison between the postvocalic sequences [kl] and [ɣl], and [kɾ] and [ɣɾ], in syllable-onset position, on the one hand, and between the heterosyllabic sequences [lk] and [lɣ], and [rk] and [rɣ], on the other hand. As to the presence of [ɣ] in the postvocalic sequences [ɣl] and [ɣɾ] and in the postconsonantal ones [lɣ] and [rɣ], it needs to be stated that, analogously to Spanish, Catalan has a phonological stop lenition rule according to which syllable-initial /b d g/ are realized as approximants after any non-stop segment with the exception of the sequences [ld ʎd] and to a large extent [ɾb] as well where C2 is realized as a stop because the two heterosyllabic consonants are homorganic (Recasens 1993). The consonant realizations [β ð ɣ] (which, in contrast with the International Phonetic Alphabet chart, are meant to be not fricative but approximant) are produced with non-turbulent airflow exiting the vocal tract through a relatively wide central constriction, which may occur at the velar zone (for [ɣ]), at the upper teeth ([ð]) or at the lips ([β]). However, as the studies for the Iberian languages mentioned below (see also Martínez-Celdrán 2004) and others for Tuscan Italian (Giannelli and Savoia 1978; Dalcher 2006) have shown, stop lenition is in fact a gradient process such that these approximants may differ a great deal in constriction opening degree especially when occurring postconsonantally, thus varying from more fricative-like to more vowel-like. Thus, it has been found for Spanish, Catalan and Galician in this respect that constriction narrowing for postconsonantal [β ð ɣ] is very much determined by the degree of constriction for the preceding consonant such that more open, less constricted approximant realizations are more prone to occur after sonorants than after fricatives and when the two consecutive consonants are heterorganic than when they are (quasi)-homorganic (Martínez-Celdrán and Regueira 2008; Hualde et al. 2011; Recasens 2015).

Through a comparison of the sequence pairs [lk]-[lɣ] and [rk]-[rɣ], the goal of this part of the study is to elicit whether some approximation of the tongue back to the pharyngeal wall or the velar region for contextual [l] and [r] may render the adjacent approximant [ɣ] fricative-like and therefore nearly as constricted as [k] or at least more constricted than postvocalic [ɣ] occurring in the sequences [ɣl] and [ɣɾ]. Several specific indications about the phonetic implementation of the lateral and the rhotic should be given at this stage. In Catalan, the alveolar lateral /l/ used to be dark, while at the present time darkness degree depends on the speaker and dialectal zone taken into consideration (Recasens 2014, pp. 175–204; Simonet 2010). While dark and clear /l/ share an apicoalveolar place of articulation, the two consonant varieties differ regarding tongue body configuration: clear /l/ is produced with a fronted tongue body and a raised blade and predorsum, which need to be somewhat braced laterally to allow for the passage of airflow through the mouth sides; dark /l/, on the other hand, exhibits a relatively lowered tongue front immediately behind the constriction location and a somewhat retracted postdorsum which approaches the velar and/or pharyngeal articulatory zones (Delattre 1965, p. 89; Ladefoged and Maddieson 1996, pp. 183–93). As to the alveolar rhotic, in Catalan, it is implemented as a tap in syllable-onset clusters and also morpheme-medial intervocalically (i.e., [ɾ] in the sequences [kɾ] and [ɣɾ], as in *acritud* 'acrimony' and *agrair* 'to express one's gratitude to', and also in *pare* 'father'), as a full trill absolute stem initially and after a heterosyllabic consonant (as in *rus* 'Russian', *pre-romà* 'pre-Roman', and also in *honra* 'honor') and as a short trill or as a tap in syllable-coda position (i.e., [r] in the sequences [rk] and [rɣ], as in *arca* 'arch' and *murga* 'lots of noise') (Recasens 1993, pp. 176–78). Regarding the articulatory characteristics of the two rhotic types, the alveolar trill is articulated with two or more fast apicoalveolar contacts (which may be reduced to a single long contact preconsonantally) and, analogously to dark /l/, a somewhat lowered tongue front and some tongue body retraction, while the production of the alveolar tap involves one fast ballistic apicoalveolar contact and a similar tongue body configuration to that for clear /l/ (see Recasens and Espinosa 2007 for Catalan and Proctor 2009 for the two rhotics in Spanish). Given a trend for the dark lateral and the alveolar trill to involve some tongue body retraction towards the velar and pharyngeal zones, it could be that the constriction degree for the velar approximant approaches that for the velar stop or at least that it is greater postconsonantally than postvocalically and therefore for [lɣ] and [rɣ] than for [ɣl] and [ɣɾ].

The first part of the study is also concerned with other aspects about the articulation of the velar approximant [ɣ]. In particular, we will explore the extent to which differences in constriction degree between the velar stop and the velar approximant cooccur with differences in tongue body position behind and in front of the velar zone. It may be that the formation of a wider constriction for [ɣ] than for [k] is assisted by shaping the tongue body back and the tongue predorsum in a certain way and that this consonant-specific lingual configuration serves to facilitate the passage of air throughout the vocal tract for the approximant. This research issue is all the more relevant since, to our knowledge, there are no data on overall tongue body position for [β ð ɣ] in the phonetics literature. On the other hand, we will look into whether there are changes in tongue position during the lateral and the rhotic, which depend on whether these consonants are flanked by the velar approximant or by the velar stop. Regarding the velar approximant, those consonant-to-consonant coarticulatory effects could seek to ease the production of [ɣ] by anticipating its overall tongue posture during the preceding liquid. Results from this research topic should improve our knowledge about the coarticulatory patterns which may occur in consonant sequences by showing that, not differently from stops, the syllable-initial approximants [β ð ɣ] in Catalan and Spanish are produced with active lingual or labial gestures which may overlap with the adjacent phonetic segments in the speech chain.

The present study will also look into whether sequences made of a velar stop C1 and a lateral or a rhotic C2 exhibit differences in lingual configuration depending on syllable affiliation and, therefore, on whether the two consonants occur in syllable-onset position or in different syllables. The motivation for this part of the study is to ascertain whether a syllable boundary effect operates in Catalan such that consonantal sequences which may occur syllable initially are produced less coherently in C#C sequences than word internally. The relationship between tongue configuration and syllable affiliation will be explored by comparing [k#l] to onset [kl] (as in *trec làmines* 'I take out metal plates' vs. *està clar* 'it is clear') and also [k#r] to onset [kɾ] (*sac rodó* 'round sack' vs. *ho creu* 'he/she believes it'). In Catalan, the sequences [k#l] and [k#r] are supposed to exhibit voicing throughout the velar stop due to the application of a regressive assimilation voicing rule, though in this particular case, the velar stop tends to be partly or fully voiceless (Recasens 2014, pp. 342–45). While substantial cross-linguistic research has been carried out on the articulatory implementation of the syllable-initial cluster [kl] (Gibbon et al. 1993; Pouplier et al. 2022), articulatory data about the ways in which this cluster differs from [k#l] have been collected mainly for English and German. Articulatory studies carried out on these Germanic languages reveal that prosodic boundary strength affects the degree of gestural overlap between consecutive consonants such that the stop and the lateral overlap less and lengthen more in the sequence [k#l] than in the onset cluster [kl] (Hardcastle 1985 and Byrd and Choi 2010 for English and Bombien et al. 2013 for German). These articulatory data are consistent with the claim made by Articulatory Phonology (Browman and Goldstein 1990, 1992) that an increase in gestural overlap should result in a decrease in gestural magnitude, thus implying that the stop ought to be longer, less overlapped, and more constricted in [k#l] than in [kl]. Acoustic data for other English consonant sequences also reveal that segmental duration is a systematic correlate of juncture, i.e., consonants lengthen when flanked by a word or a strong morphological boundary in comparison to when they are not (Lehiste 1960; Hoard 1966). While no data for this specific boundary effect appears to be available for either Catalan or other Romance languages, acoustic data for velar + /l/ sequences reported in Redford (2007) show that the effect in question is not universal: she found that the stop closure is enhanced across a word boundary ([k#l]) vs. syllable initially ([kl]) in English but less so or not at all in Finnish and Russian. Taking these data into consideration, the present study will look into whether the presence of a syllable/word boundary between two heterosyllabic consonants affects the articulatory manifestation and duration of at least the velar stop in Catalan such that it has a longer duration and is produced with a more extensive tongue dorsum gesture and thus a narrower dorsovelar constriction in [k#l] and [k#r] than in [kl] and [kɾ], respectively.

Another factor could play a role in the phonetic implementation of velar + rhotic clusters depending on whether occurring syllable initially or across a word boundary. The fact that the rhotic is realized as a tap in the onset cluster ([kɾ]) and as a full trill in the heterosyllabic sequence ([k#r]) could impinge on the articulatory configuration of the velar stop and on the characteristics of the C1-to-C2 lingual movement. In particular, in so far as the overall tongue body configuration for velars resembles that for [r] rather than that for [ɾ], one would expect less postdorsal movement and a more constricted velar stop in the sequence with the former rhotic ([k#r]) than in that with the latter ([kɾ]). If this is the case, both the syllable affiliation and the C2-to-C1 coarticulatory effects could lead to a similar tongue configuration during the velar consonant. It follows from this that the boundary effect under investigation should be ascertained in a more straightforward way in sequences with /l/ than in those with the rhotic in so far as in Catalan, the rhotic shows relevant articulatory differences associated with manner of articulation in the sequences [kɾ] and [k#r].

To recapitulate, the present investigation will look into differences in constriction degree and overall tongue configuration for [ɣ] as a function of contextual liquids sharing a similar tongue body configuration at the back of the vocal tract and into the extent to which the lingual gesture for the velar approximant is anticipated during the preceding liquid. The second major research topic deals with the syllable/word boundary effect on consonant articulation in the sequence pairs [kl]/[k#l] and [kɾ]/[k#r]. Results from this investigation should contribute to a better understanding of the articulatory characteristics and coarticulatory behavior of the velar approximant while throwing some light onto the production constraints on approximants of other places of articulation as well.

Ultrasound should allow for these issues to be looked into since it provides data on the configuration of the body of the tongue within the vocal tract and, consequently, about tongue postdorsum placement at the velar and pharyngeal regions. This recording and analysis technique has been previously used by ourselves in order to examine the tongue configuration characteristics of Catalan consonants (see, for example, Recasens and Rodríguez 2017), which puts us in a good position to interpret the articulatory data for the velar sequences with a liquid subject to analysis in the present investigation. In comparison to other techniques which are commonly used for the study of the articulation of consonants, such as electropalatography, electromagnetic articulography, and magnetic resonance (Kochetov 2020), an advantage of ultrasound is that it provides data on tongue postdorsum position at the upper and lower pharynx and at the velar zone and on tongue predorsum height at the palatal zone. A problem with this recording technique is that it barely provides data at the tongue tip and the tongue root since the mandible and hyoid bones refract the sound before it reaches the tongue surface, thus creating a black region where the tongue tip and the tongue root are located (see Stone 2005). This, however, should not be of great concern to the present investigation since much of the tongue configuration data of interest correspond to the palatal, velar, and upper pharyngeal regions of the vocal tract.

## 2. Materials and Methods

Tongue configuration and acoustic data were collected for the following pairs of consonant sequences: the syllable-onset clusters [kl]-[ɣl] and [kɾ]-[ɣɾ]; the heterosyllabic sequences [l#k]-[l#ɣ], [r#k]- [r#ɣ], and [k#l]-[k#r]. All sequences of consonants occurred in intervocalic position and were embedded in the meaningful Catalan sentences listed in Table 1. Sentences had four syllables and lexical stresses falling on the vowels preceding and following C1 and C2, respectively.

**Table 1.** List of Catalan sentences with English glosses. The consonant sequences under analysis appear underlined.

| | | |
|---|---|---|
| 1. [kl] | un dinar clau | 'an important dinner' |
| 2. [ɣl] | ell pujà glaç | 'he brought up ice' |
| 3. [kɾ] | i menjà cranc | 'and (s)he ate crab' |
| 4. [ɣɾ] | ell pujà grades | 'he climbed steps' |
| 5. [l#k] | animal car | 'expensive animal' |
| 6. [l#ɣ] | va matar el gall | '(s)he killed the rooster' |
| 7. [r#k] | un radar car | 'an expensive radar' |
| 8. [r#ɣ] | és un bar gal | 'it is a French bar' |
| 9. [k#l] | és un frac làbil | 'it is an outmoded tailcoat' |
| 10. [k#r] | un atac ràpid | 'a fast attack' |

Sentences were recorded six times by five native Catalan speakers, i.e., two men (DR, RO) and three women (ES, JU, IM) of 40–60 years of age who speak Catalan on a regular basis in their everyday lives. Ultrasound recordings were performed with an Echo Blaster unit type EB128CEXT from TELEMED and a microconvex Echo Blaster 128 CEXT transducer with a 2 to 4 MHz frequency range and a central curvature of 20 mm. The ultrasound images were acquired using a probe with a 100% of 104° field of view and a frequency of 2 MHz, which was attached to a transducer holder positioned under the subject's chin in an Articulate Instruments Stabilization Headset. The recording sampling rate was 54 frames per second, yielding one image every 18.5 ms. Image streams were recorded synchronously with the audio signal sampled at 22,050 Hz with an AKG-D70 microphone. Contours of the back of the alveolar zone and hard palate were also recorded by asking speakers to press the tongue against their hard palate. Tongue contours were tracked automatically at all temporal frames along each C#C sequence token for each speaker using the Articulate Assistant Advanced (AAA) software and adjusted manually by the paper author. Data points for all tongue contours were exported in ASCII files as x-y coordinates with their origin located at the bottom-left corner of the ultrasound image towards the rear of the vocal tract. Acoustic files were also exported in .wav format in order to extract segmental duration measures.

Segmentation was carried out on waveform and spectrographic displays. The boundary between C1 and C2 was taken to occur at the following events: in stop/approximant + liquid sequences, two temporal points were identified in case a C1 stop burst was present (i.e., C1 closure offset before the short stop burst, C2 onset immediately after the C1 stop burst), and a single temporal point at the beginning of the formant structure for the liquid in the absence of a burst; in liquid + stop/approximant sequences, we identified a single point at the offset of the C1 formant structure or else two temporal points at C1 offset and C2 onset whenever there was a vocalic portion at the offset of a rhotic C1. Depending on whether the boundary between C1 and C2 consisted of one or two temporal points, the lingual splines were processed at six points in time (at C1 onset, C1 midpoint, C1 offset, C2 onset, C2 midpoint, C2 offset) or at five points (at C1 onset, C1 midpoint, C1 offset/C2 onset, C2 midpoint, C2 offset). In the former case and as referred to above, the interval between C1 offset and C2 onset could be occupied by a stop burst or by a vocalic portion.

Tongue spline data points were converted from Cartesian to polar coordinates (Mielke 2015) by shifting the origin of the ultrasound image to approximately the center of the ultrasound probe, which was located at X = 86.7 mm and Y = 0 mm. SSANOVA smoothed splines consisting of strings of points separated by 0.01 radians with the associated standard errors were computed across the splines for all tokens of each consonant sequence using the R package gss to find a best-fit curve (Davidson 2006; Gu 2014). The rightmost and leftmost edges of the smoothed splines were determined by entering into the SSANOVA computation procedure the corresponding mean angle radian values across all tokens of the clusters of interest.

In order to elicit differences in tongue position at different regions of the tongue, the length of the SSANOVA splines displayed in Cartesian coordinates was divided into four portions which correspond to different articulatory zones, namely, alveolar (ALV), palatal (PAL), velar (VEL) and pharyngeal (PHAR), separately for each subject. This subdivision procedure was carried out by applying the same criterion as in the previous study Recasens and Rodríguez (2017) since the data for the clusters subject to analysis in the two studies were acquired in the same recording session. As described next, the criterion for determining the four articulatory zones involved the use of a different set of consonants and consonantal sequences from those subjected to analysis in this manuscript: the boundary between the alveolar and the dental zone was identified at an inflection point occurring at the spline front edge during dental /t/ in the sequence /pt/ and that between the alveolar and palatal zones at another inflection point located at the back alveolar area during the trill /r/ in the sequence /pr/ (/r/ is postalveolar in Catalan); the boundary between the palatal and velar zones was placed at the closure location for the velar stop in the sequence /iki/ which according to EPG data is articulated at the postpalatal zone, just in front of the soft palate, in Catalan; the length of the velar zone was taken to be 1.25 and 1.51 times that of the palatal zone in the case of the male and female speakers, respectively, as reported by Fitch and Giedd (1999); finally, the pharyngeal zone extended between the left edge of the velar zone all the way until the bottom edge of the lingual splines.

Distances between each of the four lingual regions and the origin of the ultrasound field of view were measured at all temporal points selected for analysis. The distance values at the velar (VEL) and palatal (PAL) zones were obtained by averaging the distances between the five central points at each zone and the origin. A different evaluation criterion was applied to the two extreme zones, alveolar (ALV) and pharyngeal (PHAR), in view of the fact that the splines for the consonant sequences subject to analysis could differ in length: in this case, we averaged the distances between the origin and five points located not at the zone midpoint but further away from the outer edges of the two zones in question, namely, at the upper third of the pharyngeal zone and at the leftmost third of the alveolar zone.

Linear Mixed Model (LMM) statistical tests were run on the distance values gathered at the midpoints of C1 and C2 with speaker as a random effect. Separate tests were performed on data for each of the cluster pairs of interest, i.e., [kl]-[ɣl], [kɾ]-[ɣɾ], [lk]-[lɣ], [rk]-[rɣ], [k#l]-[kl], and [k#r]-[kɾ]. The LMM tests had the fixed effects 'sequence' (with 2 levels for each pair of consonant sequences, as for example [kl] and [ɣl] in the case of the pair [kl]-[ɣl]), 'C place' (with 2 levels corresponding to the velar and alveolar places of articulation) and 'zone' (with the 4 levels ALV, PAL, VEL, and PHAR). Additional LMM tests, one for each cluster pair, were also carried out on the cluster duration data using the duration values for all cluster tokens with 'sequence' and 'C place' as fixed effects. Least Significance Difference (LSD) post-hoc tests were run on all main effects and significant interactions in order to find out whether numerical differences between pairs of levels of a given statistical variable reached significance or not. Given the large number of tests involved in the LMM analyses, the Benjamini–Hochberg (BH) correction procedure for adjusting the false discovery rate was applied to those variable comparisons that were of relevance to the present investigation. In all statistical tests, the significance level was set at $p < 0.05$.

## 3. Results

Statistical results for the main effects and factor interactions obtained from the LMM tests run on the distance and duration values are given for all pairs of clusters in Tables 2–5. These statistical results will be commented on with reference to the lingual position and duration data and to the tongue profiles displayed in Figures 1–6.

Statistical differences among distance values for different articulatory zones, and thus for the main zone effect and for the C place × zone interaction, will not be reported. It suffices to say in this respect that distances between the tongue and the origin of the ultrasound field of view were consistently larger at the velar (VEL) and palatal (PAL) zones than at the more extreme pharyngeal (PHAR) and alveolar (ALV) zones. Moreover, they were larger at the velar zone for velars than for alveolars and the reverse at the front of the vocal tract, which is in agreement with differences in place of articulation between the two consonant classes.

### *3.1. Effect of Manner of Articulation*

#### 3.1.1. Lingual Configuration

Table 2 shows the statistical results for the distance values between the tongue and the origin of the ultrasound field of view in the case of clusters with the velar stop and the velar approximant. They indicate a significant main effect of sequence and C place of articulation for the [rk]-[rɣ] pair, which is associated with larger distances for [rk] vs. [rɣ] and for the velar vs. the rhotic. There are two relevant significant interactions, i.e., sequence × zone and sequence × zone × C place. According to the former interaction, tongue distances are significantly greater for sequences with a voiced approximant than for those with a voiceless stop at PHAR ([kɾ]-[ɣɾ], [lk]-[lɣ]) and the reverse at PAL ([kɾ]-[ɣɾ], [lk]-[lɣ], [rk]-[rɣ]). The triple sequence × zone × C place interaction, which achieves significance only for [lk]-[lɣ], is associated with larger distances between [k] and [ɣ] at PHAR and PAL when compared to the two other articulatory zones.

Data for the distance values at successive points in time for the four pairs of sequences plotted in Figure 1 reveal indeed the presence of larger distances at the pharynx for the sequences with the voiced approximant (discontinuous lines) than for those with the voiceless stop (continuous lines), and the opposite relationship at the velar and palatal zones. They also show that, while differences in tongue position between the two sequence pairs occur mainly during the velar consonant (at C1 in the case of [kl]-[ɣl] and [kɾ]-[ɣɾ], at C2 in the case of [lk]-[lɣ] and [rk]-[rɣ]), they may also extend into the adjacent liquid. This coarticulatory effect applies mostly to the coda consonants [l] and [r], which tend to be either more retracted at PHAR and thus at the back of the vocal tract (mostly [l]) or lower at PAL and thus at the palatal zone (mostly [r]) whenever occurring before [ɣ] than before [k].

**Table 2.** F and *p* values for the main effects and factor interactions obtained from LMM analyses run on the distance values between the tongue surface and the origin of the ultrasound field of view for cluster pairs with a velar stop and a velar approximant (*, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$). NS: non-significant effect.

| | kl-ɣl | kɾ-ɣɾ | lk-lɣ | rk-rɣ |
|---|---|---|---|---|
| Sequence | NS | NS | NS | 4.86 (1, 52) * |
| C place | NS | NS | NS | 65.13 (1, 52) *** |
| Zone | 8.93 (3, 54) ** | 9.02 (3, 44) ** | 9.33 (3, 48) ** | 6.80 (3, 52) ** |
| Sequence × zone | NS | 6.63 (3, 44) *** | 11.19 (3, 48) *** | 3.50 (3, 52) * |
| C place × zone | 12.83 (3, 54) *** | 70.67 (3, 44) *** | 22.56 (3, 48) *** | 29.55 (3, 52) *** |
| Sequence × C place | NS | NS | NS | NS |
| Sequence × zone × C place | NS | NS | 4.53 (3, 48) ** | NS |

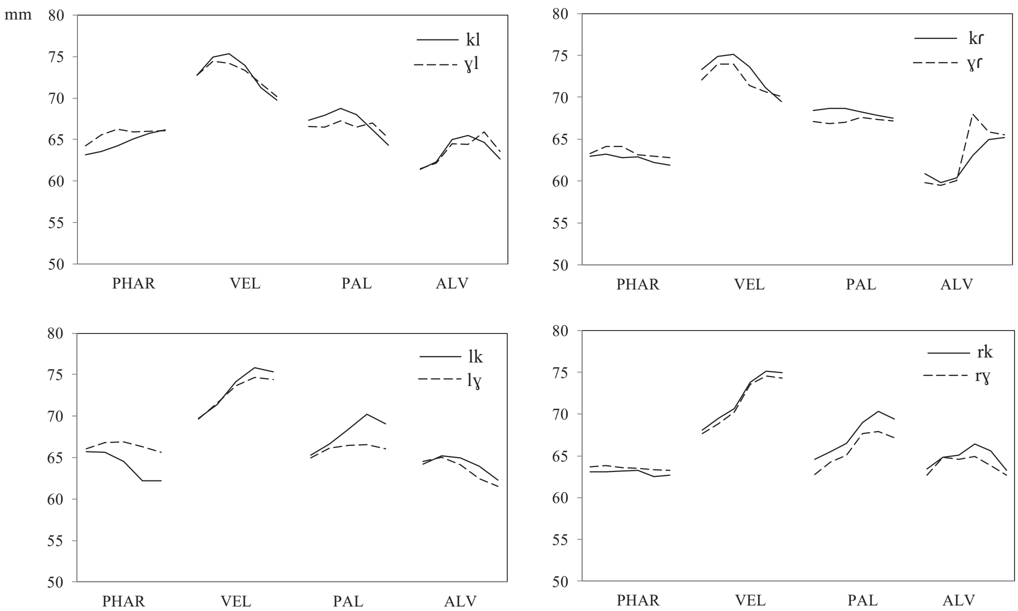



**Figure 1.** Cross-speaker distance measures between the tongue surface and the origin of the ultrasound field of view sampled at consecutive temporal points during the sequence pairs [kl]-[ɣl], [kɾ]-[ɣɾ], [lk]-[lɣ], and [rk]-[rɣ]. The distance trajectories for each articulatory zone proceed from C1 onset (leftmost edge) to C2 offset (rightmost edge) through intermediate temporal points. PHAR = pharyngeal, VEL = velar, PAL = palatal, ALV = alveolar.

The distance data for [kl]-[ɣl] and [kɾ]-[ɣɾ] plotted in Figure 1 are in agreement with the lingual configurations represented in Figure 2 for speaker JU. Figure 2 reveals indeed that, in comparison with C1 = [k], C1 = [ɣ] tends to be produced with a lower tongue dorsum position and thus a wider dorsal constriction at the palatal zone and a more retracted postdorsum at the pharynx.

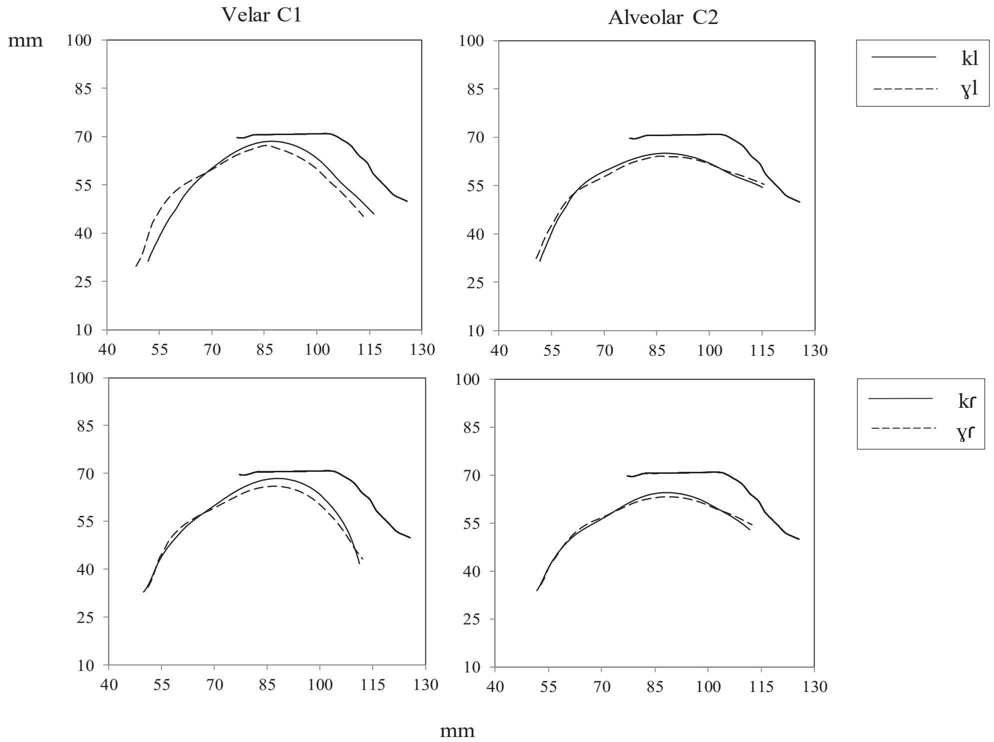



**Figure 2.** Tongue configurations at the midpoint of C1 and C2 for [kl] and [ɣl] (**top**) and [kɾ] and [ɣɾ] (**bottom**) with the palate trace superimposed according to speaker JU. The front of the mouth is on the right of the graphs.

### 3.1.2. Duration

Figure 3 reports duration data for the two consonants of all consonant sequences subjected to analysis in this part of the study. According to the corresponding statistical results provided in Table 3, the clusters with a rhotic show a main sequence effect and a main C place effect, which happen to be associated, respectively, with a longer duration for [kɾ] vs. [ɣɾ] and for [rk] vs. [rɣ] and with a shorter duration for the rhotic than for the velar consonant whether in onset or coda position. There is also a significant sequence × C place interaction for the [rk]-[rɣ] sequence pair, which turned out to be related to differences in duration for velar > rhotic and for [k] > [ɣ] and a somewhat longer rhotic before [ɣ] than before [k]. Figure 3 also shows that the alveolar tap is longer after [k] than after [ɣ] in the sequence pair [kɾ]-[ɣɾ], though this difference did not achieve significance. As to the clusters with the alveolar lateral, there is only a significant sequence × C place interaction for [kl]-[ɣl], which is associated with a longer velar consonant than alveolar liquid in the case of [kl] and the reverse for [ɣl], and a longer velar for [kl] vs. [ɣl] and a longer alveolar lateral for [ɣl] vs. [kl].

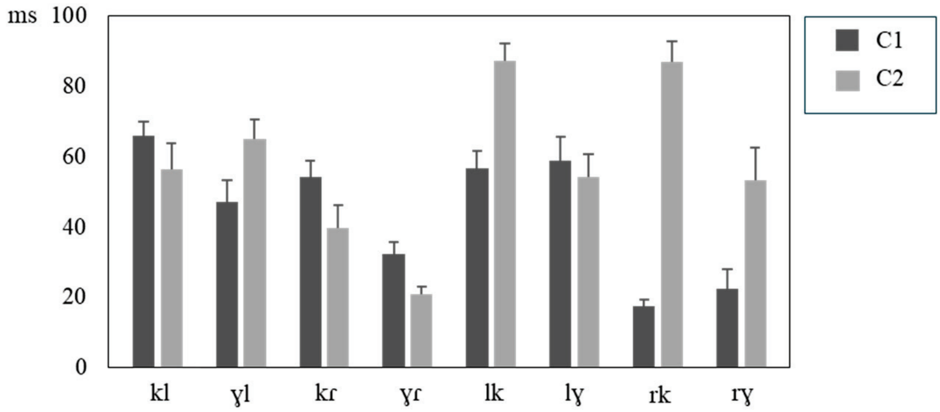



**Figure 3.** Consonant durations and standard deviation values for cluster pairs differing in manner of articulation for the velar consonant.

**Table 3.** F and *p* values for the main effects and factor interaction obtained from LMM analyses run on the duration values for cluster pairs with a velar stop and a velar approximant (*, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$). NS: non-significant effect.

| | kl-ɣl | kɾ-ɣɾ | lk-lɣ | rk-rɣ |
|---|---|---|---|---|
| Sequence | NS | 50.84 (1, 100) ** | NS | 12.59 (1, 100) * |
| C place | NS | 16.96 (1, 100) * | NS | 149.5 (1, 100) *** |
| Sequence × C place | 14.7 (1, 100) * | NS | NS | 10.16 (1, 100) * |

A longer rhotic before [ɣ] than before [k] and a longer lateral after [ɣ] than after [k] indicate that the C1-to-C2 closing movement for [rk] and [rɣ] and the C1-to-C2 opening movement for [kl] and [ɣl] proceed more slowly in sequences with the approximant [ɣ] than in those with the stop [k].

### *3.2. Syllable Affiliation Effect*

#### 3.2.1. Lingual Configuration

Statistical results for the distance data between the tongue surface and the origin of the ultrasound field of view for the heterosyllabic stop + liquid sequences are given in Table 4. Results from the post-hoc test run on the main sequence effect reveal significantly greater distances across zones for [k#l] than for [kl] but no differences between [k#r] and [kɾ], and those performed on the main C place effect, larger distances for the velar than for the liquids. Moreover, a significant sequence × zone interaction turned out to be associated with a more distant tongue position from the origin for [k#r] than for [kɾ] at ALV and PHAR and the reverse at PAL. These sequence- and consonant-dependent differences may be seen in the tongue distance trajectories over time plotted in Figure 4 and are also apparent in the lingual splines for speaker RO displayed in Figure 5. There is then a trend for the voiceless velar stop to be produced with a more extreme articulation in heterosyllabic than in tautosyllabic sequences. As suggested in the Introduction section, in sequences with the alveolar rhotic, this difference in tongue configuration could be associated not only with the presence vs. absence of a syllable juncture but also with differences in C2-to-C1 coarticulation in so far as the alveolar trill is articulated with a somewhat lower front dorsum and a more retracted tongue body than the alveolar tap.

**Table 4.** F and *p* values for the main effects and factor interactions obtained from LMM analyses run on the distance values between the tongue surface and the origin of the ultrasound field of view for cluster pairs differing in word boundary availability (*, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$). NS: non-significant effect.

| | k#l-kl | k#r-kɾ |
|---|---|---|
| Sequence | 6.82 (1, 52) * | NS |
| C place | 6.16 (1, 52) * | 7.39 (1, 50) ** |
| Zone | 7.84 (3, 52) ** | 9.16 (3, 50) ** |
| Sequence × zone | NS | 5.79 (3, 50) ** |
| C place × zone | 11.94 (3, 52) *** | 32.66 (3, 50) *** |
| Sequence × C place | NS | NS |
| Sequence × zone × C place | NS | NS |

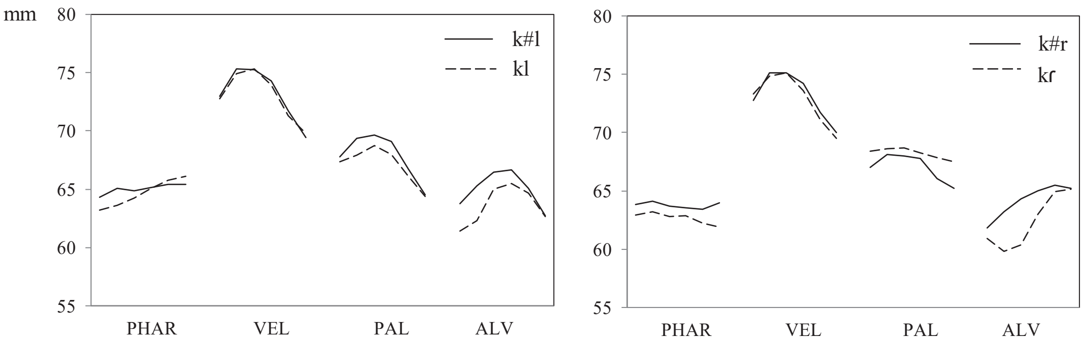



**Figure 4.** Cross-speaker distance measures between the tongue surface and the origin of the ultrasound field of view sampled at consecutive temporal points during the sequence pairs [k#l]-[kl] and [k#r]-[kɾ]. The distance trajectories for each articulatory zone proceed from C1 onset (leftmost edge) to C2 offset (rightmost edge) through intermediate temporal points. PHAR = pharyngeal, VEL = velar, PAL = palatal, ALV = alveolar.

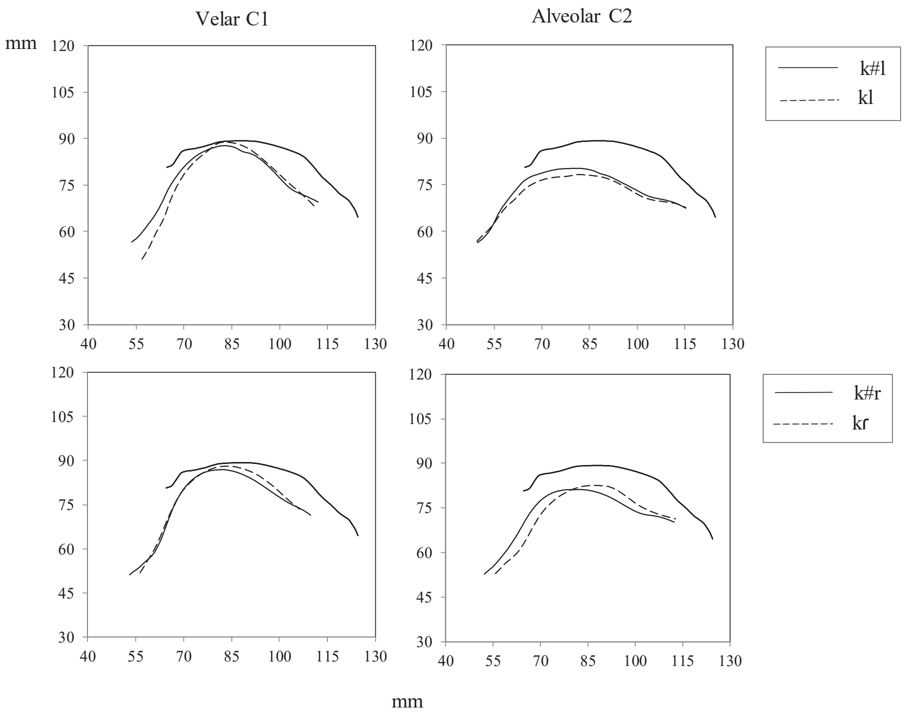



**Figure 5.** Tongue configurations at the midpoint of C1 and C2 for [k#l] and [kl] (**top**) and [k#r] and [kɾ] (**bottom**) with the palate trace superimposed according to speaker RO. The front of the mouth is on the right of the graphs.

3.2.2. Duration

Mean duration values for the cross-boundary and syllable-onset consonant sequences and the corresponding statistical results are provided in Figure 6 and Table 5, respectively. According to the data reported in the figure and the table, while [k#l] and [kl] do not differ significantly from each other, there is a significantly greater duration for [k#r] than for [kɾ] (main sequence effect), which is associated with a much longer realization for the trill [r] than for the tap [ɾ] (sequence × C place interaction).

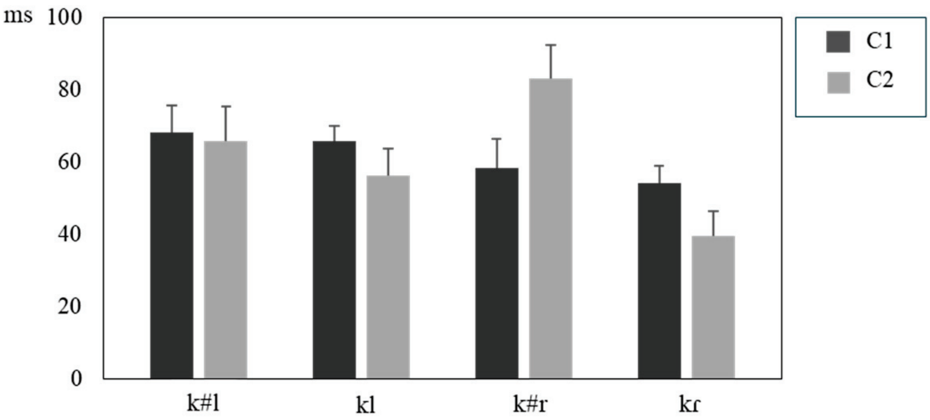



**Figure 6.** Consonant durations and standard deviation values for cluster pairs differing in word boundary availability.

**Table 5.** F and *p* values for the main effects and factor interaction obtained from LMM analyses run on the duration values for cluster pairs differing in word boundary availability (*, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$). NS: non-significant effect.

| | k#l-kl | k#r-kɾ |
|---|---|---|
| Sequence | NS | 15.32 (1, 100) * |
| C place | NS | NS |
| Sequence × C place | NS | 10.77 (1, 100) * |

## 4. Discussion

Articulatory data evaluated in terms of spatial distance for the cluster pairs [kl]-[ɣl], [kɾ]-[ɣɾ], [lk]-[lɣ], and [rk]-[rɣ] show that the (shorter) velar approximant is less constricted than the (longer) voiceless velar stop at the velar/palatal zone while exhibiting a more retracted tongue body at the pharynx as a general rule. Therefore, it appears that the loss of central contact at the velar constriction location (and also at the palatal zone where the predorsum is located) for the approximant as compared to the voiceless stop occurs concomitantly with some postdorsum backing. This finding about tongue position behind and in front of the constriction location for the velar approximant, which has not been reported previously in the phonetics literature, may result from the formation of a wide velar constriction while also being associated with the need to allow continuous airflow through the oral cavity. It is important to emphasize that differences in overall tongue body configuration between [k] and [ɣ] are not the same as those between [k] and [g], which should be associated with the voicing contrast between the two stops but not with a difference in manner of articulation. In comparison to voiceless stops, voiced stops of the same place of articulation in English and other languages exhibit an active expansion of the supraglottal cavity system, which is achieved mainly by fronting the tongue body and thus enlarging the pharyngeal cavity and, albeit less consistently, by lowering the tongue front and thus enlarging the oral cavity (Westbury 1983; Ahn 2018). These changes in lingual configuration facilitate the maintenance of voicing by lowering the intraoral pressure level above the glottis. This piece of evidence confirms that a more retracted tongue body (and, to a large extent, a lower tongue predorsum position) for [ɣ] with respect to [k] depends not on differences in voicing but on differences in the manner of articulation between the two consonantal realizations.

Visual inspection of the distance trajectories reported in Figure 1 reveals essentially the same tongue position for [ɣ] after a vowel and after the lateral and the rhotic. Thus, while the lateral and the rhotic may involve some tongue body retraction, this retraction movement does not appear to be prominent enough to cause the constriction narrowing degree for [ɣ] to approach that for [k], which does not match with the hypothesis set in the Introduction that there should be differences in dorsovelar constriction degree between postvocalic and postconsonantal [ɣ]. This finding may be related to the fact that there is no central closure at the velar zone for [l] or [r], which could block the passage of air through the dorsal constriction for [ɣ]. Moreover, the fact that constriction degree is essentially the same postvocalically as postconsonantally may also be related to the difficulty involved in forming a complete dorsal closure at the soft palate for true velars, which may result in air leakage through the dorsal constriction, most especially when /g/ is preceded by a non-stop consonant (Kingston 2008, p. 21). Therefore, it turns out that /g/ is likely to be realized either as a stop or as a quite unconstricted approximant, much independent of the degree of opening for the preceding phonetic segment. In line with this possibility, previous studies have reported a greater lenition degree for /g/ than for /b/ and /d/ (see, for example, Dalcher 2006 for Florentine Italian and Tang et al. 2023 for Spanish).

Manner of articulation-dependent differences in tongue configuration between the stop and the approximant were found to extend into the preceding liquid in clusters with a coda lateral or rhotic, whether at the back of the vocal tract or at the palatal zone. Moreover, [k] turned out to be longer than [ɣ], as expected, while the liquid was longer when flanked by [ɣ] than by [k], which may be attributed to slower rates of articulatory movement as the lateral moves towards and out of an approximant vs. a stop. Evidence for a much slower rate of increase of the constriction area for approximants than for stops has been reported in the literature (Stevens 1999, p. 532). These findings provide some support for the implementation of syllable-initial [β ð ɣ] in Catalan and Spanish by means of labial or lingual gestures, which may be activated during the preceding phonetic segment in the speech chain.

In agreement with English and German data provided in the Introduction section (Byrd and Choi 2010; Bombien et al. 2013), the Catalan voiceless velar stop + liquid sequences were produced with a more extreme lingual configuration and thus less gestural overlap whenever occurring across a word boundary than syllable initially. This more prominent lingual configuration is implemented differently in sequences with an alveolar lateral and in those with an alveolar rhotic due presumably to the articulatory requirements involved in the production of the two liquids. Indeed, in stop + liquid sequences across a word boundary, the tongue body for [k] is more retracted at the pharynx (PHAR) and may be higher at the palatal zone (PAL) whenever C2 is [l], and also more retracted at PHAR but lower at PAL when C2 is the rhotic. Even though the ultrasound data does not provide information about the degree of dorsal contact at the velar place of articulation, differences in tongue configuration between [k#l] and [kl], i.e., a higher predorsum and a more advanced postdorsum in the former sequence than in the latter, suggest that the velar stop is produced with a greater dorsovelar closure area across a word boundary than word internally. As to the consonantal sequences with the rhotic, rhotic type may account for a lower predorsum position and a more retracted postdorsum for C1 in the case of [k#r] vs. [kɾ] since the alveolar trill is expected to be articulated with less predorsum height and more postdorsum backing than the alveolar tap (see Section 1).

In contrast with data for the Germanic languages showing duration differences for [k#l] > [kl], in Catalan, differences in articulatory displacement between velar stop + liquid sequences with and without a syllable boundary turned out not to be clearly correlated with the corresponding duration data in so far as [k#l] was not statistically longer than [kl] while duration differences between [k#r] and [kɾ] could be due to rhotic type rather than to syllable affiliation. The finding that /k#l/ differs from /kl/ in articulation but not in duration suggests that when produced in a VC#CV string, consonantal sequences that may occur syllable initially are produced more coherently and thus overlap to a larger extent in Catalan (and perhaps in other Romance languages as well) than in English or German even though the syllabification pattern is essentially the same, i.e., (VC)(CV), in all cases. This finding, which awaits further investigation, could be associated with a stronger trend for phonetic segments to lengthen and shorten and thus to be less isochronous, depending on the position that they occupy within the foot unit or the interstress interval in stress-timed languages like German and English than in syllable-timed languages like Catalan (Dauer 1983; Roseano et al. 2022).

The findings of the present investigation may be summarized as follows. The first part of the paper brings about new data for Catalan on the articulation of the velar approximant, mostly regarding its overall lingual configuration characteristics. The ultrasound data reveal that [ɣ] is implemented through a specific tongue configuration, which could serve an aerodynamic goal and small changes in constriction width as a function of context, which may follow from the difficulty involved in forming a complete dorsal closure at the soft palate. Moreover, coarticulatory effects in segmental articulation and duration suggest that [ɣ] is produced with an active articulatory gesture, which is prepared ahead of time during the preceding liquid. As to the effect of a syllable/word boundary on sequences composed of /k/ and a following liquid, the velar stop turned out to exhibit a more extreme lingual configuration but not a longer closure when flanked by a syllable/word boundary, a finding which in the case of rhotic clusters goes hand in hand with differences in manner of articulation for the rhotic. All in all, it is hypothesized that the boundary effect in question may be more straightforward in English and German than in Catalan and perhaps other Romance languages and that this could be related to differences in rhythmic organization between stress-timed and syllable-timed languages. Future research could explore context-dependent differences in articulatory implementation (both in constriction opening and in overall tongue body configuration) for stops vs. approximants of labial, dental, and velar places of articulation and differences in the boundary effect on articulation for velar + liquid clusters among different language types.

**Funding:** This research was funded by project PGC2018-096877-B-I00 of the Spanish Ministry of Science, Innovation and Universities and by the research group 2017 SGR 34 of the Catalan Government.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Additional data supporting the experimental results reported in the present study are unavailable.

**Acknowledgments:** I would like to thank Clara Rodríguez for her help with the data processing and statistical analysis and three reviewers for their insightful comments on a previous version of the manuscript.

**Conflicts of Interest:** The author declares no conflicts of interest.

## References

Ahn, Suzy. 2018. The role of tongue position in laryngeal contrasts: An ultrasound study of English and Brazilian Portuguese. *Journal of Phonetics* 71: 451–67. [CrossRef]

Bombien, Lasse, Christine Mooshammer, and Philip Hoole. 2013. Articulatory coordination in word-initial clusters of German. *Journal of Phonetics* 41: 546–61. [CrossRef]

Browman, Catherine P., and Louis Goldstein. 1990. Tiers in articulatory phonology, with some implications for casual speech. In *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*. Edited by John Kingston and Mary E. Beckman. Cambridge: Cambridge University Press, pp. 341–76.

Browman, Catherine P., and Louis Goldstein. 1992. Articulatory phonology: An overview. *Phonetica* 49: 222–34. [CrossRef]
Byrd, Dani, and Susie Choi. 2010. At the juncture of prosody, phonology, and phonetics. The interaction of phrasal and syllable structure in shaping the timing of consonant gestures. In *Laboratory Phonology 10*. Edited by Cécile Fougeron, Barbara Kühnert, Mariapaola D'Imperio and Nathalie Vallée. Berlin and New York: De Gruyter, pp. 31–59.
Dalcher, Christina V. 2006. Consonant Weakening in Florentine Italian: An Acoustic Study of Gradient and Variable Sound Change. Ph.D. dissertation, Georgetown University, Washington, DC, USA.
Dauer, Rebecca M. 1983. Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics* 11: 51–62. [CrossRef]
Davidson, Lisa. 2006. Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *Journal of the Acoustical Society of America* 120: 407–15. [CrossRef] [PubMed]
Delattre, Pierre. 1965. *Comparing the Phonetic Features of English, French, German and Spanish: An Interim Report*. Heidelberg: Julius Gross Verlag.
Fitch, W. Tecumseh, and Jay Giedd. 1999. Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *Journal of the Acoustical Society of America* 106: 1511–22. [CrossRef] [PubMed]
Giannelli, Luciano, and Leonardo M. Savoia. 1978. L'indebolimento consonantico in Toscana I. *Rivista Italiana di Dialettologia* 2: 25–58.
Gibbon, Fiona, William Hardcastle, and Katerina Nicolaidis. 1993. Temporal and spatial aspects of lingual coarticulation in /kl/ sequences: A cross-linguistic investigation. *Language and Speech* 36: 261–77. [CrossRef] [PubMed]
Gu, Chong. 2014. Smoothing spline ANOVA models: R Package gss. *Journal of Statistical Software* 58: 1–25. [CrossRef]
Hardcastle, William J. 1985. Some phonetic and syntactic constraints on lingual coarticulation during /kl/ sequences. *Speech Communication* 4: 247–63. [CrossRef]
Hoard, James E. 1966. Juncture and syllable structure in English. *Phonetica* 15: 96–109. [CrossRef]
Hualde, José Ignacio, Ryan Shosted, and Daniel Scarpace. 2011. Acoustics and articulation of Spanish /d/ spirantization. Paper presented at XVII International Congress of Phonetic Sciences, Hong Kong, China, August 17–21; Edited by Wai Sum Lee and Eric Zee. Hong Kong: City University of Hong Kong, pp. 906–9.
Kingston, John. 2008. Lenition. In *Selected Proceedings of the 3rd Conference on Laboratory Approaches to Spanish Phonology*. Edited by Laura Colantoni and Jeffrey Steele. Somerville: Cascadilla Press, pp. 1–31.
Kochetov, Alexei. 2020. Research methods in articulatory phonetics I: Introduction and studying oral gestures. *Language and Linguistics Compass* 14: e12368. [CrossRef]
Ladefoged, Peter, and Ian Maddieson. 1996. *The Sounds of the World's Languages*. Oxford: Blackwell.
Lehiste, Ilse. 1960. An acoustic-phonetic study of internal open juncture. *Phonetica* 5 S1: 5–54. [CrossRef]
Martínez-Celdrán, Eugenio. 2004. Problems in the classification of approximants. *Journal of the International Phonetic Association* 34: 201–9. [CrossRef]
Martínez-Celdrán, Eugenio, and Xosé L. Regueira. 2008. Spirant approximants in Galician. *Journal of the International Phonetic Association* 38: 51–68. [CrossRef]
Mielke, Jeff. 2015. An ultrasound study of Canadian French rhotic vowels with polar smoothing spline comparisons. *Journal of the Acoustical Society of America* 137: 2858–69. [CrossRef] [PubMed]
Pouplier, Marianne, Manfred Pastätter, Philip Hoole, Stefania Marin, Ioana Chitoran, Tomas Lentz, and Alexei Kochetov. 2022. Language and cluster-specific effects in the timing of onset consonant sequences in seven languages. *Journal of Phonetics* 93: 1–28. [CrossRef]
Proctor, Michael I. 2009. Gestural Characterization of a Phonological Class: The Liquids. Ph.D. dissertation, Yale University, New Haven, CT, USA.
Recasens, Daniel. 1993. *Fonètica i fonologia*. Barcelona: Enciclopèdia Catalana.
Recasens, Daniel. 2014. *Fonètica i fonologia experimentals del català*. Barcelona: Institut d'Estudis Catalans.
Recasens, Daniel. 2015. The effect of contextual consonants on voiced stop lenition: Evidence from Catalan. *Language and Speech* 59: 1–23. [CrossRef]
Recasens, Daniel, and Aina Espinosa. 2007. Phonetic typology and positional allophones for alveolar rhotics in Catalan. *Phonetica* 63: 1–28. [CrossRef] [PubMed]
Recasens, Daniel, and Clara Rodríguez. 2017. Lingual articulation and coarticulation for Catalan consonants and vowels: An ultrasound study. *Phonetica* 74: 125–56. [CrossRef]
Redford, Melissa. 2007. Word-internal versus word-peripheral consonantal duration patterns in three languages. *Journal of the Acoustical Society of America* 121: 1665–78. [CrossRef] [PubMed]
Roseano, Paolo, Patrícia Marsà Morales, and Laura Alañá Vilas. 2022. Is Catalan syllable timed? An answer based on cluster analysis. In *Propuestas en fonética experimental: Enfoques metodológicos y nuevas tecnologías*. Edited by Beatriz Blecua, Jordi Cicres, Marina Espejel and María Machuca. Girona: Universitat de Girona, pp. 270–74.
Simonet, Miquel. 2010. Dark and clear laterals in Catalan and Spanish: Interaction of phonetic categories in early bilinguals. *Journal of Phonetics* 38: 663–78. [CrossRef]
Stevens, Kenneth N. 1999. *Acoustic Phonetics*. Boston: The MIT Press.
Stone, Maureen. 2005. A guide to analysing tongue motion from ultrasound images. *Clinical Linguistics and Phonetics* 19: 455–501. [CrossRef] [PubMed]

Tang, Kevin, Ratree Wayland, Fenqi Wang, Sophia Vellozzi, Rahul Sengupta, and Lori Altmann. 2023. From sonority hierarchy to posterior probability as a measure: The case of Spanish stops. *Journal of the Acoustical Society of America* 153: 1191–203. [CrossRef]

Westbury, John R. 1983. Enlargement of the supraglottal cavity and its relation to stop consonant voicing. *Journal of the Acoustical Society of America* 73: 1322–36. [CrossRef]

*languages*

MDPI

*Article*

# The Interplay between Syllabic Duration and Melody to Indicate Prosodic Functions in Brazilian Portuguese Story Retelling

**Plinio A. Barbosa * and Luís H. G. Alvarenga**

Department of Linguistics, Universidade Estadual de Campinas (UNICAMP), Campinas 13083-970, SP, Brazil; l255365@dac.unicamp.br
* Correspondence: pbarbosa@unicamp.br

**Abstract:** This paper investigates the relationship between syllabic duration and F0 contours for implementing three prosodic functions. Work on rhythm usually describes the evolution of syllable-sized durations throughout utterances, rarely making reference to melodic events. On the other hand, work on intonation usually describes linear sequences of melodic events with indirect references to duration. Although some scholars have explored the relationship between these two parameters for particular functions, to our knowledge, there has been no investigation on the systematic correlation between syllabic duration and F0 values throughout narrative sequences. Based on a corpus of story retelling with nine speakers of Brazilian Portuguese from two regions, our work investigated the interplay between syllabic duration and melody to signal three prosodic functions: terminal and non-terminal boundary marking and prominence. The examination of local syllabic duration maxima and four F0 descriptors revealed that these maxima act as landmarks for particular F0 shapes: for non-terminal boundaries, the great majority of shapes were increasing and increasing–decreasing patterns; for terminal boundaries, almost all shapes were decreasing F0 patterns; and for prominence marking, the great majority of shapes were high tones across the stressed syllable. Time series analyses revealed significant correlations between duration and specific F0 descriptors, pointing to a ruled interplay between F0 and syllabic duration patterns in Brazilian Portuguese story retelling.

**Keywords:** prosodic function; prosodic annotation; story retelling

## 1. Introduction

Intonation and speech rhythm research is practically exclusively associated with the respective studies of fundamental frequency (F0) patterns and syllabic duration throughout utterances, with a few exceptions reviewed below. Those who work with the description and modeling of a melodic contour focus almost exclusively on the description of the form and progression of F0 contours and levels in time, with indirect or secondary references to syllabic duration. This is the case of the American ToBI annotation system (Silverman et al. 1992), which is largely used in phonology-based intonation research in several languages to which the ToBI system was adapted (see, for instance, G-ToBI for German and SP-ToBI for Spanish). As for prosodic breaks, ToBI only marks the strength of a break in essentially two levels (3 or 4) to direct the analyses on matters of the sequence and organization of mono- and bitonal events linked to the realization of pitch accents and boundary tones. Some criticism appeared after ten years of use of the ToBI system. The main drawbacks concern the low inter-annotator reliability for the choice of pitch accent tones (Wightman 2002), as well as the possible circularity in not dissociating form from function (Hirst 2005). These two drawbacks are avoided in the work reported here by a procedure we used some time ago (Barbosa 2010) and which is similar to the rapid prosodic transcription used by Cole and collaborators (Cole et al. 2010; Cole and Shattuck-Hufnagel 2016). The method they proposed consists of asking listeners to evaluate how the speakers split their production into smaller parts by signaling breaks and how they highlighted words by listening to the corresponding audio files. To accomplish both tasks, the listeners were invited to mark the breaks with bars (/) and highlighted words with circles. The strength of these two functions (boundary and prominence) is considered as directly related to the proportion of the listeners' choices.

From the phonetic point of view, models of intonation have relied on the realization of focus and boundary marking functions from simple phonological descriptions of the sentence with codes which refer to F0 levels and contours. Examples of such models were proposed for languages such as Swedish (Bruce 1977, 1982; Gårding and Bruce 1981), Dutch (t'Hart 1984), English (t'Hart 1984; van Santen and Möbius 2000), and Japanese (Fujisaki and Hirose 1984), among others. Botinis et al. (2001) revised some of these models, pointing out that these phonetic models combine global trends and local changes in the F0 contours for implementing focus. In particular, the model by van Santen and Möbius (2000) combined the duration of accented and the following non-accented syllables to generate the F0 contours associated with pitch accents.

Recent work on Brazilian Portuguese (BP) analyzed the F0 contour and the role of the visual modality in the realization of both wide and narrow focus in declarative and interrogative sentences. The authors evaluated the roles of F0, syllabic duration, and intensity in the signaling of the focal function (Carnaval et al. 2022; Miranda et al. 2021, 2022). Also, in BP, Teixeira et al. (2018) studied the relevance of more than 100 acoustic prosodic descriptors for signaling terminal and non-terminal boundaries, with the aim of producing an algorithm for the automatic detection and classification of prosodic boundaries based on data of the C-Oral-Brazil corpus. Their work revealed the need to combine duration and F0 to achieve a better performance in predicting prosodic boundaries in spontaneous speech.

As for work on other languages, a comparative study of Mandarin and English investigated the joint manifestation of F0 contours and syllabic duration changes associated with tone, intonation, and duration in English (Xu 2009). A more recent work modeled the joint manifestation of F0 contours and syllabic duration for focus implementation in Emirati Arabic (Alzaidi et al. 2023). Work by Christodoulides (2018), which investigates the relative importance of prosodic–acoustic parameters to signal different kinds of boundaries in French spontaneous speech, points to the higher relevance of silent pause duration. As for prominence marking in English, the work by Herment-Dujardin and Hirst (2002) pointed out the relative relevance of F0 and duration parameters, of their combination as well as of semantic content. Their work also took into consideration a corpus of spontaneous speech. With controlled experiments, work on the prosodic cues for signaling prominence in German, both in adults (Holzgrefe et al. 2012) and in 8-month infants (Wellmann et al. 2012), revealed that the combination of pitch change and preboundary lengthening is a reliable cue for perceiving a boundary in speech.

Recently, for Greek, Arvaniti et al. (2024) used functional principal component analysis (FPCA) to evaluate the trade-off between F0 shape and duration in a corpus of 13 speakers arranged in different pairs to read dialogs. The first two components of the FPCA revealed that "F0 curves with a less pronounced dip and an earlier and lower peak were associated with longer accented vowel duration" and that "lower F0 curves, particularly those low in the preaccentual region, were associated with longer duration of that region." (pp. 292 and 293). None of these works, however, analyzed the systematic correspondence of F0 descriptors extracted from the F0 trace and syllabic duration for realizing focus and for signaling prosodic boundary throughout long stretches of unscripted speech.

In the same direction, it is important to highlight that long-standing research on speech rhythm rarely makes reference to F0 contour events. This line of research has been pointing to the primacy of syllable duration for marking prosodic boundaries and prominence in different languages (see Leemann et al. 2016 in eight languages/varieties; Streefkerk 1997 in Dutch; Gussenhoven and Rietveld 1992 in English; and Barbosa 2007 in BP).

Recent work on the notion of macro-rhythm considered the timing of melodic events such as pitch accents throughout utterances but did not take into account the analysis of the duration of entire syllabic sequences regardless of their accentedness (Jun 2014; Wehrle et al. 2020).

It is the interplay between F0 and syllabic duration patterns that we propose to investigate here by examining the cross-correlation between the corresponding time series in places where terminal and non-terminal boundaries and prominence are realized in storytelling. Speech productions in two dialects (São Paulo and Rio de Janeiro dialects) are analyzed. These two dialects were chosen because they have been the object of the majority of BP intonation studies.

Time series correlations capture the systematic correspondence between two variables. If both series are identical, the correlation is 1, whereas if they are completely unrelated on statistical grounds, the correlation is 0. In between, the technique is able to capture systematic positions where both series have local maxima or minima, which is the aspect we want to investigate here. To the best of our knowledge, the previous literature did not explore systematic correlations of these paired series. This also applies for work on BP, for which very limited corpora can be found, usually investigating the realization of local functions in read utterances.

As for work on narratives and storytelling, no systematic exploration of the relationship between melodic and duration levels has been carried out. The work by Oliveira (2012) investigated narratives in BP on the role of prosodic parameters such as F0, speech rate, and pause for the segmentation of narrative sections. He pointed to the cyclical character of speech rate which accompanies the change across these sections.

Other studies on storytelling proposed prosodic modification rules from neutral speech to endow speech synthesis systems with this speaking style. They have been developed for languages such as Malay (Ramli et al. 2016), Hindi (Verma et al. 2015), and French (Doukhan et al. 2011). The first one points to the prosodic differences between storytelling and the neutral speeches of two professional storytellers, while the second one evaluates the differences regarding distinct emotions conveyed during the narratives of five laypeople telling stories to children. The work on French, on the other hand, uses a single professional storyteller and assesses differences in melodic, durational, and intensive prosodic parameters between other speaking styles, as well as across narrative sections. None of these studies explore the retelling of a story, which has a strong component of cognitive load (see Dixon and Gould 1996 for the difference between storytelling and story retelling; see Pratt et al. 1989; Skehan and Foster 1999 for the study of processing load in story retelling). The number of speakers is also lower than the number of participants in our study.

In the present study, we fill these two gaps by describing and quantifying the convergence between F0 descriptors and syllabic duration in story retelling with a corpus that includes more speakers than usually found for this kind of study as well as analyzing running speech, and not isolated utterances, as is the case in the great majority of studies investigating BP. Recently, we investigated the interplay between F0 descriptors and normalized duration maxima in three speakers from São Paulo state, Brazil, for both reading and story retelling recorded in 2009 (Barbosa 2024). The results of cross-correlations between the positions of prominence and each series of four F0 descriptors (F0 median, F0 range, F0 rise and fall rates) revealed that the two speaking styles differ in the sense that the F0 range and F0 rise/fall rate are more correlated with peaks of duration in story retelling than F0 median is in reading. In this style, F0 median and F0 rise rate peaks are mostly aligned with duration maxima, which is also due to rising contours being the most frequent shape for realizing this function in reading. A rising contour for marking prominence is the most frequent contour in story retelling, but it is preceded by F0 fall rate minima and with a large F0 range. This is related to the fact that an F0 fall with higher rates precedes the typical rising of the prominent F0 contour either inside the longer V-to-V interval or inside the immediately preceding unit, which signals that the alignment of an F0 fall is more relevant for preparing a variable following an F0 rise.

As for the cross-correlation between the non-terminal boundary positions with the F0 descriptor series, with the exception of one speaker, the F0 range and F0 rise rate maxima correlate more with duration peaks, with higher values for story retelling, which can be taken as a characteristic of this style: the peaks of risings and extended ranges may indicate in BP that the speaker has more to say. Finally, the same work pointed to the correlation between terminal boundary positions with the F0 descriptor series, showing that F0 median minima and F0 range maxima are the descriptors with higher values of correlation in reading. This can be seen as a characteristic of this style when BP speakers signal terminal boundaries.

Based on previous work, our hypotheses concerning the correlation of one or more F0 descriptor with duration maxima are as follows: (1) F0 range and F0 rise/fall rate maxima are the most relevant correlated descriptors when signaling prominence; (2) the F0 range and F0 rise rate maxima are the most relevant descriptors at non-terminal boundaries; and (3) F0 median minima and F0 range maxima are the most relevant descriptors at terminal boundaries, which are less frequent in the case of story retelling due to the nature of the task.

## 2. Methodology

### 2.1. Corpus

The BELÉM corpus is formed by readings and retellings of the story about the origin of the Portuguese Belém pastries by Brazilian and Portuguese male and female speakers with a range of 30 to 45 years of age from the state of São Paulo. For the building of this corpus, the first recording was made in 2009. In 2022, the recordings were resumed to add to the BELÉM corpus Brazilian speakers from different dialectal regions of Brazil, starting with speakers from Rio de Janeiro, as well as extending the number of speakers from São Paulo. Furthermore, a shortened version of the original text (753 words instead of 1568), which can be found in Appendix A, was used as the text for the new retellings. These new recordings include six subjects from São Paulo (3 females and 3 males) and three subjects from Rio de Janeiro (2 females and 1 male). As in 2009, this new recruitment was based in the traditional friend-of-a-friend approach in sociolinguistics started by the friends of the second author in Brazil. Their narratives had between 220 and 330 words (2 to 4 min). The additional participants were between 20 and 30 years old at the time of recording and were college students of different majors. These recordings are available in the Figshare platform [https://doi.org/10.6084/m9.figshare.25383190.v1 (accessed on 5 July 2024)] Only data from the new recordings, based on the shortened version of the Belém pastries story, are used for the analyses presented here.

Due to restrictions related to the COVID-19 pandemic, the participants themselves used the Easy Voice app on their own cell phones to make all recordings. The nine subjects read the text and soon after retold the story in their own words. Only the retellings are considered for analysis here. Because this app allows choosing among different codifications, instructions were given to record all audio files in PCM format (WAV) at a sampling rate of 48 kHz. The first author, who is a trained phonetician, further evaluated all audio files for the presence of noise that would impact computing the acoustic parameters. No recordings were discarded. The recordings were then resampled at 16 kHz and leveled to the same maximum intensity level at 65 dB.

### 2.2. Acoustic–Prosodic Parameters for Analysis

A Praat script, Prosody Descriptor Extractor (Barbosa 2020), henceforth PDE, was used to extract statistical descriptors of prosodic–acoustic parameters based on syllabic duration and fundamental frequency (F0). The script is accompanied by a manual, and examples of input and output data are also given.

A syllabic duration series was obtained by normalizing and smoothing the duration of the sequence of V-to-V intervals, that is, of all the segments delimited by two immediately consecutive vowel onsets.

The reason for using V-to-V intervals instead of phonological syllables is threefold: (1) Dogil and Braun (1988) conducted psycholinguistic experiments that showed that vowel onset tracking is a fundamental property of speech signal processing in our brain. (2) This property of the brain activity was also pointed out by Chistovich and Ogorodnikova (1982) by examining post-stimulus temporal neuronal responses to speech. They reported amplified neuronal responses to portions of energy increase typical of C-V transitions, accompanied by response suppression in regions where energy decreased (typically around V-C transitions). (3) Furthermore, a segmentation based on vowel onsets has the advantage of being detectable under moderately noisy conditions (Barbosa 2010).

The perceptual impression of speech rate is also primarily associated with the tracking of vowel onsets and not syllable onsets, as was experimentally tested by Pompino-Marschall (1991) with German subjects. This reveals that the perception of the syllable sequence relies on the detection of the nucleus of the syllables, more often occupied by vowels. Because V-to-V intervals are syllable-sized units, which mark the flow of syllables throughout utterances, we refer to the V-to-V interval duration as the "syllabic duration".

Figure 1 illustrates, in the second tier, the segmentation and labeling of the V-to-V intervals for the excerpt "e aí no fim ele dormiu. . ." (and then he eventually slept) of a story retelling carried out by a female speaker from Rio de Janeiro. The content of the other tiers is explained later on in this text.

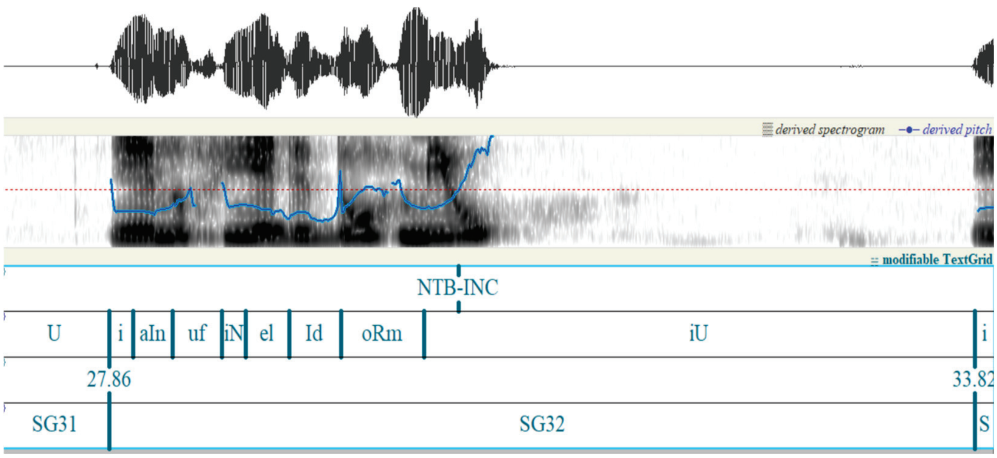



**Figure 1.** Waveform (**top**), F0 trace (blue) superimposed to a broadband spectrogram (**middle**) and annotation tiers of the excerpt "e aí no fim ele dormiu. . ." (and then he eventually slept) of a story retelling from a female speaker from Rio de Janeiro. From top to bottom, tier 1 codes the prosodic function and F0 shape, tier 2 the V-to-V intervals with their labels, and tier 3 the value of the normalized interval duration at the end of the stress group delimited in tier 4.

For normalizing the duration of the V-to-V intervals, the PDE script uses the z-score transformation given in Equation (1), where dur is the V-to-V duration in ms and the pair ($\mu_i$ and $var_i$) denotes the reference mean and variance in ms of the phones within the corresponding interval. These reference descriptors are found in the file TableOfReal included with the script.

$$z = \frac{dur - \sum_i \mu_i}{\sqrt{\sum_i var_i}} \quad (1)$$

A smoothing technique is then used which consists of serially applying a smoothing technique carried by a 5-point moving average filter given by Equation (2) to the sequence of z-scores obtained from the previous stage.

$$z^i_{smoothed} = \frac{5.z^i + 3.z^{i-1} + 3.z^{i+1} + 1.z^{i-2} + 1.z^{i+2}}{13} \quad (2)$$

This technique minimizes the effects of intrinsic duration and number of segments in the V-to-V unit, as well as attenuates the minor effects of duration variation related to the realization of lexical stress in the speech chain. Local peaks of smoothed z-scores are then detected by tracking the position for which their discrete first derivative changes from a positive to a negative value.

Previous research demonstrated good correspondence between these local peaks of syllabic duration and the perception of both prominent and pre-boundary syllables with correlations between 69 and 82% for reading (Barbosa 2008), provided that silent pauses, when applicable, were included in the corresponding V-to-V interval, as can be seen in Figure 1 for "iU". A local peak of smoothed z-scores is considered here an index of prosodic strength and the duration of the silent pause, when present, is an integral part of the signaling of this strength. It is related to the fact that the longer a silent pause is, the stronger the perception of a boundary (Sanderman and Collier 1995). The same applies when a local peak of smoothed z-scores signals the prominence of a syllable in a word. This is why these maxima were taken as indicators of the right edges of stress groups in BP, a language for which syllabic duration is the main parameter for signaling lexical and phrase stress (Fernandes 1976; Massini 1991; Barbosa 1996). When a silent pause is part of the V-to-V interval and it has a duration peak, this does not mean that it is part of the stress group but only that the stretch of sound at the left of this pause ends the stress group. For the purpose of this study, the interval between two immediately consecutive smoothed z-score maxima is called the "stress group". The stress group delimited in such a way is taken as a prosodic constituent that ends in a prominent unit (often an informational focus) or a prosodic boundary, either terminal or non-terminal.

The PDE script also delivers 12 F0 descriptors for the intervals of the tier specified by the user, here, the V-to-V intervals tier. From these 12 parameters, we selected four F0 descriptors: F0 median, F0 range (F0 maximum minus F0 minimum), and F0 rise and F0 fall mean rates. The latter two descriptors were computed as the first derivatives of smoothed and interpolated F0 contour. Smoothing and interpolation employed embedded Praat functions with 5 Hz as the cut frequency for smoothing and a quadratic function for interpolation, avoiding small oscillations and gaps of the F0 contour before computing the derivative. These four F0 descriptors in particular were chosen due to their relevance for signaling prominence and boundary in the literature (see Mittmann and Barbosa 2016 for a review). This relevance is illustrated in Figure 1, where the values of the V-to-V smoothed z-scores at the right ends of stress groups 31 and 32 (27.86 and 33.82) correspond to V-to-V units followed by long pauses. Furthermore, by the end of stress group 32, an F0 rise (INC) signals a non-terminal boundary (NTB). In those cases, the F0 range and F0 rise rate are better descriptors of F0 shape in signaling this type of boundary.

It is exactly the relation between the four F0 descriptors and duration maximum positions that realize each one of three prosodic functions, which are further discussed in the next section. We illustrate this relation in Figure 2 before presenting the prosodic functions in the next section. By examining the very first terminal boundary position at the extreme left of Figure 2, one can observe a syllabic duration peak coinciding with a local minimum of the F0 contour. The "meeting" of these two landmarks, a local maximum of syllabic duration and a local minimum of F0, unequivocally marks a terminal boundary of declarative utterances for a Brazilian listener (Moraes 1998). What is more, the fact that the relevance of both F0 minima and lower rates of F0 decrease, as we will see later on, allows the listener to distinguish the terminal boundary from the non-terminal boundary, such as the ones shown in the antepenultimate and penultimate positions in this figure. In the penultimate position, the non-terminal boundary is marked by both a local duration peak and an F0 peak, soon followed by an F0 fall. The two dashed positions, on the other hand, indicate prominences by F0 rising (LH) and falling (HL) contours, respectively. In both positions, the duration peaks are less salient in contrast to the salient local peaks of the boundary positions, as revealed by their heights.

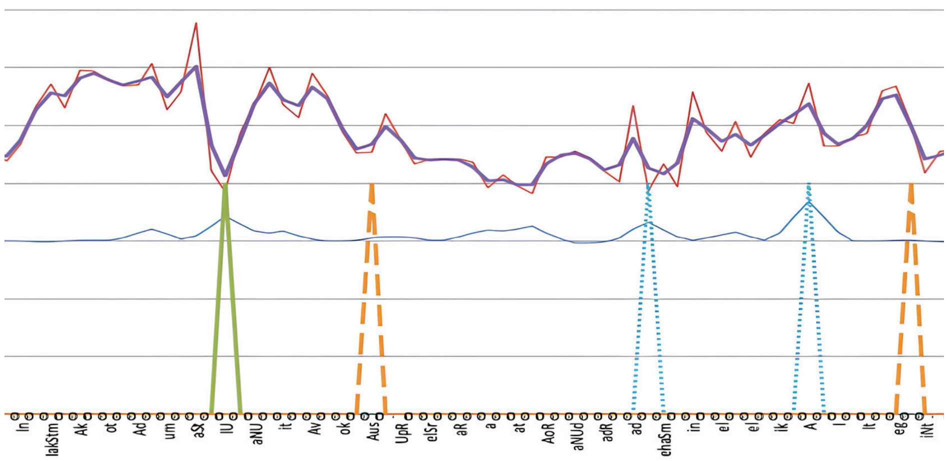



**Figure 2.** Original (above, red line) and smoothed F0 contour (above, superimposed, dark blue being the smoothed one), syllabic duration contour (light blue, middle) and positions of V-to-V duration maxima where the following functions are realized: terminal (solid line), non-terminal (dotted line), and prominence (dashed line).

Analyses of the parameters investigated here were carried out for two regions (São Paulo and Rio de Janeiro) and two genders (males and females). The reason to split between the two genders and two regions is to suggest future work on the contribution of these factors, as well as to be able to engage with the previous literature, as presented in the Discussion section.

### *2.3. Prosodic Function Annotation*

For each stress group automatically generated by the PDE script, the two authors independently annotated, based on their respective perception, which one of the three functions studied here, i.e., the terminal boundary (TLB), non-terminal boundary (NTB), or prominence (PRO), was realized at its right edge. After that stage, both authors discussed cases of initial disagreement until they achieved consensus. This was carried out for less than 1% of the labels given. Previous research revealed that even lay listeners perceive these functions with quite a fair to high degree of inter-rater reliability (Raso and Mello 2012; Cole and Shattuck-Hufnagel 2016).

The detection of the positions where one of the three functions investigated here was realized and signaled by a peak of normalized/smoothed syllabic duration had performance compatible with previous tests (e.g., success rates between 69 and 82% for reading in Barbosa 2008). As regards missing errors, 17% of the prosodic events were not detected. In addition, the procedure produced circa 6% of false alarms. A false alarm (FA) stands for a duration peak that does not correspond to any of the functions studied here. Thus, the procedure has a success rate of 77%, which means that the use of syllable-size duration for detecting both prominence and boundary is quite successful. The relative frequency of hesitations was 7% for São Paulo and 5% for Rio de Janeiro speakers. Gender-wise, the hesitations were 10% for males and 11% for females.

Hesitations were not analyzed here. If one of these three functions was realized at places not at the right edge of the stress groups, the label NDT (not detected) before the label for the performed function was used. For instance, NDT_NTB signals that the specific non-terminal boundary was not realized at the right edge of the automatically delimited stress group. The total number of labels, excluding the FA labels, was 500 for the speakers of São Paulo and 272 for the speakers of Rio de Janeiro distributed in the following way, according to the region: 34 TLB, 296 NTB, and 170 PRO for São Paulo and 39 TLB, 153 NTB, and 80 PRO for Rio de Janeiro.

For each boundary function, both terminal and non-terminal, the shape of the melodic contour that ended the stress group was labeled as follows: increasing (INC), decreasing (DEC), increasing–decreasing (INCDEC), leveled (LEV), and decreasing with a slow F0 rate

(SLDEC). All decreasing contours with a falling rate between 2.5 and 15 Hz every 50 ms were classified as a case of slow decreasing (SLDEC); those with a falling rate below 2.5 Hz every 50 ms were classified as LEV and the others as DEC.

As for the prominence function, we marked the F0 shapes by conventional symbols from intonation research (monotonal and bitonal units): leveled high F0 peak (H), falling contour (HL), and rising contour (LH), where the right tone of the bitonal contours was aligned with the stressed syllable. This choice of annotation allows for a more direct comparison with the previous literature, which uses such a convention. The H, HL, and LH labels were chosen according to criteria found in Lucente (2012) referring to the F0 shape during the stressed vowel in this way: the LH label refers to a rising F0 preceded by an F0 fall; the H label refers to a less sharp increase just before and leveled during the stressed vowel and not preceded by an F0 fall; the HL label refers to an F0 fall during the stressed vowel preceded by an F0 rise.

### *2.4. Statistical Analyses*

To quantify the systematicity of the relationships between each F0 descriptor and the positions in which a F0 shape signals a function among the three studied here, we calculated the cross-correlation of these two time series for each function: (1) a series of values of 0 and 1 along the succession of V-to-V units in which 0 meant that the particular function was not realized in the current unit, and 1 if it was realized there, and (2) a series of each one of the four F0 descriptor values computed in each V-to-V unit in Hertz. This was conducted for each speaker. To do this, we used the *ccf()* function for cross-correlation between two time series available in the *tseries* package on the R statistical platform (R Project n.d., version 4.3.1). For a similar analysis of dialogs analyzing the correspondence between F0 and intensity contours, see Buder and Eriksson (1999).

These cross-correlations were computed around a window of five V-to-V units before and after the duration peak to investigate for which lag these landmarks have the higher correlation with a particular F0 contour descriptor. A lag of 0 (zero) means that the maximum duration position and the F0 descriptor series are compared directly, without moving one series in relation to the other, that is, the F0 descriptor maximum or minimum corresponds exactly with the duration maximum position. A lag of 1 (one) means that the maximum duration position series is compared with the F0 descriptor series moved one V-to-V unit to the right, allowing us to investigate whether the series of F0 descriptor minima or maxima one V-to-V unit to the left of the duration maximum is better correlated with the latter parameter. A lag of $-1$ (minus one) means that the maximum duration position series is compared with the F0 descriptor series moved one V-to-V unit to the left, allowing us, on the other hand, to investigate whether the series of F0 descriptor minima or maxima one V-to-V unit to the right of the duration maximum achieves higher correlations. The same reasoning applies for higher lag values. The significance level for the cross-correlations was set to 0.05.

Proportion tests were used for comparing the significance of proportions of functions between the two regions studied here and between speakers, both at the 0.05 significance level. For this purpose, the *prop.test()* function in R was used.

## 3. Results

The great majority of studies on BP intonation, both from phonological and phonetic perspectives, have investigated either the dialect of São Paulo (Madureira 1994, 2016; Madureira and Fontes 1997) or of Rio de Janeiro (Moraes 1998, 2008; Carnaval et al. 2022). For the sake of comparison, in the following, we refer to significant differences and commonalities found between the two dialects. Furthermore, new findings beyond the analysis of controlled speech with isolated utterances usually found in those previous studies were obtained in the present one.

Figure 3 shows the relative frequency of the prosodic functions under scrutiny here for the two regions. Non-terminal boundaries are the most common function realized in story retelling, with proportions higher than 70%. According to a proportion test, terminal boundaries are significantly more frequently used by participants from Rio de Janeiro (14%, ranging from 11 to 16%) than participants from São Paulo (7%, ranging from 3 to 12%). The amount of data for this comparison (a total of 39 TLB in São Paulo and 34 TLB in Rio) is higher than the usual minimum number of 30 data points suggested for paired comparisons including proportion tests (Dowdy and Wearden 1991). The differences in proportion between the two regions for the other two functions are not significantly different.

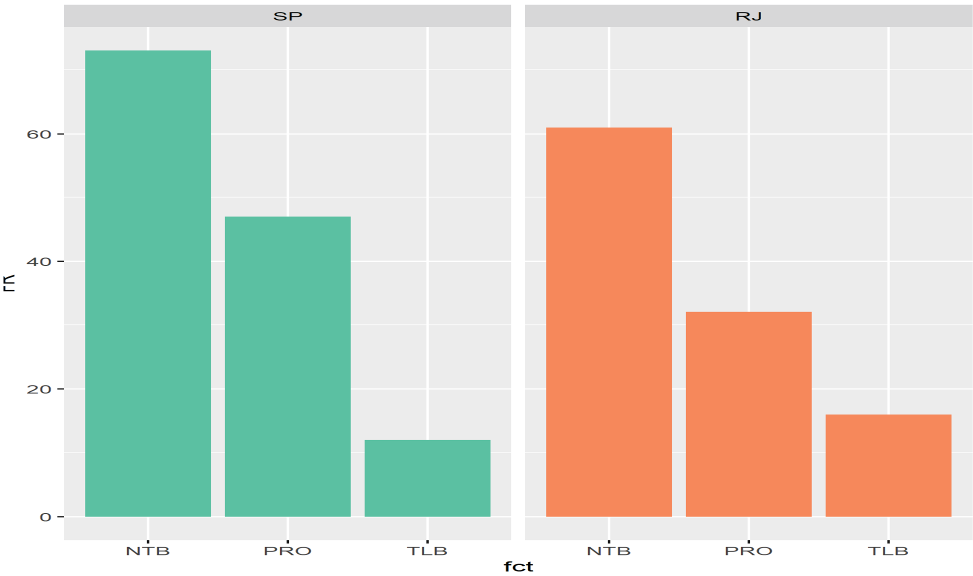



**Figure 3.** Relative frequencies of the three functions in story retelling according to dialect (**left**, SP = São Paulo, and **right**, RJ = Rio de Janeiro).

Considering the limited number of speakers, it is important to check if the results are relatively homogeneous across speakers or are an artifact of computing the average. This does not seem to be the case, as can be seen in Figure 4. There are, however, some exceptions. Figure 4 shows that male speaker AM from São Paulo has proportions that differ from the general tendency either for males or for São Paulo speakers: he uses the prominence function more than signals non-terminal boundaries. Furthermore, speakers MV (male) and SP (female) from São Paulo use a very low frequency of terminal boundaries.

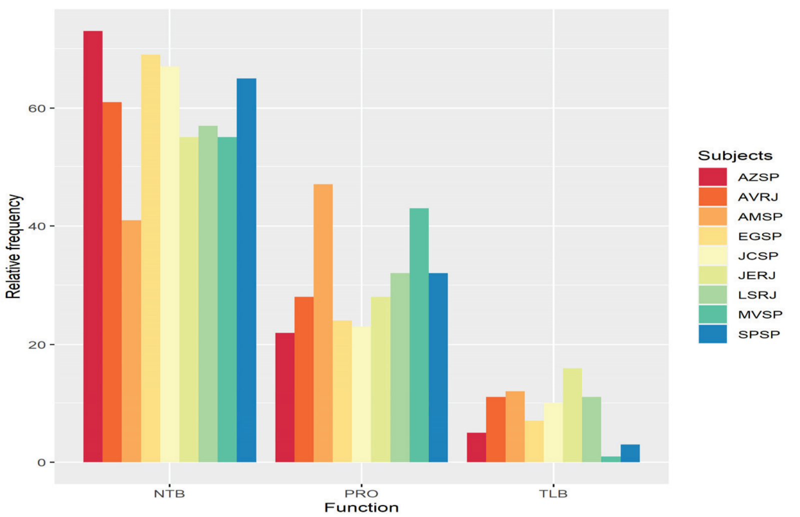



**Figure 4.** Proportion of the three functions in story retelling according to subject. The first two letters stand for the subject and the two last ones, the dialectal origin (SP or RJ).

Table 1 indicates the relative frequencies of the F0 shapes associated with each of the functions in story retelling according to dialect. The results are pooled for all speakers from each dialect. The two most frequent shapes for non-terminal boundaries are level and increasing, making up circa 86% in RJ and 80% in SP of all shapes, with a significant preference for increasing contours in São Paulo against level in Rio, as confirmed by a proportion test ($X^2 = 4.0$, $p = 0.04$ in São Paulo, $X^2 = 8.3$, $p = 0.004$ in Rio). For prominence, the most frequent shape is the high tone (H) followed by the LH contour in both dialects. Together, they account for 80% of all shapes in both RJ and SP. As for the terminal boundaries, the decreasing shape is by far the most relevant melodic form for realizing this function.

**Table 1.** Proportions (%) of the melodic forms associated with each of the functions in story retelling according to dialect. Significantly different proportions for the same function and shape in each dialect is indicated with an asterisk (*).

| Dialect | Function | Shape | Shape Proportion per Function |
|---|---|---|---|
| RJ | NTB | LEV | 51.91 * |
| | | INC | 35.24 * |
| | | SLDEC | 7.04 |
| | | INCDEC | 5.76 |
| | PRO | H | 58.75 |
| | | LH | 21.25 |
| | | HL | 13.75 |
| | | HLH | 6.25 |
| | TLB | DEC | 74.99 |
| | | INCDEC | 19.43 |
| | | SLDEC | 5.54 |
| SP | NTB | INC | 44.9 * |
| | | LEV | 36.46 * |
| | | INCDEC | 11.13 |
| | | SLDEC | 7.4 |
| | PRO | H | 43.5 |
| | | LH | 24.08 |
| | | HL | 18.79 |
| | | HLH | 13.5 |
| | TLB | DEC | 85.27 |
| | | INCDEC | 8.82 |
| | | SLDEC | 5.88 |

The only two significant differences across gender were that (1) male speakers use the slow decreasing shape more than female speakers, a property already found in a previous study in BP reading (Barbosa and Mareüil 2016), and that (2) there is a significant preference for the high tone in female speakers in the prominence position (circa 61% against circa 42% in males).

With regard to the cross-correlations between the time series whose contours were illustrated in Figure 2, only significant results are shown in Table 2, both by dialect and by gender. Only correlations within the window of lags −2 and 2 are taken into account due to the extension of the phonological words which generally include one to two prestressed and one to two post-stressed syllables. As can be seen in Table 2, higher correlations were found for lag 0, with only a single exception (TLB for females for the F0 rise, lag −2). Only the two highest significant correlations are shown because the others are lower than 5%, not taken into consideration here.

**Table 2.** The two highest significant correlations between position of syllabic duration peak and F0 descriptor value according to function, gender and dialect. When not mentioned otherwise, the correlations shown here are for lag 0. The abbreviation "ns" stands for non-significant for any of the four F0 descriptors.

| PRO | NTB | TLB |
|---|---|---|
| SP | | |
| F0range (0.06) | F0range (0.36) | F0fall (0.07) |
| F0rise (0.06) | F0rise (0.17) | F0range (0.04) |
| RJ | | |
| F0fall (0.12) | F0range (0.15) | ns |
| F0range (0.09) | F0rise (0.06) | |
| Male | | |
| F0range (0.06) | F0range (0.27) | ns |
| F0fall (0.06) | F0rise (0.17) | |
| Female | | |
| F0fall (0.14) | F0range (0.32) | F0range (0.06) |
| F0rise (0.12) | F0rise (0.12) | F0rise (0.05, lag −2) |

F0 range and F0 rise are the most relevant descriptors for non-terminal boundaries when correlations with duration maxima are concerned. With the exception of speakers from Rio de Janeiro, correlations between the F0 range and duration maxima for São Paulo and for both genders are between circa than 30% and 40%. This means that there is a tendency for the F0 range to be higher just before non-terminal boundaries, where syllabic duration maxima are positioned.

The F0 range is a relevant descriptor for the correlation of F0 and duration for the prominence function for males but not females. Looking at the data according to regions, the F0 range is the most relevant descriptor as far as correlations with duration are concerned in both regions. The higher correlation of F0 fall with duration in Rio and in both genders suggest that an F0 fall in the syllabic unit where there is a maximum of duration is a relevant cue for marking prominence in BP.

It is likely that the non-significant correlations for terminal boundaries for the less represented dialect (RJ) and for male speakers is due to the small amount of data for these cases. For São Paulo, however, the correlations of F0 range and F0 fall with duration maxima are the most relevant descriptors. This is compatible with the decreasing shape. Female speakers, on the other hand, seem to privilege the convergence of duration maxima with an F0 rise two syllabic units before the position of the local maximum, which suggests their falls are much more variable and converge less with duration maxima.

Figure 5 shows individual values of cross-correlations between the position of syllabic duration peak and the four F0 descriptor values for lags −5 to 5. The importance of the splitting of the data according to speaker is to show that the correlations shown in Table 2 are not a result of averaging but an effect found in the great majority of the speakers. The results shown here are split according to lag, function and subject. If we consider lags between −2 and 2, which would imply examining up to two syllables around the stressed syllable, where usually duration maxima are found, a lower spread of the correlation values across the individuals for a particular combination of function and F0 descriptor signals a higher consistency of the interplay between syllabic duration and F0 for that particular function. Following this rationale, for non-terminal boundaries, F0 range and F0 rise are, in that order, the most relevant descriptors, which confirms the pooled values of Table 2.

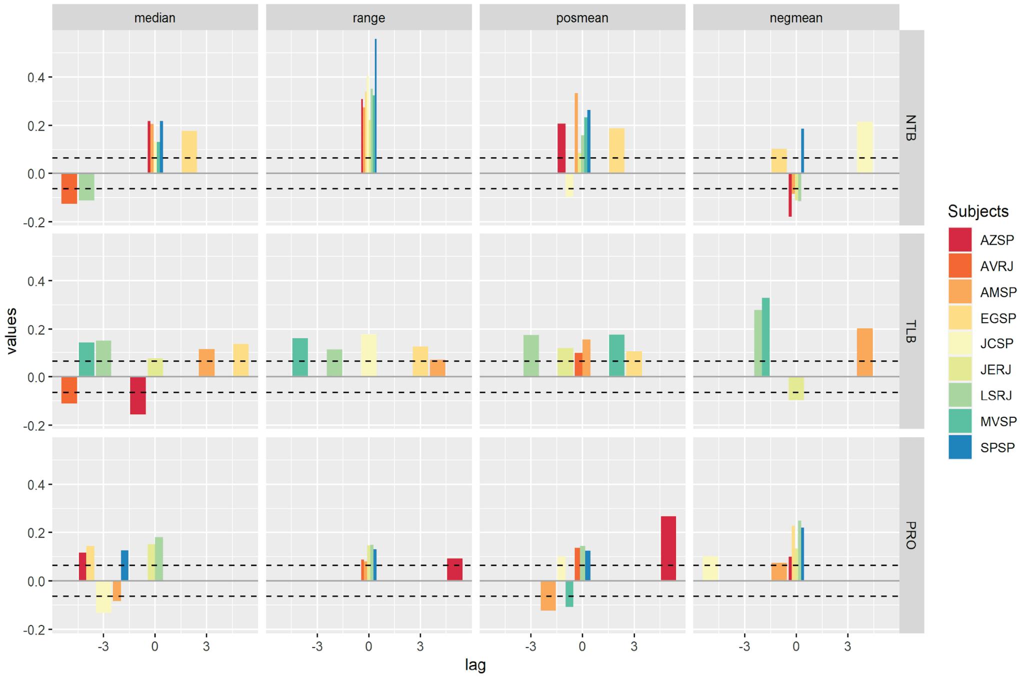



**Figure 5.** Cross-correlations between position of syllabic duration peak and each one of the four F0 descriptor values (columns) according to function (rows, labels in the right), lag (*x*-axis) and subject (color). Non-significant correlations are not presented. Median stands for F0 median, range stands for F0 range, posmean stands for mean F0 rise rate and negmean stands for mean F0 fall rate.

For prominence, the bars are more concentrated in the windows [−2, 2] for the F0 range, F0 fall, and F0 rise in that order. For terminal boundaries, the situation is more complex, presenting much more inter-individual variation with a higher concentration in the same lag window for F0 falls, as expected, with cross-correlations around 30% for the subjects with the two shades of green (LSRJ and MVSP). This value is similar to the correlations for non-terminal boundaries with the F0 range as a descriptor.

An important aspect of F0 dynamics concerns the rates of rises and falls, as well as the amount of the F0 range where significant correlations apply. For São Paulo speakers, terminal boundaries are signaled by decreasing F0 contours, with a median rate of 2.4 Hz/50 ms against 3.5 Hz/50 ms in other positions (corresponding to the points in the data series which includes positions for the two other functions and a position where none of the three functions studied here are realized), that is, F0 falls are slower when realizing terminal boundaries (see Barbosa 2024 for similar results for the same dialect).

As for F0 rises before non-terminal boundaries, the rate values for São Paulo speakers vary from 8.6 Hz/50 ms in the window where non-terminal boundaries are realized against 5.0 Hz/50 ms elsewhere, and for Rio speakers, they vary from 6.2 Hz/50 ms in the window where non-terminal boundaries are realized against 5.0 Hz/50 ms elsewhere. This means that F0 rises are faster when realizing non-terminal boundaries in both varieties but with a larger difference in São Paulo (see Barbosa 2024 for similar results for another corpus of São Paulo speakers).

A similar pattern applies for prominences for São Paulo speakers in terms of F0 rises: 7.1 Hz/50 ms in the window where prominences are realized against 5.4 Hz/50 ms elsewhere; for Rio speakers, the figures are 7.2 Hz/50 ms in the window where prominences are realized against 5 Hz/50 ms elsewhere, a very similar result in comparison with São Paulo. As for the F0 range, the patterns are as follows: 28.8 Hz in the window where prominences are realized against 20.0 Hz elsewhere for São Paulo; for Rio, the values are 31.6 Hz in the window where prominences are realized against 17.2 Hz elsewhere. As it can be seen, the figures for rates are quite close in both varieties.

In the prominent position, female speakers have F0 falls of 8.4 Hz/50 ms in the window where prominences are realized against 4.1 Hz/50 ms elsewhere and 3.8 Hz/50 ms in the window where prominences are realized against 2.7 Hz/50 ms elsewhere for male speakers.

## 4. Discussion

With respect to the previous literature on the prosody of BP, the results of this study add to the current knowledge on the matter, first by analyzing a higher number of speakers and by exploring spontaneous speech in the story retelling style. In fact, with just a few exceptions, such as the study of focus by Carnaval et al. (2022) with four speakers from Rio de Janeiro, the majority of past and recent studies on BP relies on results obtained from isolated utterances (see, for instance, Moraes 1998, 2008; Madureira 1994; Miranda et al. 2021, 2022). Moreover, none of the studies on spontaneous speech investigated the frequency of F0 shapes and the relation of F0 descriptors to syllabic duration maxima in a systematic way. The work by Lucente (2012) with speakers from São Paulo, for instance, occasionally pointed to some aspects of the shapes of F0 at boundary position, while the work by Teixeira et al. (2018), with speakers from Minas Gerais, studied the relative importance of duration and F0 descriptors for the signaling of boundaries but not the amount of their correlation. Although in their material, instances of narrative excerpts can be found, their analysis did not consider these instances as an independent factor of investigation.

Our results point to the fact that in story retelling, the most realized function is non-terminal boundary marking, followed by the signaling of prominence. The preference for non-terminal boundaries is compatible with the need to chain stretches of speech to tell a story and to call attention to the most relevant remembered facts. Non-terminal boundaries and prominences represent almost 90% of the instances of the three functions studied here. Non-terminality is signaled by increasing and level F0 shapes where the first one has fast F0 rises which produces expanded F0 ranges partially synchronized with syllabic duration maxima in the unit just before the boundary. To the best of our knowledge, a new finding of this study is to report that faster F0 rises are associated with duration maxima in realizing non-terminality. The relevance of F0 rise peaks for non-terminality is a consequence of the use of increasing contours for realizing non-terminal boundaries.

An important aspect of the F0 dynamics for terminal boundaries which is significant, despite the limited number of data (about 30 data points per dialect), is the use of slower falls for signaling terminality in comparison to the rate of falls elsewhere, a finding not pointed out by previous studies on BP intonation.

As for prominence, the F0 range is expanded when realizing this function accompanied by earlier synchronization between the F0 fall and duration maxima within the syllabic unit where prominence is realized mainly by an H tone. This fall allows F0 to reach low values which are associated with a lengthened syllabic unit. This behavior is similar to the findings by Arvaniti et al. (2024) on the trade between duration and F0 around accented units, where F0 valleys accompany longer stressed syllables when realizing pitch accent.

A certain amount of interspeaker variability is part of prosodic studies on speaking style (see Perkell et al. 2002; Yoon 2014; Barbosa 2022, inter alia) and this is not different for story retelling, as shown in Figures 4 and 5, already commented upon here. A further investigation of prosodic differences across speakers, including the ones studied here, associated with the study of the effect of different retellings to the listeners could contribute to coaching on storytelling, making listening to stories a more pleasant experience to a target audience. Studies on the relation of poetry declamation and pleasantness (Wagner and Betz 2023 for German; Barbosa 2022 for Brazilian and European Portuguese) have results in this direction.

The rising shape (LH), the second most frequent shape for signaling prominence in the present study, is described by Moraes (2008) as a default realization of narrow focus for the dialect of Rio de Janeiro and later on by Lucente (2012) for the São Paulo dialect, the former for read speech and the latter for spontaneous speech. Nevertheless, both authors only described the LH shape as being the most common in the dialect they studied, without computing its frequency, as we showed in Table 1: circa 21% in Rio and 24% in São Paulo. In our study, it is the second most frequent and not the most frequent, as in Moraes' and Lucente' studies. As for the terminal boundaries, the decreasing shape is the most relevant melodic form for realizing this function. It is proposed as the prototypical realization at the right end of neutral declaratives in BP by Moraes (1998, 2008) in read speech. Based on the present study, terminal boundaries are realized mostly with the same deceasing shape in story retelling.

Some findings from the current study deserve further investigation. One of them is the significantly distinct proportions of instances of terminal boundaries between the two regions, with a higher proportion for Rio de Janeiro (14% against 7% in São Paulo, see Figure 3). Having additional data from the two regions could reveal, if this difference in proportion is confirmed, that speakers from Rio de Janeiro complete thematic excerpts of the story being told more often than speakers of São Paulo. Another finding is related to interspeaker differences specially referring to the frequency of terminal and non-terminal boundaries. This could benefit studies of the effect of different storytelling and story retelling on listeners in terms of different degrees of pleasantness depending on the uses of the types of boundaries (see Figure 4 for some differences across speakers). The differences across gender, like the finding that female speakers are faster in signaling prominences by a previous sharp fall also deserve a more extended investigation with more speakers and a balanced corpus in this respect. As the study by Barbosa and Mareüil (2016) showed, this could contribute to a perception of more musical prosody in female speakers in BP.

## 5. Conclusions

The results of this study derive from the use of a new methodology for investigating the relation between the two main prosodic parameters (syllable duration and F0 contours) for signaling prominence and boundaries. The results presented here stress the importance of investigating the correspondence between F0 contours and syllabic duration contours to further understanding how prosodic functions are realized.

The examination of local syllabic duration maxima and the four F0 descriptors revealed that these maxima act as landmarks for particular F0 shapes: for non-terminal boundaries, the great majority of shapes were increasing and increasing–decreasing patterns; for terminal boundaries, almost all shapes were decreasing F0 patterns; and for prominence marking, the vast majority of shapes were high tones across the stressed syllable.

Time series analyses revealed significant correlations between duration and specific F0 descriptors pointing to a ruled interplay between F0 and syllabic duration patterns in Brazilian Portuguese story retelling. The cross-correlation values obtained in our analysis of the data indicate that the right edge of stress groups in BP, primarily marked by peaks of normalized duration of syllable-size units (the V-to-V unit) is the characteristic place where duration and F0 landmarks meet.

What is more, expanded F0 ranges and faster or slower rates of F0 contours are significant aspects of the dynamics of boundary marking in BP story retelling, findings that could stimulate cross-linguistic work on the prosody of storytelling and story retelling.

**Author Contributions:** Methodology, P.A.B.; software, P.A.B.; validation, L.H.G.A.; data curation, L.H.G.A.; writing—original draft, P.A.B.; writing—review and editing, L.H.G.A.; supervision, P.A.B.; funding acquisition, P.A.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Council for Scientific and Technological Development, grant number 302194/2019-3.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki and approved by the Comissão de Ética em Pesquisa of the University of Campinas (protocol codes 51747521.2.0000.8142 on 12 October 2021 and 70135123.9.0000.814 on 5 August 2023).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The original contributions presented in the study are included in the article. Further inquiries can be directed to the corresponding author.

**Acknowledgments:** We thank all our speakers, and three anonymous reviewers. We are in debt with the editor of the special issue "Phonetics and Phonology of Ibero-Romance Languages", Rebeka Campos-Astorkiza, who commented and suggested changes that contributed to improving this paper.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Appendix A

Manuel desceu a escada, maldizendo a sua vida. Por mais que se esforçasse, estava sempre se metendo em problemas. Tristonho, foi ver o irmão Bernardo de Santa Maria. E este deu um sorriso, com uma ternura bondosa. Muito baixo, gordo, de careca brilhante e duas bochechas encarnadas como maçãs maduras, ele poderia ser uma figura rústica se não fossem os olhos cor de bronze, mais brilhantes do que estrelas. Olhos de um homem sereno, inteligente e feliz.

Bernardo, em vez de fazer perguntas, o mandou trabalhar ao lado do forno onde os irmãos faziam grandes pães redondos. Um calorzinho bom devolveu-lhe as forças. E colocou as mãos à obra sem saber o porquê sentia a cabeça girar e um vazio no estômago. Mas o motivo era bem simples. O pão quente e o leite morno exalavam um cheiro de dar água na boca. Não se atreveu, no entanto, a pedir nada. Afinal de contas, estava de castigo.

Para enganar a fome, enterrou as mãos na massa até ao cotovelo e fez uma grande bola. Depois olhou em volta de soslaio. Talvez não reparassem se ele comesse um bocado de massa crua. Ninguém reparou. Então, em vez de fazer os movimentos certos, começou a amassar ao acaso dobrando e redobrando aquela mistura de farinha e água que iria se transformar em pão. Uma ideia genial passou pela sua cabeça. Se fosse buscar outros mantimentos; Em vez de pão poderia fazer uma bola para rechear de carne. Ou um doce. Assim poderia ir comendo bocadinhos disto e daquilo sem que ninguém percebesse.

Foi dito e feito! Guloso como era, começou pela manteiga, depois foi ao leite, raspas de limão, ovos. . . O frei Bernardo o olhava discretamente sem dizer nada. Manuel tinha recuperado as cores e girava pela cozinha na maior correria.

Em cima da mesa, em vez de um pão redondo, tinha várias tigelinhas de massa cheias de creme. Meteu tudo no forno e foi se sentar. Estava cansadíssimo! Alguns minutos de descanso não seriam notados. Ele se encostou na parede, fechou os olhos e deixou que uma dormência suave fosse tomando conta do seu corpo e espírito. Ele estava tão bem ali! Flutuando dentro de si mesmo que não ouviu os sinos anunciarem que amanhecia. Nem foi à igreja para rezar a Prima. Escorregou para debaixo da mesa e dormia profundamente deitado no chão.

Quando, algumas horas depois, a voz do prior se fez ouvir na cozinha, Manuel quase desmaiou de susto. Não pensava em nenhuma desculpa possível para se justificar. Levantou-se passado de vergonha, com a roupa cheia de manchas, o cabelo bagunçado e os olhos inchados. E foi naquela triste figura que o prior entrou no refeitório. Ali estava o frei Diogo, os outros monges e irmãos, todos com ar muito solene. Com certeza iam expulsá-lo do convento. Abaixou a cabeça, esperando ouvir palavras terríveis, e só então reparou que, na mesa estava um tabuleiro cheio de pastéis que alguém polvilhou de canela.

– Este rapaz é um exemplo para todos nós – exclamou o prior, apontando para o tabuleiro com um gesto breve. – Ele fez uma bela ação, merece uma recompensa.

Manuel ainda olhou à procura do tal rapaz que seria recompensado, e ficou perplexo quando percebeu que era ele próprio, porque não se lembrava de ter feito nada excepcional.

– Ele passou a noite trabalhando sem descanso – continuou o prior no mesmo tom de aprovação.

– Com certeza não poupou esforços para inventar os melhores pastéis que já comi na minha vida. Podem até ser vendidos para fora, o que ajudará muito as finanças do mosteiro.

Até frei Diogo acenou que sim. E não é que ele parecia contente? O prior se virou com um sorriso aberto:

– A partir de hoje você será o nosso doceiro chefe! O pobre rapaz ficou sem saber se ria ou se chorava. Nunca lhe tinham feito um elogio! As palavras do prior encheram seu peito com uma alegria nova, desconhecida. Mas o pior é que não se lembrava da receita inventada por acaso na noite anterior. Era preciso confessar a verdade. E coragem?

Foi o irmão Bernardo quem o socorreu. Primeiro, fez sinal para que ficasse quieto. Depois, a sós, explicou que tinha reparado em todas as voltas que ele deu, em tudo o que usou e que juntos fariam uma nova dose nesse mesmo dia.

– Fui eu quem colocou a canela – confessou. – E ficaram tão deliciosos que só por minha conta já comi seis! Você será famoso, Manuel. Você e os seus pastéis de Belém!

## References


Alzaidi, Muhammad S. A., Yi Xu, Anqi Xu, and Marta Szreder. 2023. Analysis and computational modelling of Emirati Arabic intonation: A preliminary study. *Journal of Phonetics* 98: 101236. [CrossRef]

Arvaniti, Amalia, Argyro Katsika, and Na Hu. 2024. Variability, overlap, and cue trading in intonation. *Language* 100: 265–307. [CrossRef]

Barbosa, Plinio A. 1996. At least two macrorhythmic units are necessary for modeling Brazilian Portuguese duration. Paper presented at 1st ESCA Tutorial and Research Workshop on Speech Production Modeling: From Control Strategies to Acoustics & 4th Speech Production Seminar: Models and Data, Autrans, France, May 20–24; pp. 85–88.

Barbosa, Plinio A. 2007. From syntax to acoustic duration: A dynamical model of speech rhythm production. *Speech Communication* 49: 725–42. [CrossRef]

Barbosa, Plinio A. 2008. Prominence- and boundary-related acoustic correlations in Brazilian Portuguese read and spontaneous speech. Paper presented at Speech Prosody 2008 Conference, Campinas, Brazil, May 6–9; pp. 257–60.

Barbosa, Plinio A. 2010. Automatic Duration-Related Salience Detection in Brazilian Portuguese Read and Spontaneous Speech. Paper presented at Speech Prosody, Chicago, IL, USA, May 10–14; p. 100067.

Barbosa, Plinio A. 2020. Prosody Descriptor Extractor [Software Program]. Available online: https://github.com/pabarbosa/prosody-scripts/tree/master/ProsodyDescriptorExtractor (accessed on 5 July 2024).

Barbosa, Plinio A. 2022. Pleasantness and wellbeing in poem declamation in European and Brazilian Portuguese depends mostly on pausing and voice quality. *Frontiers in Communication* 7: 855177. [CrossRef]

Barbosa, Plinio A. 2024. The interplay between syllabic duration and melody to signal prosodic functions in reading and story retelling in Brazilian Portuguese. Paper presented at Speech Prosody 2024 Conference, Leiden, The Netherlands, July 2–5; pp. 1075–79. [CrossRef]

Barbosa, Plinio A., and Philippe B. Mareüil. 2016. Pics mélodiques prétoniques en portugais brésilien: Une étude quantitative. Paper presented at Actes de la Conférence Conjointe JEP-TALN-RECITAL 2016, Paris, France, July 4–8; pp. 527–53.

Botinis, Antonis, Bjorn Granström, and Bernd Möbius. 2001. Developments and paradigms in intonation research. *Speech Communication* 33: 263–96. [CrossRef]

Bruce, Gösta. 1977. *Swedish Word Accents in Sentence Perspective*. Lund: Gleerup.

Bruce, Gösta. 1982. *Developing the Swedish Intonation Model*. Working Papers. Lund: Department of Linguistics and Phonetics, Lund University, p. 22.

Buder, Eugene H., and Anders Eriksson. 1999. Time-series analysis of conversational prosody for the identification of rhythmic units. Paper presented at 14th International Congress of Phonetic Sciences, San Francisco, CA, USA, August 1–7, vol. 2, pp. 1071–74.

Carnaval, Manuela, João A. Moraes, and Albert Rilliard. 2022. Focus types in Brazilian Portuguese: Multimodal production and perception. *DELTA. Documentação de Estudos em Linguística Teórica e Aplicada* 38: 1–34. [CrossRef]

Chistovich, Ludmilla A., and Elena A. Ogorodnikova. 1982. Temporal processing of spectral data in vowel perception. *Speech Communication* 1: 45–54. [CrossRef]

Christodoulides, George. 2018. Acoustic Correlates of Prosodic Boundaries in French. A Review of Corpus Data. *Revista de Estudos da Linguagem* 26: 1531–49. [CrossRef]

Cole, Jennifer, and Stephanie Shattuck-Hufnagel. 2016. New methods for prosodic transcription: Capturing variability as a source of information. *Laboratory Phonology* 7: 8. [CrossRef]

Cole, Jennifer, Yoonsook Mo, and Soondo Baek. 2010. The role of syntactic structure in guiding prosody perception with ordinary listeners and everyday speech. *Language and Cognitive Processes* 25: 1141. [CrossRef]

Dixon, Roger A., and Odette N. Gould. 1996. Adults telling and retelling stories collaboratively. In *Interactive Minds: Life-Span Perspectives on the Social Foundation of Cognition*. Edited by Paul B. Baltes and Ursula M. Staudinger. Cambridge: Cambridge University Press, pp. 221–41.

Dogil, Grzegorz, and Gunter Braun. 1988. *The PIVOT Model of Speech Parsing*. Wien: Verlag.

Doukhan, David, Albert Rilliard, Sophie Rosset, Martine Adda-Decker, and Christophe d'Alessandro. 2011. Prosodic analysis of a corpus of tales. Paper presented at Twelfth Annual Conference of the International Speech Communication Association, Florence, Italy, August 27–31; pp. 3129–32.

Dowdy, Shirley, and Stanley Wearden. 1991. *Statistics for Research*, 2nd ed. New York: John Wiley & Sons.

Fernandes, Norma H. 1976. Contribuição para uma análise instrumental da acentuação e intonação do português. Master's thesis, State University of Sao Paulo, São Paulo, Brazil.

Fujisaki, Hiroya, and Keikichi Hirose. 1984. Analysis of voice fundamental frequency contours for declarative sentences of Japanese. *Journal of the Acoustical Society of Japan* 5: 233–42. [CrossRef]

Gårding, Eva, and Gösta Bruce. 1981. A presentation of the Lund model for-Swedish intonation. In *Nordic Prosody II*. Edited by T. Fretheim. Trondheim: Tapir, pp. 33–39.

Gussenhoven, Carlos, and Antonius C. Rietveld. 1992. Intonation contours, prosodic structure and pre boundary lengthening. *Journal of Phonetics* 20: 283–303. [CrossRef]

Herment-Dujardin, Sophie, and Daniel J. Hirst. 2002. Emphasis in English: A perceptual study based on manipulated synthetic speech. Paper presented at Speech Prosody 2002, Aix-en-Provence, France, April 11–13; pp. 379–82.

Hirst, Daniel J. 2005. Form and function in the representation of speech prosody. *Speech Communication* 46: 334–47. [CrossRef]

Holzgrefe, Julia, Caroline Schröder, Barbara Höhle, and Isabell Wartenburger. 2012. Processing of prosodic boundary cues as revealed by event-related brain potentials. Paper presented at the 13th Conference on Laboratory Phonology (LabPhon 13), Stuttgart, Germany, July 27–29.

Jun, Sun-Ah. 2014. Prosodic typology: By prominence type, word prosody, and macro-rhythm. In *Prosodic Typology II*. Oxford: Oxford University Press, pp. 520–39.

Leemann, Adrian, Marie-José Kolly, Yang Li, Ricky Chan, Geraldine Kwek, and Anna Jespersen. 2016. Towards a typology of prominence perception: The role of duration. Paper presented at 8th International Conference on Speech Prosody (SP2016), Boston, MA, USA, May 31–June 3; pp. 445–49.

Lucente, Luciana. 2012. Aspectos dinâmicos da fala e da entoação do português brasileiro. Ph.D. thesis, Universidade Estadual de Campinas, Campinas, Brazil.

Madureira, Sandra. 1994. Pitch Patterns in Brazilian Portuguese: An Acoustic-Phonetic Analysis. Paper presented at Fifth Australian International Conference on Speech Science and Technology, Perth, Australia, December 6–8; pp. 156–59.

Madureira, Sandra. 2016. Intonation and variation: The multiplicity of forms and senses. *Dialectologia: Revista Electrònica* Special Issue VI: 57–74.

Madureira, Sandra, and Mário A. Fontes. 1997. Fundamental contours in Brazilian Portuguese words. Paper presented at Intonation: Theories, Models and Applications, Athens, Greece, September 18–20; pp. 211–14.

Massini, Gladis. 1991. A duração no estudo do acento e do ritmo em português. Master's thesis, State University of Campinas, Campinas, Brazil.

Miranda, Luma S., João A. Moraes, and Albert Rilliard. 2022. Effects of F0 movements, intensity, and duration in the perceptual identification of Brazilian Portuguese wh-questions and wh-exclamations. *DELTA. Documentação de Estudos em Linguística Teórica e Aplicada* 38: 31–29. [CrossRef]

Miranda, Luma S., Marc Swerts, João A. Moraes, and Albert Rilliard. 2021. The Role of the Auditory and Visual Modalities in the Perceptual Identification of Brazilian Portuguese Statements and Echo Questions. *Language and Speech* 64: 3–23. [CrossRef]

Mittmann, Maryualê M., and Plinio A. Barbosa. 2016. An automatic speech segmentation tool based on multiple acoustic parameters. *CHIMERA. Romance Corpora and Linguistic Studies* 3: 133–44. [CrossRef]

Moraes, João A. 1998. Intonation in Brazilian Portuguese. In *Intonational Systems: A Survey of Twenty Languages*. Edited by Daniel Hirst and Albert Di Cristo. Cambridge: MIT Press.

Moraes, João A. 2008. The pitch accents in Brazilian Portuguese: Analysis by synthesis. Paper presented at Speech Prosody 2008, Campinas, Brazil, May 6–9; pp. 389–97.

Oliveira, Miguel, Jr. 2012. A study on speech rate as a prosodic feature in spontaneous narrative. *Alfa: Revista de Linguística* 56: 623–51. [CrossRef]

Perkell, Joseph S., Majid Zandipour, Melanie L. Matthies, and Harlan Lane. 2002. Economy of effort in different speaking conditions. I. A preliminary study of intersubject differences and modeling issues. *The Journal of the Acoustical Society of America* 112: 1627–41. [CrossRef]

Pompino-Marschall, Bernd. 1991. The syllable as a prosodic unit and the so-called P-centre effect. *Forschungsberichte Institut für Phonetik und Sprachliche Kommunikation der Universität München* 29: 66–124.

Pratt, Michael W., Cheryl Boyes, Susan Robins, and Judy Manchester. 1989. Telling tales: Aging, working memory, and the narrative cohesion of story retellings. *Developmental Psychology* 25: 628. [CrossRef]

R Project. n.d. The R Foundation for Statistical Computing. Available online: http://www.r-project.org/ (accessed on 5 July 2024).

Ramli, Izzad, Noirani Seman, Norizah Ardi, and Nursuriati Jamil. 2016. Rule-based storytelling text-to-speech (TTS) synthesis. *MATEC Web of Conferences, EDP Sciences* 77: 04003. [CrossRef]

Raso, Tommaso, and Heliana Mello. 2012. The C-ORAL-BRASIL I: Reference corpus for informal Brazilian Portuguese. In *Lecture Notes in Computer Science*. Berlin and Heidelberg: Springer, pp. 362–67.

Sanderman, Angelien A., and René Collier. 1995. Prosodic phrasing at the sentence level. In *Producing Speech: Contemporary Issues. For Katherine Safford Harris*. Melville: American Institute of Physic, pp. 321–32.

Silverman, Kim, Mary Beckman, John Pitrelli, Mari Ostendorf, Colin Wightman, Patti Price, Janet Pierrehumbert, and Julia Hirschberg. 1992. ToBI: A standard for labelling English prosody. Paper presented at International Congress of Spoken Language Processing, Banff, AB, Canada, October 13–16; pp. 867–70.

Skehan, Peter, and Pauline Foster. 1999. The influence of task structure and processing conditions on narrative retellings. *Language Learning* 49: 93–120. [CrossRef]

Streefkerk, Barbertje M. 1997. Acoustical correlates of prominence: A design for research. *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam* 21: 131–42.

t'Hart, Johan. 1984. A Phonetic Approach to Intonation: From Pitch Contours to Intonation Patterns. In *Intonation, Accent and Rhythm: Studies in Discourse Phonology*. Berlin and New York: De Gruyter, pp. 193–202.

Teixeira, Barbara, Plinio A. Barbosa, and Tommaso Raso. 2018. Automatic detection of prosodic boundaries in Brazilian Portuguese spontaneous speech. In *International Conference on Computational Processing of the Portuguese Language*. Cham: Springer International Publishing, pp. 429–37.

van Santen, Jan. P., and Bernd Möbius. 2000. A quantitative model of F0 generation and alignment. In *Intonation: Analysis, Modelling and Technology*. Dordrecht: Springer Netherlands, pp. 269–88.

Verma, Rashmi, Parakrant Sarkar, and Krothapalli S. Rao. 2015. Conversion of neutral speech to storytelling style speech. Paper presented at 2015 Eighth IEEE International Conference on Advances in Pattern Recognition, Kolkata, India, January 4–7; pp. 1–6.

Wagner, Petra, and Simon Betz. 2023. Effects of Meter, Genre and Experience on Pausing, Lengthening and Prosodic Phrasing in German Poetry Reading. Paper presented at Interspeech 2023, Dublin, Ireland, August 20–24; pp. 2538–42.

Wehrle, Simon, Francesco Cangemi, Harriet Hanekamp, Kai Vogeley, and Martine Grice. 2020. Assessing the intonation style of speakers with autism spectrum disorder. Paper presented at 10th International Conference on Speech Prosody, Tokyo, Japan, May 25–28; pp. 809–13.

Wellmann, Caroline, Julia Holzgrefe, Hubert Truckenbrodt, Isabell Wartenburger, and Barbara Höhle. 2012. How each prosodic boundary cue matters: Evidence from German infants. *Frontiers in Psychology* 3: 33536. [CrossRef] [PubMed]

Wightman, Colin. W. 2002. ToBI or not ToBI? Paper presented at Speech Prosody 2002, Aix-en-Provence, France, April 11–13; pp. 25–29.

Xu, Yi. 2009. Timing and coordination in tone and intonation: An articulatory-functional perspective. *Lingua* 19: 906–27. [CrossRef]

Yoon, Tae-Jin. 2014. Speaker variation in English prosodic boundary. *Linguistic Research* 31: 1–23.

languages

Article

# An Exploratory Study of Yes-No Question Intonation in Bilingual Labourdin French

**Nerea Delgado**

Modern Languages and Linguistics, Florida State University, Tallahassee, FL 32306, USA;
ndelgadofernandez@outlook.com

**Abstract:** Despite the growing interest in the study of intonation in bilingual regions in France, the case of Basque French remains under-investigated. Previous research on yes-no questions in standard French has shown that rising contours are the prototypical realization, while bilingual varieties in contact with Corsican or Occitan also seem to allow falling intonation to different extents. To investigate the case of Basque French, data from 11 Basque-French bilinguals from Labourd were considered. Participants completed a contextualized reading task and the Bilingual Language Profile questionnaire, which was used to examine their linguistic profile and language dominance. The results showed that rising intonation predominated (90%), presenting two main realizations: A low rise (L+H* H%) and a high rise ((¡)H* H%). Falling contours, in turn, appeared in only 10% of the data. In a preliminary consideration of these results by language dominance group, low rising contours were found to be more common among Basque-dominant participants, while falling ones appeared more often among French-dominant participants. While surprising, this result could, at least partly, stem from the participants' personal experiences with bilingualism and their contact with other varieties of Basque. We thus conclude that, for the most part, Basque Labourdin French resembles standard French intonation. The higher prevalence of falling intonation among French-dominant speakers, however, calls for future research.

**Keywords:** intonation; yes-no questions; bilingualism; Basque Labourdin French

## 1. Introduction

This article provides a first description of the intonation of yes-no questions in the variety of Southern French spoken in the Northern/French Basque Country or *Iparralde*[1]. Administratively speaking, this region forms part of la Nouvelle Aquitaine and comprises the three Basque historic provinces of Labourd (Lapurdi), Basse-Navarre (Nafarroa Beherea) and Soule (Zuberoa). The map in Figure 1 shows the locations of these provinces in southern France.

In Iparralde, French is in contact with Basque, a language isolate also spoken in the Southern/Spanish Basque Country. The global percentage of bilingual population in Iparralde amounts to 20.1% and that of passive bilinguals to 9.4%, although there exist differences across regions (Office Publique de la Langue Basque—Euskararen Erakunde Publikoa 2023)[2]. In Labourd, the area under study in this paper, these percentages are slightly reduced, to 15.7% and 8.6%, respectively (Office Publique de la Langue Basque—Euskararen Erakunde Publikoa 2023). While French is still considered the only official language of the country (Bochman 2018), the Communauté d'Agglomération Pays Basque recognized Basque (and Gascon) as a language of the community in 2018 (Communauté d'Agglomération Pays Basque—Euskal Hirigune Elkargoa 2018), which brings some hope for the endurance of the language in the territory[3].

This first description of Labourdin French intonation is expected to serve a dual purpose: documenting an understudied variety of Southern French and contributing to our understanding of language contact outcomes involving Basque in present-day France.

Both of these findings can have important implications for both contact and intonation models alike.

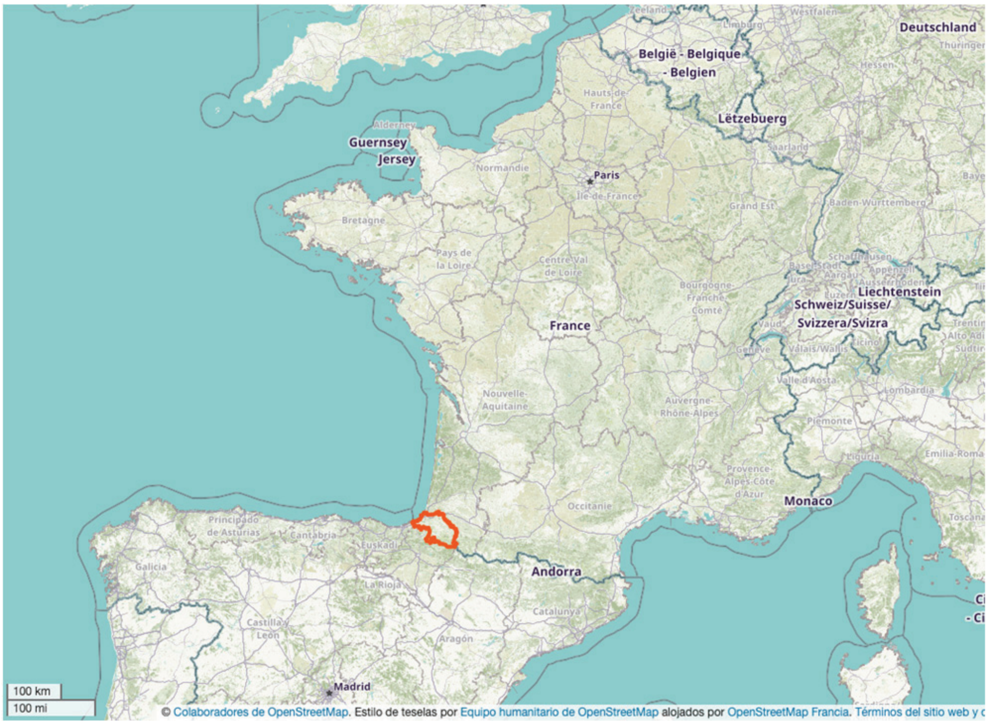



**Figure 1.** Map of Northern Basque Country or Iparralde (OpenStreetMap contributors n.d.).

## 2. Background

Despite the linguistic diversity found in Iparralde, the intonation patterns of French (and Basque) as spoken in the region remain poorly documented. To contextualize the study of intonation in the Northern Basque Country, we next consider previous intonation research on French, other bilingual regions inside and outside of France, as well as other bilingual communities outside the French-speaking world.

### *Previous Research on Intonation*

Intonation can serve different functions in the speech signal such as expressing feelings or emotions or reflecting syntactic structure. Importantly, it can also be used to mark question intonation. According to Dryer (2013), intonation is the second most common strategy to convey yes-no question meanings in the world's languages. Examples include English (Dehé and Braun 2019), German (Braun et al. 2019), Georgian (Vicenik and Jun 2014), and Persian (Sadat-Tehrani 2011), to name a few. The use of intonation to mark questions becomes crucial when both questions and declarative sentences share the same structure, which applies to French, as shown in (1).

| | | | | | |
|---|---|---|---|---|---|
| (1) | a. | Elle vient | à | la | fête. |
| | | she come-PRS.3SG | to | DET | party |
| | | 'She is coming to the party' | | | |
| | b. | Elle vient | à | la | fête ? |
| | | she come-PRS.3SG | to | DET | party |
| | | 'Is she coming to the party?' | | | |

The intonation distinction between declarative sentences/statements and yes-no questions is well-established in standard French. On the one hand, statements are typically realized with overall falling intonation (Delattre 1966; Delais-Roussarie et al. 2015; Di Cristo 1998, 2016; Jun and Fougeron 2000, 2002; Post 2000). Their general shape consists of a sequence of rises at the end of each stress group or what Di Cristo (1998) calls the "saw-tooth" pattern, and a progressive fall in tone until the end of the sentence. Yes-no questions, on the other hand, present a final rising intonation, particularly when there are no interrogative markers such as *est-ce que* or inversion (Delattre 1966; Di Cristo 1998, 2016; Fónagy and Bérard 1973; Post 2000). In a more detailed description, Di Cristo (1998, 2016) depicts yes-no questions with an initial peak followed by less prominent peaks or even a low plateau, and a final rise. Delais-Roussarie et al. (2015) annotate the contrast as L* L% for declaratives and H* H% for yes-no questions using the French ToBI (see end of section). Sichel-Bazin (2015) also proposes a low rising label (L+H* H%) for yes-no questions for a speaker from Lille.

Despite the consensus on the intonational realization of yes-no questions in standard French, it is far less clear what happens in non-standard varieties of the language. The last decade has seen growing interest in the study of intonation in bilingual regions in France and the results have shown that alternative intonations are available in the speakers' prosodic inventories. For example, Boula de Mareüil et al. (2012, 2016a) find that yes-no questions in Corsica (Corsican-French contact) are produced with falling intonation (H* L%). In areas of Occitan-French contact, falling contours have also been attested, although to different extents. Boula de Mareüil et al. (2016b) report that 5 out of their 14 participants produced falling intonation, while Delais-Roussarie et al. (2015), Sichel-Bazin et al. (2012), and Sichel-Bazin (2015) only observe a few sporadic instances.

Different intonation patterns have also been found for other sentence types in varieties of French spoken in territories outside of mainland France and that are in contact with other languages. For example, Martin (2012) documents steep rises that double in height what is attested in standard French in *continuation majeure* constructions in the varieties spoken in L'Île Maurice and La Réunion. While the source of said variation is unclear, Martin (2012) hypothesizes an influence from different creole languages spoken on these islands. Bullock (2009) also identifies an influence from American English in the expression of focus in the variety of French spoken by a group of third generation heritage speakers of French in Pennsylvania. In African varieties of French, Bordal and Lyche (2012) report the use of either stress-like or tonal-like features depending on whether these features are present in the speakers' L1 or dominant language, although neither stress nor tone were used contrastively in their French productions.

Beyond the French-speaking world, it is also not uncommon for bilingual varieties to present non-standard intonation patterns. Examples include differences in peak alignment in declarative sentences in Spanish in historical contact with Italian in Buenos Aires (Colantoni and Gurlekian 2004; Colantoni 2011), declarative and question intonation in Spanish in contact with Catalan (respectively, Simonet 2009, 2011; and Romera and Elordieta 2013), the speech of Polish immigrants in Britain (Kozminska 2019), the expression of focus in Spanish in contact with Quechua (Muntendam and Torreira 2016; O'Rourke 2004, 2009; van Rijswijsk and Muntendam 2014) or K'ichee' (Baird 2017), or pitch changes in English in contact with Inuktitut (Colantoni et al. 2023). Additionally, the creation of mixed patterns of intonation, that is, combining properties of the bilinguals' two languages (Queen 2001, 2012, for Turkish-German bilingual children) and instances of bi-directional transfer (Mennen 2004, for Dutch-Greek bilinguals) are also documented. Intonation is a highly permeable feature of speech, which explains the diverse patterns that arise in contact situations where speakers have different intonation systems at their disposal. The study of these varieties can offer important insights about the organization and interrelation of said intonation systems in the speaker's mind, and can, ultimately, help inform models of bilingual intonation.

Regarding intonation in the Northern Basque country, the sole study that has been conducted to date has focused on Labourdin Basque (Duguine and Irurtzun 2020). Duguine and Irurtzun (2020) find rising contours as a common realization of yes-no questions, which is at odds with descriptions of other varieties of Basque and Basque Spanish in the Southern/Spanish Basque Country, where falling contours are more commonly reported (see Elordieta and Hualde 2014 for Basque; Romera and Elordieta 2020, for Basque Spanish; and Delgado 2024a for both, as well as references therein). These differences between

Labourdin Basque, Southern Basque varieties, and Basque Spanish suggest a complex prosodic reality that may derive from both different dialectal and sociolinguistic influences. Indeed, the role of the sociolinguistic factors such as language dominance or linguistic attitudes has been shown to have an effect on speakers' intonation productions in the Southern/Spanish Basque Country. For instance, González and Reglero (2020) and Delgado (2024a) report a higher prevalence of rising contours in yes-no questions among Spanish-dominant bilinguals in Basque Spanish (as does Delgado 2024a for Basque). Gaminde et al. (2011) also find a divergent distribution of contours depending on the speaker's L1 (e.g., more rising for L1 Spanish). In Basque, the presence of rising intonation, though sporadic, is also associated with higher Spanish influence among younger speakers (Elordieta 2003; Elordieta and Hualde 2014). To complete this picture, not only does this study provide the first account of intonation in French as spoken in the Northern Basque Country, but it also explores the potential role of language dominance on the results.

Our study of intonation is grounded on the Autosegmental Metrical Model (Pierrehumbert 1980; Ladd 2008), which assumes that the phonetic properties of intonation result from a hierarchical organization of phonological high (H) and low (L) tones. These tones can appear as single tones or in combination (e.g., LH, HL, etc.), and are divided into pitch accents (i.e., associated with stressed syllables) and boundary tones (i.e., associated with phrase boundaries).

In French, stress is not used distinctively at the lexical level as accentuable words are always stressed on their final syllable by default (Di Cristo 1998; Astésano 2016). Southern varieties of French also allow for penultimate stress, although this difference in stress position is still non-contrastive[4]. Due to the lack of contrastivity, this form of stress serves a demarcative function, marking prosodic constituents or phrasal structure. French distinguishes three phrasal levels: The accentual phrase or AP (the lowest level), the intermediate phrase (ip), and the intonation phrase (IP) (Delais-Roussarie et al. 2015, among others). The AP may include one or more lexical items and it has been described as a series of subsequent rises annotated as /LHiLH*/ (Jun and Fougeron 2000, 2002). In this sequence of tones, the main element is a final high tone (H*), whereas the remaining low tones (L) and the initial high tone (Hi) are optional. To mark the other two levels of phrasing (ip and IP), low and high boundary tones can be used (L-, H-, and L%, H% respectively). Since the purpose of this study is on final tonal movements, our focus will be on IP boundary tones and the obligatory tone in the final or nuclear AP in the sentence.

The remainder of the paper is organized as follows: Section 3 presents the research questions guiding the study and the methodology used to answer them, while Section 4 outlines the main acoustic results and attempts a first consideration of the role of language dominance in describing the results. Section 5 connects these findings to those of previous studies conducted in the Basque Country and elsewhere in France. Finally, Section 6 concludes with some final remarks.

## 3. Materials and Methods

The overarching question that this paper seeks to answer is as follows: What are the main intonational properties of information-seeking yes-no questions in bilingual Labourdin French? Since the role of language dominance is also explored, a secondary question focuses on whether this factor has an effect on the distribution of contours in our data.

In providing a description of yes-no questions in Labourdin French, this paper aims to fill in an important gap regarding the intonation of southern varieties of French and of those in contact with Basque. While the sole study on Labourdin Basque finds rising intonation in yes-no questions (as in standard French) (Duguine and Irurtzun 2020), we hypothesize that, as in other southern/bilingual varieties of French, falling contours are a possibility. This hypothesis is also supported by findings on Spanish and Basque in the Southern Basque Country, where falling intonation in yes-no questions is more common (Delgado 2024a; Elordieta and Hualde 2014; Romera and Elordieta 2020). Regarding the role of language dominance, we predict that some differences may arise in speakers' productions, in line with previous studies conducted in the Southern Basque Country.

The methodological design implemented to respond to these research questions is described below, including information on the participants, data collection methods, and the types of analyses conducted.

*3.1. Participants*

Participants were recruited through the researcher's personal contacts and the *Centre de recherche sur la langue et les textes basques* (IKER—UMR 5478) in Bayonne/Baiona, France. A total of 11 Basque-French bilinguals (4 males and 7 females) participated in the study. They all came from towns in Labourd, and their ages ranged between 19 and 31, with a mean age of 24.72 years (SD = 4.34). Five were sequential bilinguals (three with L1 French, and two with L1 Basque), five were simultaneous, and one was a late sequential bilingual with L1 French. Participants varied in their language of dominance, which was measured through the Bilingual Language Profile (BLP) questionnaire (Birdsong et al. 2012). This questionnaire enquiries about the participants' language history, use, proficiency, and attitudes towards both French and Basque. Based on their responses, participants receive scores for each of their languages, which are then subtracted to determine the speaker's language of dominance. The BLP scores, as well as other relevant demographic information are summarized in Table 1. The participants are ordered by dominance score: From more French-dominant (positive scores) to more Basque-dominant (negative scores).

**Table 1.** Participants' demographic information.

| Participant | Age | Gender | Bilingual Type | BLP | Dom. Language | Education | City of Origin |
|---|---|---|---|---|---|---|---|
| 22 | 31 | Male | Sim. | 44 | French | PhD | Ezpeleta |
| 17 | 23 | Female | Seq (L1 F) | 37 | French | Master's | Urruña |
| 23 | 19 | Female | Seq (L1 F) | 20 | French | Some university | Hazparne |
| 25 | 29 | Male | Late Seq (L1 F) | 15 | French | Master's | Baiona |
| 20 | 25 | Female | Sim. | −6 | Basque | Master's | Kanbo |
| 18 | 25 | Female | Sim. | −14 | Basque | BA | Baiona |
| 16 | 20 | Male | Seq (L1 B) | −15 | Basque | Some university | Luhuso |
| 19 | 25 | Female | Sim. | −16 | Basque | Master's | Urruña |
| 24 | 31 | Female | Seq (L1 F) | −19 | Basque | Master's | Baiona |
| 21 | 19 | Female | Sim. | −32 | Basque | Some university | Lasse |
| 15 | 25 | Male | Seq (L1 B) | −77 | Basque | Professional training | Itsasu |

For the purposes of the statistical analysis, participants were split into groups depending on their language of dominance (French vs. Basque), although their individual scores were still considered.

*3.2. Data Collection and Procedure*

Data collection took place in Bayonne/Baiona, France, in the spring of 2022. Recordings were performed in a quiet room at the IKER lab with a Zoom H4N Pro Digital Multitrack recorder set at 44,100 Hz in mono. Participants first completed the BLP questionnaire through Qualtrics in either Basque or French prior to the experimental session and then met with the researcher at IKER. For the experimental session, participants completed a contextualized reading task consisting of a voiced-over PowerPoint (based on previous works by González and Reglero 2018, 2020). As they went through the PowerPoint, participants heard a series of daily situations and had to respond to them by reading either a statement (broad or narrow focused) or a question (yes-no or Wh-questions). There were

12 target sentences and 2–3 practice items per condition, although only the yes-no questions are reported here. All materials were piloted prior to the administration of the experiment to ensure they elicited speech in a natural way.

Target items were controlled for syntactic structure and consisted of no more than three components, which could include a subject, a verb, and an object/complement. In addition, voicing of words in sentence-final position was also controlled for to guarantee the visibility of the intonation contour during this final portion of the sentence. An example of a context and corresponding target sentence is provided below. For a complete list of target questions, the reader is referred to Appendix A.

| | | | | | |
|---|---|---|---|---|---|
| (2) | Context: | *Ton nouveau colocataire est irlandais, mais ça fait 6 ans qu'il habite au Pays Basque. Demande-lui s'il retourne souvent en Irlande. Lis la question qui apparaît dans la diapositive qui suit:* 'Your new roommate is Irish, but he has been living in the Basque Country for 6 years now. Ask him if he travels back to Ireland often. Read the question that appears on the following slide'. | | | |
| | Target sentence: | *Tu* | *retournes* | *souvent* | *en Irlande?* |
| | | you.SG | return.PRS.2SG | often | to Ireland |
| | | 'Do you travel back/return to Ireland often?' | | | |

The task was completed in both Basque and French and it lasted approximately 50 min (~25 min per language). During the session, all interactions were conducted in French with some minor code-switching to Basque. Additionally, the order in which participants began was alternated to avoid any potential language effects. In total, 132 sentences were elicited (12 target sentences $\times$ 11 participants). Three of those items were discarded due to final devoicing, thus leaving 129 sentences for analysis.

### *3.3. Data Analysis*

All data were coded and analyzed in Praat (Boersma and Weenink 2021) and annotated according to the French ToBI (F_ToBI, Delais-Roussarie et al. 2015). Following its conventions, the main accent of the AP was marked with an asterisk [*], while boundary tones were marked with the percentage sign [%]. Since this project is part of a larger study also investigating Spanish and Basque in Gipuzkoa, annotations also considered the conventions of the Spanish ToBI (Sp_ToBI; Aguilar et al. 2009; Beckman et al. 2002; Hualde 2003; Face and Prieto 2007) for comparison purposes.

Intonation contours were classified into three main categories: Low rising, high rising, and falling. The rationale behind having two types of rising (low vs. high) comes from differences in the general shape of the contours. While in low rising intonation the final rise is preceded by a low tone, in high rising contours the F0 is at a high level well before the final nuclear syllable. Typically, this rise takes place at the beginning of the question and plateaus thereafter (see Section 4.1. for more information). In the section that follows, the acoustic results of these three main contours, as well as their distribution by language dominance group, are presented.

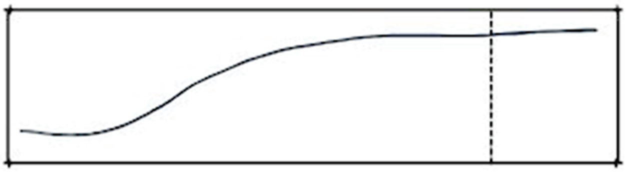
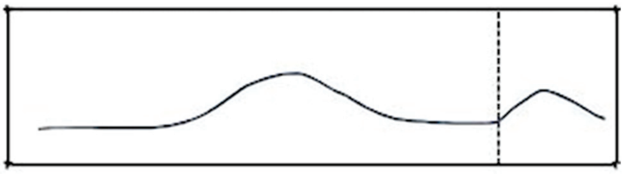
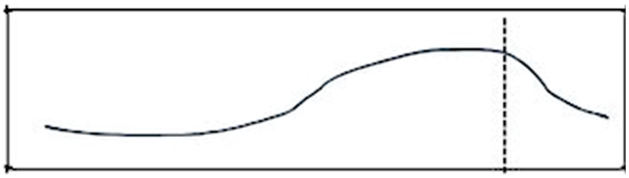
## 4. Results

### *4.1. Acoustic Results*

A representation of each of these contours is provided in Table 2. While the main focus of this work lies on the final portion of the sentence, the general shape of the intonation curve is also provided. Note that the final part of the sentence is marked with a dashed line.

As shown, two alternatives are provided for falling contours: (a) One with a rise from a low tone and a subsequent fall (rising-falling), and (b) another one with a fall from a high tone (high falling). Neither of them, however, was very common. Variations in the shapes of the three major contours proposed were possible (for example, some sentences presented two pre-final prominences), but all final tonal movements abided by what is shown in Table 2.

**Table 2.** Main contour categories in Labourdin French yes-no questions.

| Type of Contour | General Shape | ToBI Labeling |
|---|---|---|
| Low rising |  | L(+H)* H% |
| High rising |  | (¡)H* H% |
| Falling | (a) Rising-falling  | L+H* (H)L% |
| | (b) High falling  | (¡)H* L% |

Overall, questions ending in a high boundary tone (H%) predominated (90% of cases), with low rising being the more common realization (72.1%). This was followed by high rising, attested in 17.9%. Falling contours ((H)L%), in turn, amounted to only 10% of the data. The specific ToBI annotations of these contours together with their frequency of occurrence, are displayed in Table 3.

**Table 3.** Distribution of nuclear configurations according to the type of contour.

| Type of Contour | ToBI Annotation | Number of Sentences | Percentage of Usage | Total |
|---|---|---|---|---|
| Low rising | L+H* H% | 89 | 69% | 72.1% |
| | L* H% | 4 | 3.1% | |
| High rising | (¡)H* H% | 23 | 17.9% | 17.9% |
| Falling/circumflex | L+H* L% | 9 | 7% | 10% |
| | L+H* HL% | 2 | 1.5% | |
| | (¡)H* L% | 2 | 1.5% | |
| Total | | 129 | 100% | 100% |

As displayed in Table 3, low rising contours could present either a rising (L+H*) or a low (L*) nuclear tone depending on whether the rise took place at the onset of the stressed syllable or after. High rising contours, in turn, could be upstepped. To be considered a high upstepped tone, a rise (>7 Hz) took place from an already high tone. Both the onset and offset of said rise were close to the speaker's highest pitch point. For falling contours, there could be variation in the presence of a low tone in the nuclear syllable ((L+)H*) or a bitonal boundary tone (HL% as opposed to L%). To instantiate these contours, representative examples are provided in Figures 2–4 below. For clarity in the presentation of contours, questions ending in paroxytone words are displayed, although the same contours also applied to sentences ending in oxytone words. The percentage of yes-no questions ending in paroxytone words amounted to 36.4% (47 sentences), but the location of stress did not affect the participants' contour choices[5].

First, Figure 2 illustrates the low rising contour L+H* H% for the question *Tu as téléchargé l'épisode?* 'Have you downloaded the episode?'. The final word of the sentence *épisode* 'episode' is characterized by an initial peak on the first syllable, which could correspond to the optional initial accent Hi, a subsequent fall, and a rise on the final stressed syllable <so> in *épi<u>so</u>de* 'episode'. The rise on this syllable stems from a lower tone at the onset of the stressed syllable (L+H*) and culminates in a high (H%) boundary tone.

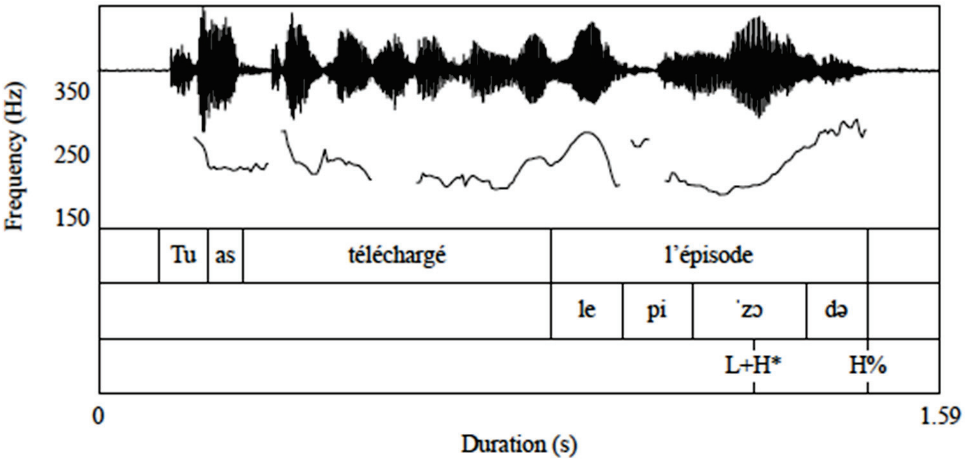



**Figure 2.** Example of L+H* H% for the question *Tu as téléchargé l'épisode?* 'Have you downloaded the episode?' Participant 21.

Figure 3 shows an instance of a high rising (¡)H* H% contour. In this case, the pitch rises early in the sentence and remains at a high level throughout the question. In contrast with the example in Figure 2, the tone is already high at the onset of the final stressed syllable (<der> in *mo<u>der</u>ne* 'modern').

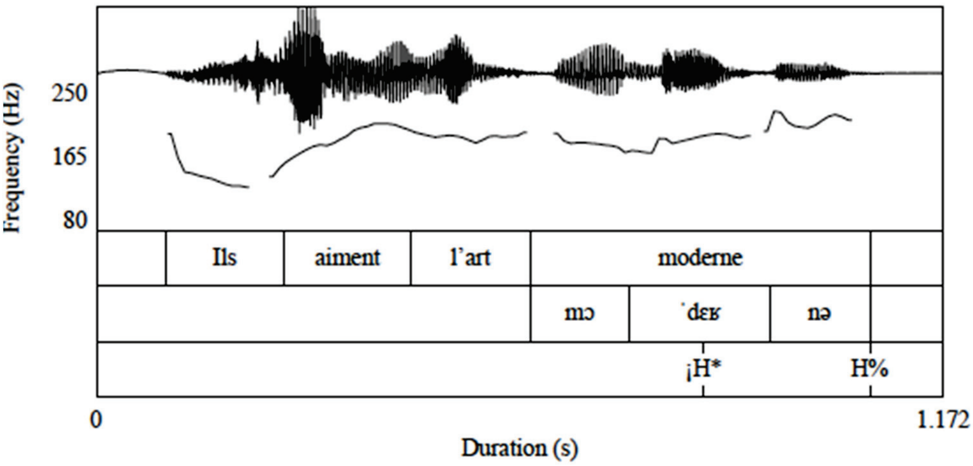



**Figure 3.** Example of (¡)H* H% for the question *Ils aiment l'art moderne?* 'Do they like modern art?' Participant 16.

Finally, questions produced with a falling configuration were scarce, but as a way of illustration, Figure 4 shows an example of the most common falling contour in our data:

L+H* L%. The nuclear accent (<lan> in *Irlande* 'Ireland') is characterized by a rise from a low tone (marked as L+H*). After reaching the peak of this tone, the pitch falls, thereby ending the question in a low boundary tone L%. An interesting aspect of the example in Figure 4 is the peak at the end of the word *souvent* 'often'. This is due to the presence of an intermediate phrase (ip).

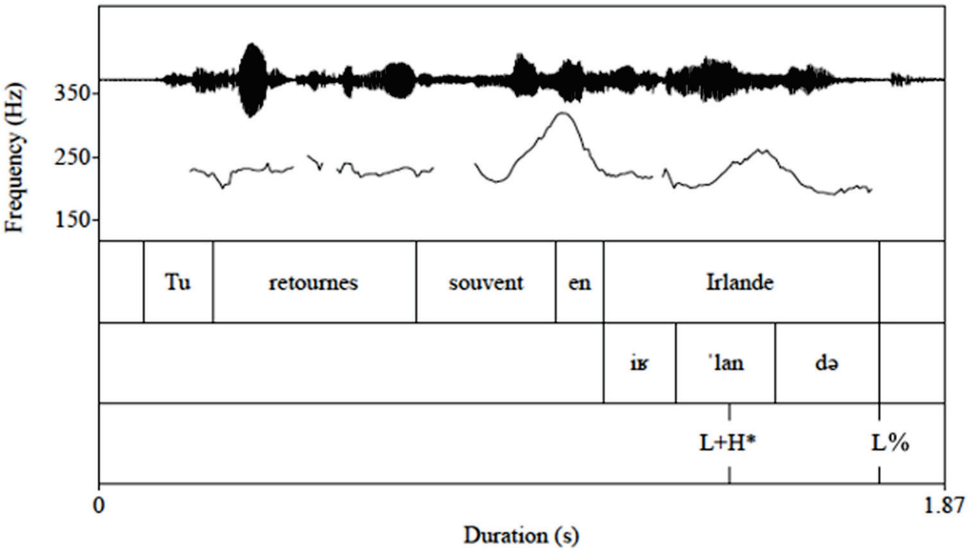



**Figure 4.** Example of L+H* L% for the question *Tu retournes souvent en Irlande?* 'Do you go back to Ireland often?' Participant 17.

### *4.2. Participant Differences Based on Language Dominance*

In the present participant pool, there were seven Basque-dominant participants and four French-dominant participants (see Table 1 for details). The distribution of contours by language dominance group is shown in Table 4.

**Table 4.** Contour distribution and frequency across language dominance groups.

| | French-Dominant | | Basque-Dominant | |
|---|---|---|---|---|
| **Contour Type** | **Sentences** | **%** | **Sentences** | **%** |
| Low rising | 27 | 57.4% | 66 | 80.5% |
| High rising | 9 | 19.1% | 14 | 17% |
| Falling | 11 | 23.5% | 2 | 2.5% |
| Total | 47 | 100% | 82 | 100% |

In both groups, rising contours predominated. While the use of high rising intonation appeared similar across groups, large differences are observed in the low rising category (57.4% for French-dominant vs. 80.5% for Basque-dominant speakers). Substantial differences were also noted in the falling category (23.5% for French-dominant vs. 2.5% for Basque-dominant). To examine the significance of the percentage differences displayed in Table 4, chi-squared tests were performed in R (R Core Team 2022). The results showed significant differences for both the low rising and falling categories, meaning that low rising contours were significantly more common in the Basque-dominant group ($X^2 = 6.231$, df = 1, $p < 0.001$) and falling ones appeared more frequently in the French-dominant group ($X^2 = 16.355$, df = 1, $p < 0.05$).

These differences were further explored by considering individual participants' productions together with their dominance scores. Figure 5 presents this distribution. In this graph, participants are ordered according to their dominance score, with the most French-dominant speaker being placed on the left-hand edge of the graph, and the most

Basque-dominant speaker on the right-hand edge of the graph. The closer participants are to the middle point separating both groups, the more balanced bilinguals they are.

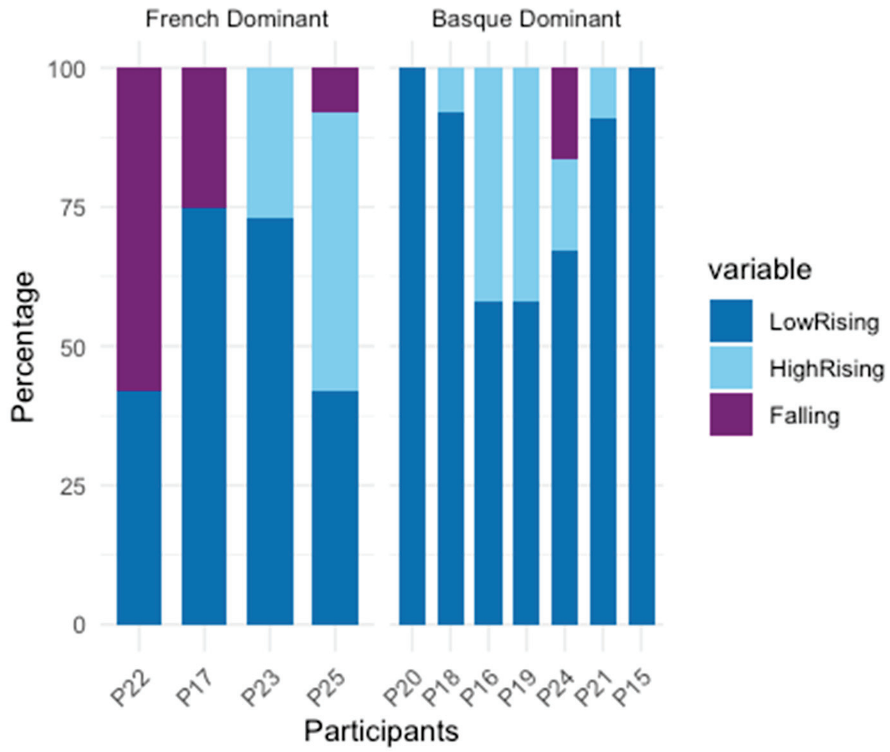



**Figure 5.** Participant distribution of contours by language dominance score.

With the exception of participant 22, all participants showed a preference for rising contours (attested in ≥75% vs. 42% for participant 22). Participant 25 exhibited a slightly higher preference for high rising contours, while the rest of the participants produced more low rising contours. Falling contours were produced by three out of four participants in the French-dominant group, although they were more common in participants 22's productions.

To further explore the relationship between language dominance scores and the likelihood of different intonation contours appearing in the data, a multinomial logistic regression analysis was conducted in R (R Core Team 2022) using the multinon() function within the "nnet" package (Ripley and Venables 2022) and OpenAI (2023) to assist with code writing and result interpretation. The dependent variable was the type of intonation contour, which had three levels, low rising, high rising, and falling, and language dominance score was included as an independent numeric predictor. Different models were used to try to account for participant and item number as random effects, in line with current approaches for mixed-effects models (Baayen et al. 2008), but due to the small size of the sample this led to a singularity problem. Therefore, the model was run without accounting for random factors. The results should thus be taken cautiously and replicated with a larger and more balanced participant group in the future. For our participant groups, Table 5 shows the summary of the coefficients, standard errors, and *p*-values yielded by the analysis[6]. "Low rising" was chosen as the reference category as it was the most commonly found contour in both groups. The coefficients listed thus refer to the likelihood of finding contours other than low rising (i.e., falling, high rising) in the data depending on dominance score.

**Table 5.** Results from multinomial logistic regression analysis with "low rising" as reference category.

| Outcome Category | Coefficient (Intercept) | SE (Intercept) | *p*-Value[7] (Intercept) | Coefficient (Dscore) | SE (Dscore) | *p*-Value (Dscore) |
|---|---|---|---|---|---|---|
| Falling | −3.15 | 0.637 | <0.001 | 0.064 | 0.017 | <0.001 |
| High rising | −1.38 | 0.234 | <0.001 | 0.007 | 0.008 | 0.359 |

The intercept coefficients suggest that, when the dominance score is zero, there is a negative relationship with both falling and high rising intonation, meaning that the occurrence of these contours is overall less likely ($p < 0.001$). When looking at the coefficients for dominance score (Dscore), that is, those that consider our speakers' actual dominance scores, only the falling category reached significance ($p < 0.001$). The positive sign of the coefficient suggests that falling contours are more likely to appear as dominance score increases, that is, among French-dominant speakers. On the other hand, there were no significant differences regarding the presence of high rising contours ($p = 0.359$). Although in a limited way, these results corroborate our X-squared results and suggest a nuanced effect of language dominance on intonation preferences that merits further investigation.

## 5. Discussion

This study has provided a first account of the intonation of yes-no questions in bilingual Labourdin French, a variety of French spoken in Iparralde (southern France). Since there are no previous descriptions of this variety, the previous literature on standard and bilingual varieties of French is considered below for comparison purposes. Furthermore, our preliminary results on language dominance are compared to those of other studies conducted in the Southern Basque Country.

### *5.1. Labourdin French and Other Varieties Compared*

Our data showed that rising contours (low and high rising) were the preferred realization for yes-no questions in Labourdin French. This concurs with what has been reported for standard varieties of French across different studies (Delais-Roussarie et al. 2015; Sichel-Bazin et al. 2012; Sichel-Bazin 2015; among others). The most common realization in our data consisted of a low rise that usually took place on the final syllable of the last word of the question, resulting in a L+H* H% contour. In cases of penultimate accentuation, the rise could begin on the penultimate syllable, although, as noted above, this did not incur in changes to the general shape of the contour. Another common type of contour was the high rise (¡)H* H%. Questions with this configuration were characterized by a rise at the beginning of the question that plateaued until the end. Delais-Roussarie et al. (2015) also propose this contour type as a common realization of yes-no questions. We do observe, however, some subtle differences. In our data, the rise in pitch takes place early in the sentence, while in the example provided by Delais-Roussarie et al. (2015), the pitch rises slowly throughout the sentence and it is not until the last accented syllable that it rises more steeply. Whether these rises are significantly or meaningfully different is an issue that calls for further exploration.

As in bilingual varieties of French in contact with Corsican (Boula de Mareüil et al. 2012, 2016a) and Occitan (Boula de Mareüil et al. 2016b; Sichel-Bazin et al. 2012; Sichel-Bazin 2015), falling contours were also attested in our data, although their presence only represented 10% of all cases. This contrasts with French in contact with Corsican, where Boula de Mareüil et al. (2012, 2016a) found falling contours to predominate. These authors observed similar contours in Corsican, thus suggesting that the presence of falling contours could be influenced by this language. Similarly, research conducted in Occitan-French-speaking areas provides further evidence for this interrelation of languages. For example, Sichel-Bazin (2015) finds similar intonational patterns in both Occitan and French in contact. Outside of the French-speaking world, Fernández Rei (2019) and Robles-Puente (2012) also find the same intonation contours for yes-no questions across both of the bilinguals' languages (respectively: Galician and Spanish; Basque and Spanish).

In our own Basque data for the same participants (Delgado 2024b), the use of falling contours was sporadic (2.7% of sentences), which falls in line with Duguine and Irurtzun's (2020) findings on Labourdin Basque. One could surmise that the limited presence of falling intonation reflects a linguistic reality where French is the predominant language. However, the higher presence of falling contours among French-dominant participants is an issue that calls for further investigation. It is worth noting that, in our data for Labourdin Basque, high rising contours were significantly more common in Basque than in French.

### *5.2. A Comparison of Labourdin French and Spanish and Basque in the Southern Basque Country*

As noted in Section 2, Labourdin French is in close proximity to Spanish and Basque varieties in the Southern/Spanish Basque Country. Our finding that rising contours predominate in Labourdin French contrasts with descriptions of other varieties of Basque and Basque Spanish in the Southern Basque Country. In both of these languages, falling contours have been reported to be more common than rising contours across varieties (for Basque see Aurrekoetxea et al. 2011; Delgado 2024a; Elordieta 2003; Elordieta and Hualde 2014; Robles-Puente 2012; for Basque Spanish see Delgado 2024a; Elejabeitia et al. 2008; Elordieta and Romera 2020; Gaminde et al. 2011; González and Reglero 2020; Robles-Puente 2011, 2012; Romera and Elordieta 2020). The presence of rising contours, in turn, has often been associated with more Spanish language dominance (for Spanish see Delgado 2024a; González and Reglero 2020; for Basque see Delgado 2024a; Elordieta 2003; and Elordieta and Hualde 2014). One could have thus expected to find rising intonation in the speech of French-dominant participants. However, preliminary results from the present study exhibited the reverse trend: falling contours were more common in the French-dominant group. This was indeed a surprising finding, although we believe it might be, at least in some cases, related to the participants' personal experiences with bilingualism rather than with language dominance per se. For instance, participant 22, who was the participant that produced the highest number of falling contours (58%), had spent some time in the Southern Basque Country performing research for his doctorate. Thus, his productions could have been influenced by the varieties of Basque and Spanish spoken there, both of which are characterized by falling intonation (Delgado 2024a; Elordieta and Hualde 2014; and references therein). The case of participant 17, who produced 25% of falls, could potentially be explained by her town of origin, Urruña (Urrugne), which is located closer to the Southern Basque Country. Our study included speakers from different towns in Labourd to provide an initial description of intonation in the province, but future studies should further compare bilingual speakers' productions within and across towns to determine the role of origin in intonation patterns in the Northern Basque Country. Lastly, for participants 24 (Basque-dominant) and 25 (French-dominant), the use of falling intonation seemed sporadic (16.5% and 8% respectively). These results do not show a clear trend within the French-dominant group, which could be related to the small size of the participant sample. These preliminary findings should thus be expanded in the future with larger and more balanced participant groups. It could also be the case that there are other factors at play. For example, Elordieta and Romera (2020) and Romera and Elordieta (2020) do not observe an effect of language dominance in their yes-no question intonation data for urban and rural towns in Bizkaia and Gipuzkoa but do find that linguistic attitudes and contact with the Basque ethnolinguistic group have an effect on the contours found in urban areas (i.e., less spread of bilingualism). Further exploration of these factors is also needed.

## 6. Conclusions

In conclusion, the present study presents the first description of the intonation of yes-no questions in Labourdin French. Our results showed that, as in standard French, rising contours predominated across participants with the preferred realization being low rising (L+H* H%), followed by high rising ((¡)H* H%). A preliminary consideration of language dominance indicated that falling intonation was more common among French-dominant participants. However, this finding could have been influenced by the participants' individual experiences with bilingualism and their town of origin. A replication of this study is thus needed with a larger participant pool and more control over the participants' town of origin. This will also help to overcome the limitations of the statistical model used.

Future research should continue to explore the intonation of French (and Basque) in the Northern Basque Country to deepen our knowledge of prosodic differences and similarities across varieties and to obtain a better understanding of any existent cross-linguistic influences. Not only will this fill an important gap in the research of intonation, but it will also advance our understanding of bilingual varieties in present-day France.

**Funding:** This material is based upon research supported by the Chateaubriand Fellowship in Humanities and Social Sciences of the Embassy of France in the United States.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board (or Ethics Committee) of Florida State University (STUDY0000214, 26 March 2021).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The dataset presented in this article is not readily available because of the conditions specified in the IRB approval for the project.

**Acknowledgments:** I would like to thank all participants in the study for their time and everyone at the *Centre de Recherche sur la Langue et les Textes Basques* in Baiona for welcoming me into their lab in the spring of 2022 and for their assistance in the data collection process. A special thanks goes to the three anonymous reviewers who, through their feedback and comments, helped to improve the quality of this paper. All errors, of course, remain mine. Finally, I would like to acknowledge OpenAI as it was a valuable resource when performing all statistical analyses.

**Conflicts of Interest:** The author declares no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## Appendix A

List of target items:

- *La bague est neuve?* 'Is the ring new?'
- *Tu retournes souvent en Irlande?* 'Do you go back to Ireland often?'
- *Tu as de la pomade à la maison?* 'Do you have ointment at home?'
- *Vous avez du pain aux graines?* 'Do you have bread with seeds?'
- *L'amie d'Edurne va venir?* 'Is Edurne's friend coming?'
- *Il y a du beurre dans les madeleines?* 'Is there butten in the madeleines?'
- *Vous avez le numéro d'Ainhoa?* 'Do you have Ainhoa's number?'
- *Tu aimes le jus de mangue?* 'Do you like mango juice?'
- *Il y a eu des inondations dans la région?* 'Have there been floods in the region?'
- *Tu as téléchargé l'episode?* 'Have you downloaded the episode?'
- *Tu t'es deguisée pour le carnaval?* 'Did you dress up for carnival?'
- *Ils aiment l'art moderne?* 'Do they like modern art?'

## Notes

1. Iparralde is a Basque word that translates to 'to the north'. Geografically, it refers to the part of the Basque Country north of the Spain-France political border.
2. The VII Sociolinguistic Survey (Office Publique de la Langue Basque—Euskararen Erakunde Publikoa 2023) separates Bayonne-Anglet-Biarritz (BAB) from the rest of Lapurdi in their statistics. Since our participants were from both the BAB and the province of Lapurdi, we have decided to combine both and provide averages here. The reader is referred to the original survey for these separate statistics.

3 As a result of a constitutional reform in 2008, Basque (as well as other minority languages) came to be recognized as part of the patrimony of the country. This change is part of the *LOI constitutionnelle no 2008-724 du 23 juillet 2008 de modernisation des institutions de la V$^{e}$ République (1)* (Secrétariat Général du Gouvernement 2008).

4 Penultimate stress in Southern varieties of French occurs due to the presence of a schwa sound in word final position (Coquillon 2005; Eychenne 2009; Eychenne 2019; Sichel-Bazin et al. 2012). This schwa increases the number of syllables and, as a result, stress appears on penultimate position. Irrespective of its position, stress is still non-contrastive in Southern French.

5 Chi-squared tests were performed to test for any differences in contour distribution based on the location of stress. Results were not significant ($p > 0.05$).

6 *p* values were calculated using the "broom" package in R (Robinson and Hayes 2022; R Core Team 2022).

7 See note 6.

## References


Aguilar, Lourdes, Carme de-la-C. Mota, and Pilar Prieto. 2009. Sp_ToBI Training Materials. Available online: http://prosodia.upf.edu/sp_tobi/ (accessed on 15 February 2024).

Astésano, Corine. 2016. Prosodic characteristics of reference French. In *Varieties of Spoken French*. Edited by Sylvain Detey, Jacques Durand, Bernard Laks and Chantal Lyche. Oxford: Oxford University Press, pp. 68–85.

Aurrekoetxea, Gotzon, Iñaki Gaminde, and Aitor Iglesias. 2011. Corpus- based prosodic variation in Basque: Y/N questions marked with the particle al. *Estudios de Fonética Experimental* 20: 11–31.

Baayen, Rolf Harald, Douglas. J. Davidson, and Douglas M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59: 390–412. [CrossRef]

Baird, Brandon O. 2017. Prosodic transfer among Spanish-K'ichee' bilinguals. In *Multidisciplinary Approaches to Bilingualism in the Hispanic and Lusophone World*. Edited by Kate Bellamy, Michael W. Child, Paz González, Antje Muntendam and M. Carmen Parafita Couto. Amsterdam: John Benjamins Publishing Company, pp. 149–74.

Beckman, Mary E., Manuel Díaz-Campos, Julia Tevis McGory, and Terrell A. Morgan. 2002. Intonation across Spanish, in the Tones and Break Indices Framework. *International Journal of Romance Linguistics* 14: 9–36. [CrossRef]

Birdsong, David, Libby M. Gertken, and Mark Amengual. 2012. *Bilingual Language Profile: An Easy to-Use Instrument to Assess Bilingualism*. Austin: University of Texas at Austin—The Center for Open Educational Resources and Language Learning.

Bochman, Klaus. 2018. Language policies in the Romance-speaking countries of Europe. In *Manual of Romance Sociolinguistics*. Edited by Wendy Ayres-Bennett and Janice Carruthers. Berlin: De Gruyter, pp. 433–61.

Boersma, Paul, and David Weenink. 2021. Praat: Doing Phonetics by Computer [Computer Program]. Version 6.1.56. Available online: http://www.praat.org/ (accessed on 20 June 2022).

Bordal, Guri, and Chantal Lyche. 2012. Regard sur la prosodie du français d'Afrique à la lumiere de la L1 des locuteurs. In *La variation prosodique régionale en français*. Edited by Anne Catherine Simon. Louvain-la-Neuve: De Boeck Supérieur, pp. 179–98.

Boula de Mareüil, Philippe, Albert Rilliard, and Hélène Maynard. 2016a. Questions à intonation descendante: An example d'isolat corse? In *Lingue delle isole, isole linguistiche*. Edited by Stella Retali-Medori. Alessandria: Edizioni dell'Orso, pp. 3–19.

Boula de Mareüil, Philippe, Albert Rilliard, Paolo Mairano, and Jean-Pierre Lai. 2012. Corsican French questions: Is there a prosodic transfer from Corsican to French and how to highlight it? Paper presented at the Sixth International Conference on Speech Prosody, Shanghai, China, May 22–25.

Boula de Mareüil, Philippe, Fanny Ivent, Albert Rilliard, and Varvara Kozhevina. 2016b. Une étude prosodique comparative des questions en français en contact avec l'occitan et le catalan. *Languages* 202: 75–91. [CrossRef]

Braun, Bettina, Nicole Dehé, Jana Neitsch, Daniela Wochner, and Katharina Zahner. 2019. The prosody of rhetorical and information-seeking questions in German. *Language and Speech* 62: 779–807. [CrossRef] [PubMed]

Bullock, Barbara E. 2009. Prosody in contact in French: A case study from a heritage variety in the USA. *International Journal of Bilingualism* 13: 165–94. [CrossRef]

Colantoni, Laura. 2011. Broad-focus declaratives in Argentine Spanish contact and non-contact varieties. In *Intonational Phrasing in Romance and Germanic: Cross-Linguistic and Bilingual Studies*. Edited by Conxita Lleó and Christoph Gabriel. Amsterdam: John Benjamins Publishing Co., pp. 183–212.

Colantoni, Laura, and Jorge Gurlekian. 2004. Convergence in intonation: Historical evidence from Buenos Aires Spanish. *Bilingualism: Language and Cognition* 7: 107–19. [CrossRef]

Colantoni, Laura, Gabrielle Klassen, Matthew Patience, Malina Radu, and Olga Tararova. 2023. Perception and Production of Sentence Types by Inuktitut-English Bilinguals. *Languages* 7: 193. [CrossRef]

Communauté d'Agglomération Pays Basque—Euskal Hirigune Elkargoa. 2018. La Politique Lingusitique. Available online: https://www.communaute-paysbasque.fr/basque-et-gascon/la-langue-basque/ressources/la-politique-linguistique (accessed on 25 February 2024).

Coquillon, Annelise. 2005. Caractérisation Prosodique du Parler de la Région Marseillaise. Doctoral dissertation, Université Aix-Marseille, Marseille, France.

Dehé, Nicole, and Bettina Braun. 2019. The prosody of rhetorical questions in English. *English Language and Linguistics* 24: 607–35. [CrossRef]

Delais-Roussarie, Elisabeth, Brechtje Post, Mathieu Avanzi, Carolin Buthke, Albert Di Cristo, Ingo Feldhausen, Sun-Ah Jun, Philippe Martin, Trudel Meisenburg, Annie Rialland, and et al. 2015. Intonational phonology of French: Developing a ToBI system for French. In *Intonation in Romance*. Edited by Sónia Frota and Pilar Prieto. Oxford: Oxford University Press, pp. 63–74.

Delattre, Pierre. 1966. Les dix intonations de base du français. *The French Review* 40: 1–14.

Delgado, Nerea. 2024a. The intonation of yes-no questions in Gipuzkoan Spanish and Basque from Irun. In *Recent Developments in Hispanic Linguistics: Studies in Structure, Variation, and Bilingualism*. Edited by Sean McKinnon and Michael Gradoville. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Delgado, Nerea. 2024b. *The Intonation of Spanish and Basque in Gipuzkoa and French and Basque in Labourd: An Exploratory Study*. Tallahassee: Florida State University.

Di Cristo, Albert. 1998. Intonation in French. In *Intonation Systems: A Survey of Twenty Languages*. Edited by Daniel Hirst and Albert Di Cristo. Cambridge, MA: Cambridge University Press, pp. 195–218.

Di Cristo, Albert. 2016. *Les musiques du français parlé: Essais sur l'accentuation, la métrique, le rythme, le phrasé prosodique et l'intonation du français contemporain*. Boston: De Gruyter.

Dryer, Matthew S. 2013. Polar Questions. In *WALS Online [Data Set]*. Edited by Matthew S. Dryer and Martin Haspelmath. Available online: https://zenodo.org/records/7385533 (accessed on 25 February 2024).

Duguine, Maia, and Aritz Irurtzun. 2020. Prosody and Language Contact: An Experimental Investigation of Interrogative Strategies in Navarro-Labourdin Basque. Paper presented at the Tenth International Conference on Speech Prosody, Tokyo, Japan, May 25–28.

Elejabeitia, Ana, Alexander Iribar, and Rosa Miren Pagola. 2008. La prosodia del castellano en tres ciudades vascas: Oraciones con y sin expansión. *Language Design* 10: 147–54.

Elordieta, Gorka. 2003. Intonation. In *A Grammar of Basque*. Edited by José Ignacio Hualde and Jon Ortiz de Urbina. Berlin: De Gruyter, pp. 72–113.

Elordieta, Gorka, and José Ignacio Hualde. 2014. Intonation in Basque. In *Prosodic Typology II: The Phonology of Intonation and Phrasing*. Edited by Sun-Ah Jun. Oxford: Oxford University Press, pp. 405–463.

Elordieta, Gorka, and Magdalena Romera. 2020. The influence of social factors on the prosody of Spanish in contact with Basque. *International Journal of Bilingualism* 25: 286–317. [CrossRef]

Eychenne, Julien. 2009. La prononciation du français au Pays Basque. In *Phonologie, Variation et Accents du Français*. Edited by Jacques Durand, Bernard Laks and Chantal Lyche. France: Hermès Science Publications, pp. 231–58.

Eychenne, Julien. 2019. On the deletion of word-final schwa in Southern French. *Phonology* 36: 355–89. [CrossRef]

Face, Timothy, and Pilar Prieto. 2007. Rising accents in Castilian Spanish: A revision of Sp_ToBI. *Journal of Portuguese Linguistics* 6: 117–46. [CrossRef]

Fernández Rei, Elisa. 2019. Galician and Spanish in Galicia: Prosodies in contact. In *Sociolinguistic Patterns and Processes of Convergence and Divergence in Spanish*. Edited by Isabel Molina Martos, Florentino Paredes García and Ana M. Cestero Mancera. Amsterdam/Philadelphia: John Benjamins Publishing Company, pp. 438–61.

Fónagy, Ivan, and Eva Bérard. 1973. Questions totales simples et implicatives en français parisien. *Studia Phonetica* 8: 53–79.

Gaminde, Iñaki, Asier Romero, Urtza Garay, and Aintzane Etxebarria. 2011. Los tonos de frontera de las oraciones interrogativas absolutas producidas por hablantes bilingües vasco-español. *Estudios de Lingüística Aplicada* 29: 61–79.

González, Carolina, and Lara Reglero. 2018. Dime una cosa: Are wh-in-situ questions different in Spanish? Evidence from intonation. In *Romance Languages and Linguistic Theory 14. Selected Papers from the 46th Linguistic Symposium on Romance Languages (LSRL)*. Edited by Lori Repetti and Francisco Ordóñez. Amsterdam/Philadelphia: John Benjamins Publishing Company, pp. 85–102.

González, Carolina, and Lara Reglero. 2020. The intonation of yes-no questions in Basque Spanish. In *Language Patterns in Spanish and Beyond*. Edited by Juan J. Colomina-Almiñana and Sandro Sessarego. London: Routledge, pp. 223–42.

Hualde, José Ignacio. 2003. El modelo métrico-autosegmental. In *Teorías de la entonación*. Edited by Pilar Prieto. Barcelona: Ariel, pp. 155–84.

Jun, Sun-Ah, and Cécile Fougeron. 2000. A phonological model of French intonation. In *Intonation: Analysis, Modeling and Technology*. Edited by Antonis Botinis. Dordrecht: Kluwer Academic Publishers, pp. 209–42.

Jun, Sun-Ah, and Cécile Fougeron. 2002. Realizations of accentual phrase in French intonation. *Probus* 14: 147–72. [CrossRef]

Kozminska, Kinga. 2019. Intonation, identity and contact induced change among Polish-speaking migrants in the U.K. *Journal of Sociolinguistics* 23: 29–53. [CrossRef]

Ladd, D. Robert. 2008. *Intonational Phonology*. Cambridge, MA: Cambridge University Press.

Martin, Philippe. 2012. Les contours de continuation majeure dans l'Océan Indien. In *La variation prosodique régionale en français*. Edited by Anne Catherine Simon. Louvain-la-Neuve: De Boeck Supérieur, pp. 199–211.

Mennen, Ineke. 2004. Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics* 32: 543–63. [CrossRef]

Muntendam, Antje, and Francisco Torreira. 2016. Focus and prosody in Spanish and Quechua: Insights from an interactive task. In *Intonational Grammar in Ibero-Romance: Approaches Across Linguistic Subfields*. Edited by Meghan E. Armstrong, Nicholas Henriksen and María del Mar Vanrell. Amsterdam/Philadelphia: John Benjamins Publishing, pp. 69–90.

Office Publique de la Langue Basque—Euskararen Erakunde Publikoa. 2023. Résultats de la VII Enquête Sociolinguistique pour le Pays Basque Nord. Available online: https://www.mintzaira.fr/fr/outils/les-actualites/actualite/article/resultats-de-la-viieme-enquete-sociolinguistique-pour-le-pays-basque-nord.html (accessed on 25 February 2024).

OpenAI. 2023. ChatGPT-3.5 [Computer Software]. Available online: http://openai.com (accessed on 10 December 2023).

OpenStreetMap contributors. n.d. *OpenStreetMap*. Available online: http://openstreetmap.org (accessed on 20 February 2024).
O'Rourke, Erin. 2004. Peak placement in two regional varieties of Peruvian Spanish intonation. In *Contemporary Approaches to Romance Linguistics: Selected Papers from the 33rd Linguistic Symposium on Romance Languages (LSRL)*. Edited by Julie J. Auger, J. Clancy Clements and Barbara S. Vance. Amsterdam/Philadelphia: John Benjamins Publishing, pp. 321–41.
O'Rourke, Erin. 2009. The realization of contrastive focus in Peruvian Spanish intonation. *Lingua* 122: 494–510. [CrossRef]
Pierrehumbert, Janet. 1980. The Phonology and Phonetics of English Intonation. Doctoral dissertation, MIT, Cambridge, MA, USA.
Post, Brechtje. 2000. *Tonal and Phrasal Structures in French Intonation*. The Hague: Thesus.
Queen, Robin M. 2001. Bilingual intonation patterns: Evidence of language change from Turkish-German bilingual children. *Language in Society* 30: 55–80. [CrossRef]
Queen, Robin M. 2012. Turkish-German bilinguals and their intonation: Triangulating evidence about contact-induced language change. *Language* 88: 791–816. [CrossRef]
R Core Team. 2022. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. Available online: www.R-project.org/ (accessed on 20 November 2023).
Ripley, Brian, and William Venables. 2022. Package 'nnet'. The Comprehensive R Archive Network (CRAN). Available online: https://cran.r-project.org/web/packages/nnet/nnet.pdf (accessed on 20 November 2023).
Robinson, David, and Alex Hayes. 2022. Package 'Broom'. The Comprehensive R Archive Network (CRAN). Available online: https://cran.r-project.org/web/packages/broom/broom.pdf (accessed on 20 November 2023).
Robles-Puente, Sergio. 2011. Absolute questions do not always have a rising pattern: Evidence from Bilbao Spanish. In *Selected Proceedings of the 5th Conference on Laboratory Approaches to Romance Phonology*. Edited by Scott. M. Alvord. Somerville: Cascadilla Proceedings Project, pp. 98–107.
Robles-Puente, Sergio. 2012. Two languages, two intonations? Statements and yes/no questions in Spanish and Basque. *International Journal of Basque Linguistics and Philology* 46: 252–62.
Romera, Magdalena, and Gorka Elordieta. 2013. Prosodic accommodation in language contact: Spanish intonation in Majorca. *International Journal of the Sociology of Language* 221: 127–51. [CrossRef]
Romera, Magdalena, and Gorka Elordieta. 2020. Information-seeking question intonation in Basque Spanish and its correlation with degree of contact and language attitudes. *Languages* 5: 70. [CrossRef]
Sadat-Tehrani, Nima. 2011. The intonation patterns of interrogatives in Persian. *Linguistic Discovery* 9: 105–36. [CrossRef]
Secrétariat Général du Gouvernement. 2008. LOI Constitutionnelle no 2008-724 du 23 juillet 2008 de Modernisation des Institutions de la V[e] République (1). Légifrance. Available online: https://www.legifrance.gouv.fr/download/pdf?id=PEy7_HFQtChq59a_W_8fBZ96WAkxk7JLKoUd3uP63A4= (accessed on 25 February 2024).
Sichel-Bazin, Rafèu. 2015. Prosodic Systems in Contact: Occitan and French. Doctoral dissertation, Universität Osnabrük, Osnabrük, Germany. Universitat Pompeu Fabra, Barcelona, Spain.
Sichel-Bazin, Rafèu, Carolin Buthke, and Trudel Meisenburg. 2012. The prosody of Occitan-French bilinguals. In *Multilingual Individuals and Multilingual Societies*. Edited by Christoph Gabriel and Kurt Braumuller. Amsterdam/Philadelphia: John Benjamins Publishing Company, pp. 349–64.
Simonet, Miquel. 2009. Nuclear pitch accents in Majorcan Catalan declaratives: Phonology, phonetics, diachrony. *Studies in Hispanic and Lusophone Linguistics* 2: 77–115. [CrossRef]
Simonet, Miquel. 2011. Intonational convergence in language contact: Utterance-final F0 contours in Catalan-Spanish early bilinguals. *Journal of the International Phonetic Association* 41: 157–84. [CrossRef]
van Rijswijsk, Remy, and Antje Muntendam. 2014. The prosody of focus in the Spanish of Quechua-Spanish bilinguals: A case study on noun phrases. *International Journal of Bilingualism* 18: 614–32. [CrossRef]
Vicenik, Chad, and Sun-Ah Jun. 2014. An auto-segmental metrical analysis of intonation in Georgian. In *Prosodic Typology II: The Phonology of Intonation and Phrasing*. Edited by Sun-Ah Jun. Oxford: Oxford Academic, pp. 154–86.

*languages*

MDPI

*Article*

# Phonation Patterns in Spanish Vowels: Spectral and Spectrographic Analysis

**Carolina González *, Susan L. Cox and Gabrielle R. Isgar**

Department of Modern Languages and Linguistics, Florida State University, Tallahassee, FL 32306, USA; slcox@fsu.edu (S.L.C.); gisgar@fsu.edu (G.R.I.)
* Correspondence: cgonzalez3@fsu.edu

**Abstract:** This article provides a detailed examination of voice quality in word-final vowels in Spanish. The experimental task involved the pronunciation of words in two prosodic contexts by native Spanish speakers from diverse dialects. A total of 400 vowels (10 participants × 10 words × 2 contexts × 2 repetitions) were analyzed acoustically in Praat. Waveforms and spectrograms were inspected visually for voice, creak, breathy voice, and devoicing cues. In addition, the relative amplitude difference between the first two harmonics (H1–H2) was obtained via FFT spectra. The findings reveal that while creaky voice is pervasive, breathy voice is also common, and devoicing occurs in 11% of tokens. We identify multiple phonation types (up to three) within the same vowel, of which modal voice followed by breathy voice was the most common combination. While creaky voice was more frequent overall for males, modal voice tended to be more common in females. In addition, creaky voice was significantly more common at the end of higher prosodic constituents. The analysis of spectral tilt shows that H1–H2 clearly distinguishes breathy voice from modal voice in both males and females, while H1–H2 values consistently discriminate creaky and modal voice in male participants only.

**Keywords:** vowels; phonation; modal; non-modal; creaky voice; breathy voice; devoicing; H1–H2

## 1. Introduction

This article investigates the occurrence of non-modal voicing in word-final vowels in Spanish. Modal voicing is characterized by periodic vibration of the vocal folds (Bissiri et al. 2011; Garellek 2014), while non-modal voicing involves non-periodicity and/or a degree of noise (Laver 1980; Esling et al. 2019). Modal voicing is commonly referred to in the literature as "voicing", while non-modal voicing encompasses phonatory qualities such as voicelessness, creaky voice and breathy voice. Voicelessness is defined by a lack of vocal fold vibration, while creaky voice (or creak) involves a glottal constriction, a low rate of vocal fold vibration (i.e., low pitch), and/or irregular F0 (Ladefoged 1971; Gordon and Ladefoged 2001; Garellek 2019, among others). Breathy voice involves voicing in addition to noise, and concentration of acoustic energy in the F3 region (Laver 1980; Keating et al. 2015; Garellek 2014, 2019; Esling et al. 2019). These phonation qualities relate to the relative degree of the vocal fold aperture, which is most open for voicelessness, then breathy voice, and most constricted for modal voice, then creaky voice (see, for example, Gordon and Ladefoged 2001).

In Spanish, as in many other languages, vowels are typically voiced, but non-modal realizations are attested in specific phonological environments and dialects. For example, vowel sequences across words (as in *la uva* 'the grape') can have creaky voice (Lorenzo Criado 1948; Lope Blanch 1987; González and Weissglass 2017). This is often the case when Spanish is in a contact situation with a language that has phonological glottalization, such as Yucatec Maya, Guaraní, or Arabic (Thon 1989; Valentín-Márquez 2006; Chappell 2013; Michnowicz and Kagan 2016; McKinnon 2018; Mohamed et al. 2019; Gynan and

Almada 2020). Vowels can also have creaky voice word-finally in various dialects, including Peninsular, Chilean, and Mexican Spanish (Morrison and Escudero 2007; Garellek and Keating 2015; Bolyanatz 2023; cf. Kim 2017). Fewer studies focusing on breathy voice are available for Spanish. Mendoza et al. (1996) and Trittin and Lleó (1995) find that Spanish-speaking females have breathier voices than males. This is also the case in English and other languages, and it is considered to result from anatomical differences in males and females: the larynx in females tends to have a longer opening, particularly posteriorly, which is conducive to aspiration noise in the F3 region (Klatt and Klatt 1990). Breathy vowels are also reported to occur utterance-finally in Andalusian Spanish (O'Neill 2005). As for voiceless vowels, they are also attested word- and utterance-finally in Spanish, often when preceded (or followed) by voiceless consonants (Delforge 2009, 2012; Torreira and Ernestus 2011; Sessarego 2012; Dabkowski 2018).

In a previous study, González et al. (2022) examined the occurrence of creaky voice in word-final Spanish vowels. Creaky voice was analyzed via visual inspection of waveforms and spectrograms, following Dilley et al. (1996), Redi and Shattuck-Hufnagel (2001) and Ladefoged and Maddieson (1996). Specifically, creaky voice was coded when one or more of the following acoustic cues was present: aperiodicity (non-regular duration of pulses), creak (gradual pulse widening with F0 lowering and damping), diplophonia (variable pulse amplitude or shape), glottal squeak (sudden F0 increase), or the presence of a glottal stop. González et al. (2022) found a prevalence of creaky voice word-finally for vowels across several Spanish dialects, particularly for men, and at the end of higher prosodic constituents. The data were originally examined for creaky voice only, but further inspection showed that at least some of the vowels coded as creaky tended to end in a noisy and/or non-periodic interval which could be consistent with breathy voice and/or devoicing.

In this study, we re-examine our original dataset to provide a more nuanced analysis of non-modal phonation in word-final Spanish vowels, focusing not just on creaky voice, but also on breathy voice and devoicing. We provide a more detailed analysis of non-modal phonation using waveforms and spectrograms, and, unlike in González et al. (2022), we additionally include a measure of spectral tilt, i.e., the degree of energy present in lower vs. higher frequencies. We focus in particular on H1–H2 values, i.e., the relative amplitude of the first harmonic (H1) (corresponding to the fundamental frequency or F0) compared to the second harmonic (H2), since previous studies show that H1–H2 correlates well with differences between modal voice, breathy voice, and creaky voice (see, for example, Klatt and Klatt 1990; Hillenbrand et al. 1994; Trittin and Lleó 1995; Gordon and Ladefoged 2001; Kreiman and Gerratt 2012; Keating et al. 2015; Kim 2017; Garellek 2019). Specifically, breathy voice has a higher spectral tilt (and, therefore, a higher H1–H2 amplitude) than modal voice, and modal voice has in turn a higher spectral tilt than creaky voice. These differences spring from open quotient differences related to phonation. Creaky voice, for example, has a low open quotient since it involves glottal constriction and increased medial vocal fold thickness. Open quotient is high for modal voice, and highest for breathy voice, since the vocal folds have a wider opening in the latter.

This study focuses on two research questions: (1) what is the prevalence of non-modal phonation beyond creaky voice in word-final vowels, and (2) does the distribution of phonation type differ in terms of speaker sex (i.e., male vs. female) or prosodic context (i.e., at the end of full or intermediate intonational phrases). Regarding (1), we expect to find examples of both breathy voice and devoicing in addition to creaky voice in our dataset. For (2), we hypothesize that non-modal phonation will be more prevalent at the end of full intonational phrases (IPs) than at the end of intermediate ones (ips), and that creaky voice will be more prevalent for males and breathy voice in females.

The remainder of this paper is organized as follows. Section 2 outlines the methodology of the study; experimental findings are presented in Section 3. Section 4 provides a discussion, and Section Section 5 closes with concluding remarks.

## 2. Methodology

This study is part of a larger project and builds on the analysis of creaky voice reported in González et al. (2022). The participants were 10 native Spanish speakers from a range of different dialects (Argentina, Bolivia, Colombia (2), Cuba (2), Peru, Puerto Rico, Spain, Venezuela). The participants from Spain, Argentina, Peru, and one from Colombia were male; the rest were female. All were 20–39 years old at the time of recording and had spent between 0 and 14 years in the US. All were raised in Spanish-speaking countries and spoke Spanish daily in their personal and professional life.

The study was approved by the IRB board of Florida State University. Participant data were collected in the phonetics laboratory after obtaining written informed consent. A digital recorder with a high-quality cardioid condenser microphone and a presence boost adapter was used to record audio data. Recordings were obtained in .wav format, in mono, with a sampling rate of 44,100 Hz.

This phase of the project involved a picture identification task comprising 12 images that participants had to name and frame in a short sentence. For example, when participants were shown a picture of a window, they were asked to say *Ventana. Es una Ventana*. ('Window. It's a window.'). A short training phase preceded the task. The picture identification task involved 10 token words and two distractors, one at the beginning and one at the end, to avoid list intonation effects. The task was conducted twice per participant.

Stimuli had penultimate stress, were two or three syllables long, and ended in /a/ or /o/ (Table 1). The final vowel of each token word occurred at the end of two different prosodic contexts: (i) a lower prosodic constituent, corresponding to an intermediate intonational phrase (ip) in the Spanish Tones and Breaks Indices framework (Sp_ToBI; see Beckman et al. 2002; Sosa 2003; Aguilar et al. 2009; Estebas-Vilaplana and Prieto 2009); and (ii) a higher prosodic constituent, corresponding to a full intonational phrase (IP) (1). The ends of full intonational phrases in our data involve a distinct pause coinciding with the end of the participant's turn and tend to be realized with a final low boundary tone (L%). In contrast, the ends of intermediate phrases are cued by a slight or no pause and often involve a rise (a high boundary tone H-) to indicate the message is continuing (Frota et al. 2007; Aguilar et al. 2009; Baxter 2017).

**Table 1.** Stimuli.

| Ending in /a/ | | Ending in /o/ | |
|---|---|---|---|
| Luna | 'moon' | Toro | 'bull' |
| Mesa | 'table' | Libro | 'book' |
| Casa | 'house' | Dinero | 'money' |
| Ventana | 'window' | Camino | 'road' |
| Pregunta | 'question' | Círculo | 'circle' |

| | | | | | |
|---|---|---|---|---|---|
| (1) | a. | [Ventana.]$_{ip}$ | [Es | una | ventana.]$_{IP}$ |
| | | Window. | (It) is | a | window. |

A total of 400 words were analyzed (10 participants × 10 tokens × 2 prosodic contexts × 2 repetitions) using Praat (Boersma and Weenink 2023). All measurements were taken by hand. Acoustic analysis involved inspection of the waveform and spectrogram for cues of modal and non-modal voicing. As in González et al. (2022), modal voice was characterized by periodicity, while creaky voice was characterized by one or more of the following: (i) irregular F0 ('aperiodicity'), (ii) F0 lowering, (iii) changes in pulse amplitude or shape ('diplophonia'), and/or (iv) presence of silence followed by a stop burst ('glottal stop') (Dilley et al. 1996; Docherty and Foulkes 2005; Gordon and Ladefoged 2001; Huber 1988; Keating et al. 2015; Ladefoged and Maddieson 1996; Redi and Shattuck-Hufnagel 2001). All token vowels were also examined for intervals involving lack of voicing—which were coded as devoiced—and for intervals involving noise in the F3 region, coded as breathy voice (Laver 1980; Keating et al. 2015; Garellek 2014, 2019; Esling et al. 2019). H2-H1 measurements were also taken via FFT spectra. Afterwards, token vowels were analyzed and divided into phonation intervals of modal voicing, creaky voice, breathy voice, and voicelessness based on visual cues from the spectrogram and waveform. FFT spectra were generated at the middle of each phonation interval to measure the relative amplitude of the first harmonic (H1) compared to the second harmonic (H2). As indicated in Section 1, H1–H2 is highest for breathy voice and lowest for creaky voice (Garellek 2019).

Figures 1–3 provide examples of waveforms, spectrograms, and FFT spectra for vowels fully realized as breathy, modal, and creaky (see Figures 4 and 5 below for examples of vowels involving more than one phonation interval). As shown in Figure 1b, breathy voice is characterized by a much higher amplitude of the first harmonic compared to the second. For modal voice, the first harmonic is higher in amplitude than the second (Figure 2b), but not as much as with breathy voice. On the other hand, creaky voice is characterized by a higher amplitude of the second harmonic compared to the first (Figure 3a).

All coding and measurements were checked by at least two of the authors. Note that we chose to conduct the spectral analysis of phonation intervals rather than full vowels or syllables, unlike in previous studies focusing on Spanish (see, for example, Trittin and Lleó 1995; Mendoza et al. 1996; Kim 2017). This methodological approach allows for a fine-grained exploration of voice quality, capturing cases where vowels have two or more phonation type sequences or 'dynamic combinations of non-modal phonations' in the words of Esposito and Khan (2020, p. 2) (for example, vowels that begin as creaky but end as breathy, as shown in Figures 4 and 5 below; see also Ladefoged 1983; DiCanio 2009).

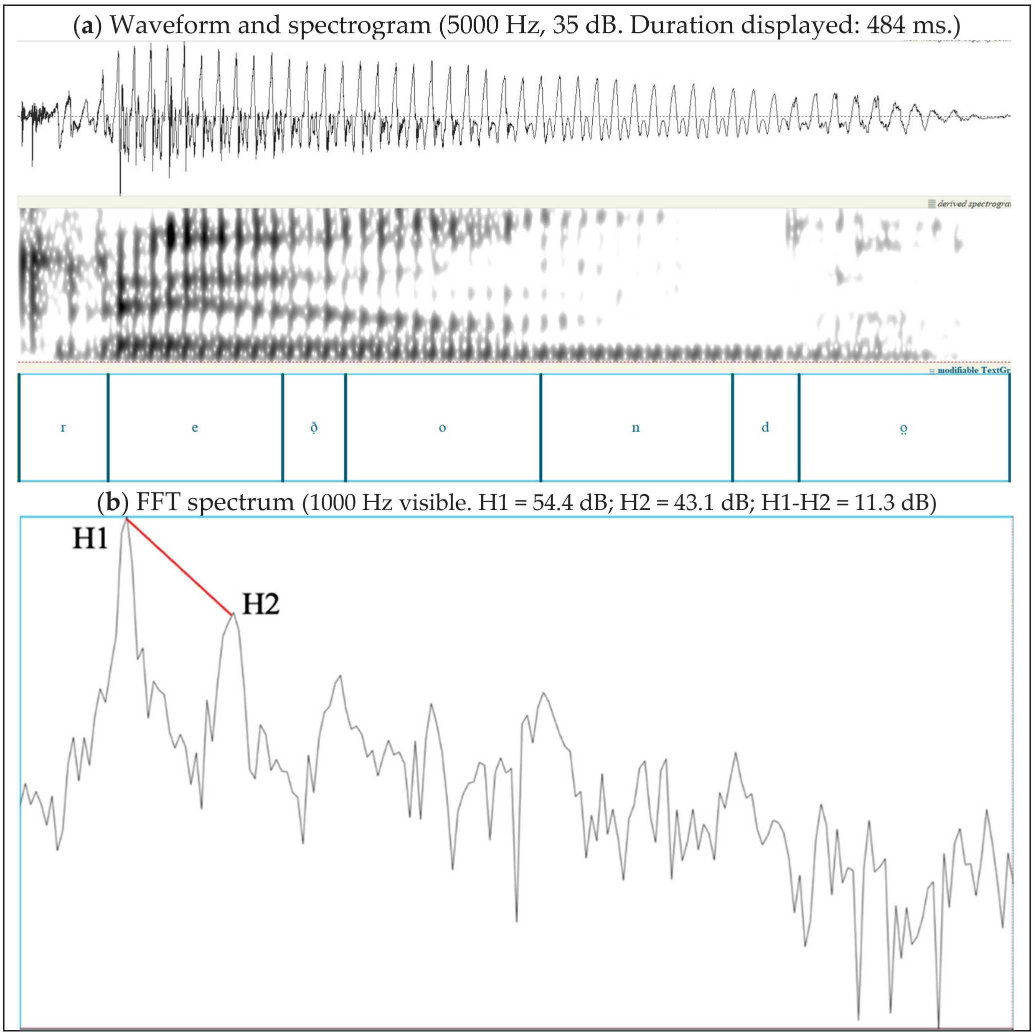



**Figure 1.** Breathy voice in word-final /o/ in *redondo* 'round', male speaker (P2).

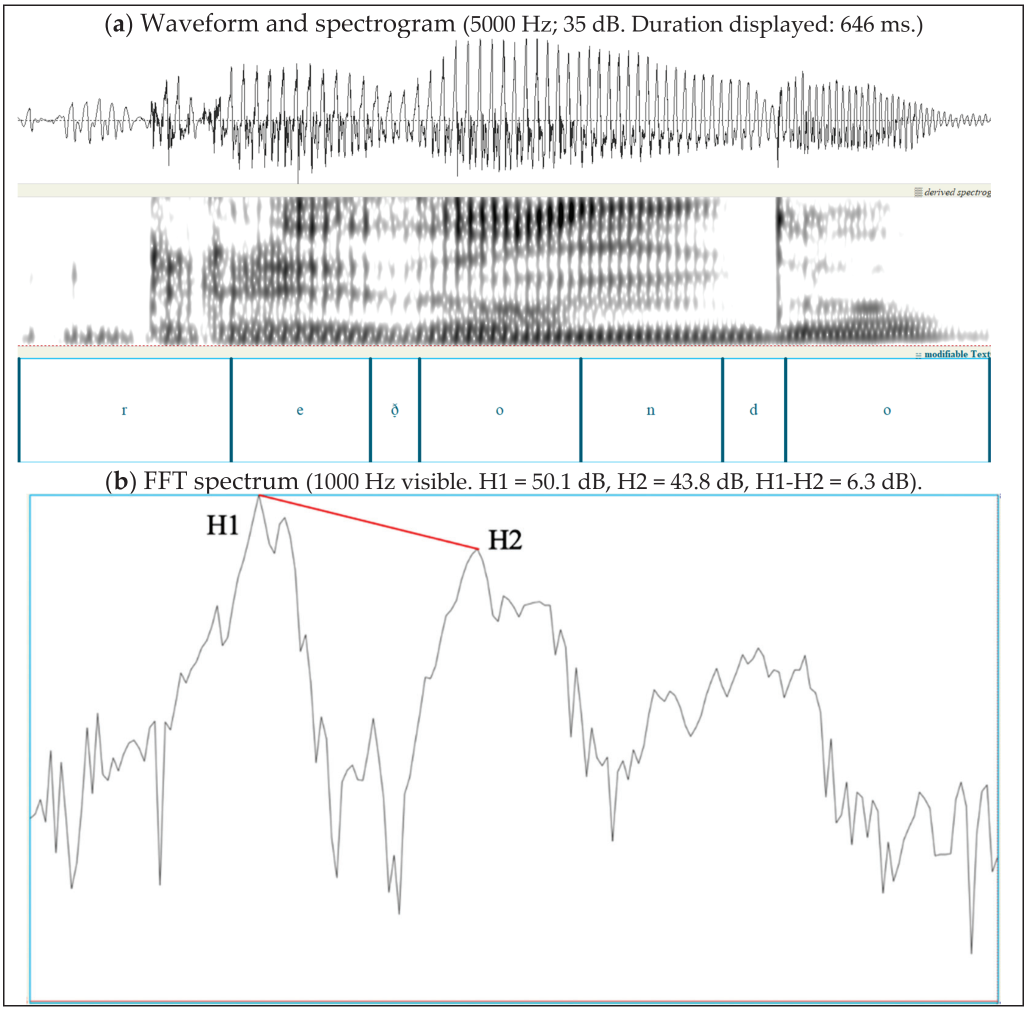



**Figure 2.** Modal voice in word-final /o/ in *redondo* 'round', male speaker (P2).

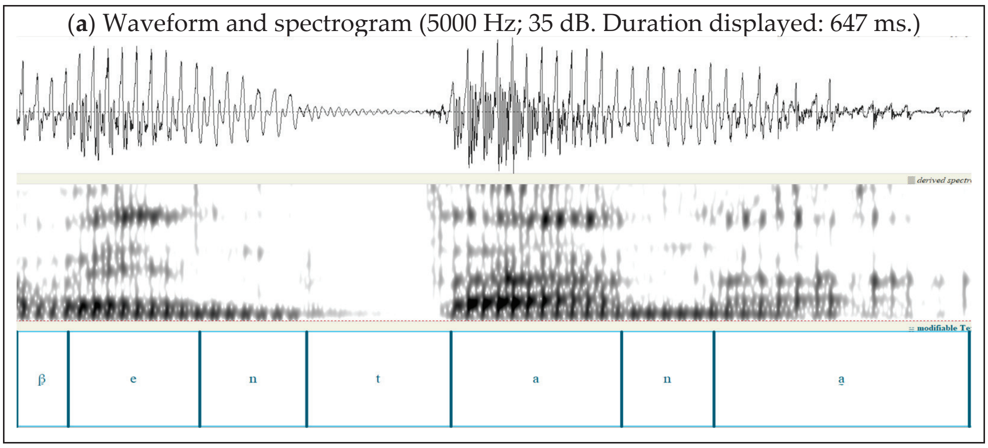



**Figure 3.** *Cont.*

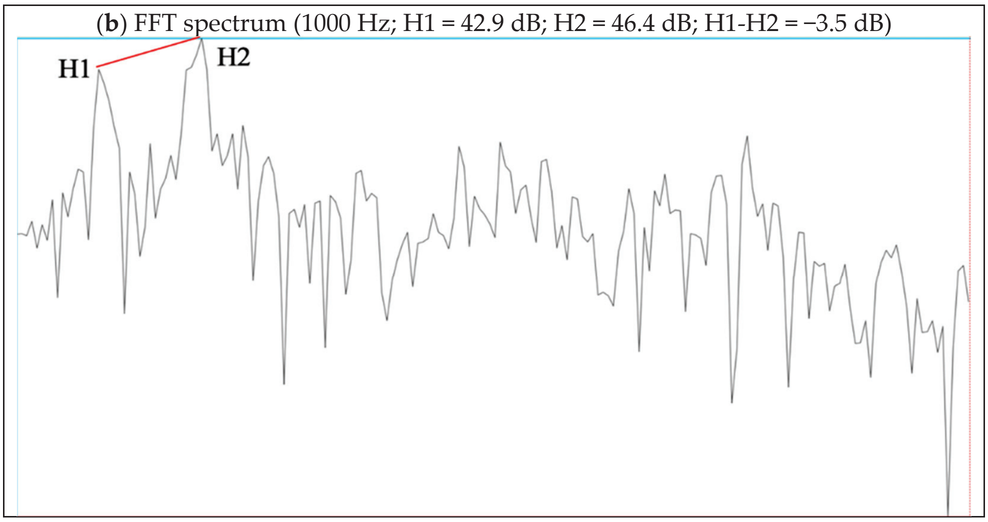



**Figure 3.** Vowel with full creak: *ventana* 'window', male speaker (P7).

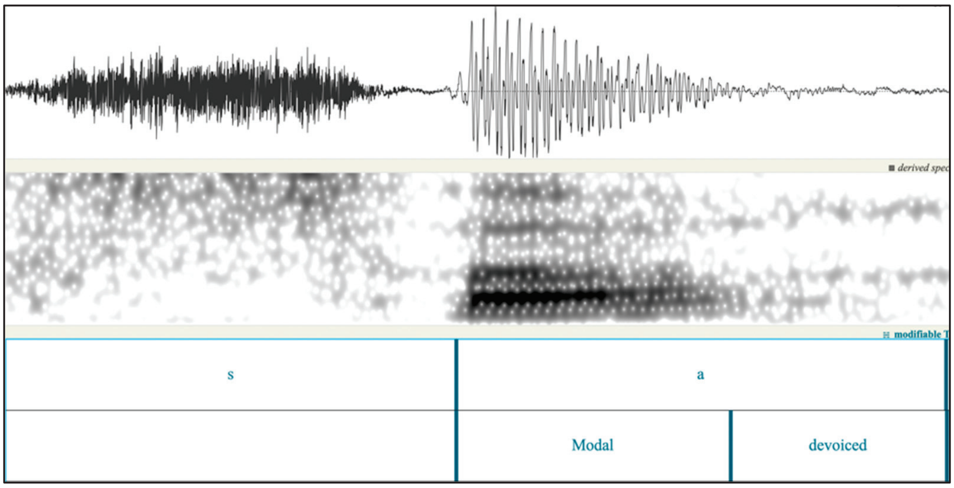



**Figure 4.** Example of double phonation. *Casa* 'house'. P5 (female), IP context. 5000 Hz, 40 dB.

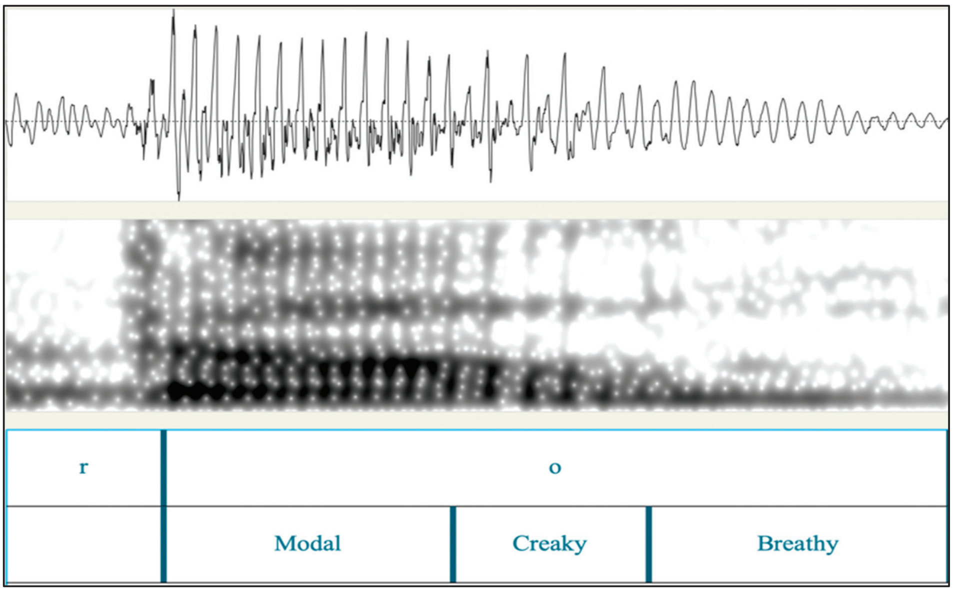



**Figure 5.** Example of triple phonation. Toro 'bull'. P3 (male), ip context. 5000 Hz, 40 dB.

For the analysis of phonation type distribution in word-final vowels, 18 tokens were discarded since they involved errors (4 tokens) or were produced as non-final in the ip context (14 tokens). Descriptive statistics were produced for 382 vowels. When determining if phonation type was associated with prosodic context and/or speaker sex, the 382 vowels were partitioned into 685 phonation intervals coded as modal, creaky, breathy, or devoiced. The analysis of H1–H2 did not include voiceless intervals ($n$ = 72) since harmonics are only present in periodic sounds. In addition, any phonation intervals occupying less than 30% of the vowel duration ($n$ = 64) were also discarded. The reason for this exclusion is two-fold: we were interested in obtaining clear H1–H2 cut-off points among modal, creaky, and breathy voice, which would be facilitated by longer analysis windows in FFT spectra. In addition, the 30% threshold has been used in other studies focusing on the analysis of creaky voice (Bolyanatz 2023), based on the fact that at least 30% of a vowel needs to have creak to be perceived as such utterance-finally (Crowhurst 2018). This resulted in the H1–H2 analysis of 549 phonation intervals (females, $n$ = 354; males, $n$ = 195).

Statistical analyses were performed using IBM SPSS Statistics (Version 29) (IBM Corp. 2022) and R (Version 4.3.1) (R Core Team 2023); these included ANOVAs, chi-squared tests, and Bonferroni post hoc tests. Statistical analyses were considered significant if $p \leq 0.05$.

## 3. Results

### *3.1. Non-Modal Phonation in Word-Final Vowels: Overall Findings*

As shown in Table 2, the most common type of non-modal phonation attested in word-final vowels in our database was creaky voice (occurring in 31% of vowels), followed closely by breathy voice (29%). Devoicing was found in 11% of cases. Creaky voice, when present, tended to encompass a larger portion of the vowel duration compared to breathy voice and devoicing (66%, 52%, and 36%, respectively).

**Table 2.** Percentage of occurrence and average relative duration by phonation type.

| Phonation Type | *n* | % | Vowel Duration Ratio (%) | |
|---|---|---|---|---|
| | | | Mean | *SD* |
| Non-Modal | 489 | 71% | 51% | 21% |
| Creaky | 215 | 31% | 66% | 28% |
| Breathy | 201 | 29% | 52% | 23% |
| Devoiced | 72 | 11% | 36% | 13% |
| Modal | 197 | 29% | 56% | 21% |
| Total | 685 | 100% | 53% | 22% |

Table 3 shows the relative occurrence of phonation combinations for all participants pooled. Overall, vowels with double phonation are the most common (58%), followed by vowels pronounced with single phonation (31%). Vowels with triple phonation are attested only in 11% of cases. The two most common phonation realizations are modal vowels ending in breathy voice (23%) and fully creaky vowels (19%). Creaky/breathy voiced vowels and modal/creaky voiced vowels are also relatively common (11% and 9%, respectively). Vowels including three different types of phonation mostly begin with modal voice, although some instances involving initial creak are also attested.

Most cases of vowels pronounced with double phonation end in breathy voice (34%), followed by devoicing (15%). In contrast, only 9% of vowels produced with double phonation end in creaky voice. For vowels involving triple phonation, most end in devoicing (5%) or breathy voice (5%). In contrast, only 1% of vowels end in creaky voice. Examples of vowels with double and triple phonation ending in breathy voice or devoicing are given in Figures 4 and 5.

Some positional generalizations for non-modal phonation were observed. Fully devoiced vowels were not attested, and devoicing did not precede any other phonation type. In vowels pronounced with double phonation, modal voice could only precede another phonation type, not follow it; and if a vowel began as breathy voiced, it could only be followed by devoicing. Finally, all cases of triple phonation involved modal or creaky voice followed by other phonation types.

**Table 3.** Distribution of single, double, and triple phonation types.

| Single | *n* | % | Double | *n* | % | Triple | *n* | % |
|---|---|---|---|---|---|---|---|---|
| Creaky | 75 | 19% | Modal/Breathy | 87 | 23% | Modal/Creaky/Breathy | 16 | 4% |
| Breathy | 27 | 7% | Creaky/Breathy | 42 | 11% | Modal/Breathy/Devoiced | 10 | 3% |
| Modal | 19 | 5% | Modal/Creaky | 35 | 9% | Modal/Creaky/Devoiced | 2 | 0.5% |
| Devoiced | 0 | 0% | Creaky/Devoiced | 29 | 8% | Modal/Breathy/Creaky | 2 | 0.5% |
| | | | Modal/Devoiced | 19 | 5% | Creaky/Breathy/Devoiced | 5 | 1% |
| | | | Breathy/Devoiced | 7 | 2% | Creaky/Modal/Creaky | 2 | 0.5% |
| | | | | | | Creaky/Modal/Breathy | 4 | 1% |
| | | | | | | Creaky/Modal/Devoiced | 1 | 0.5% |
| Total | 121 | 31% | Total | 219 | 58% | Total | 42 | 11% |

### *3.2. Speaker Sex and Prosodic Context*

#### 3.2.1. Phonation Type and Speaker Sex

Figure 6 and Table 4 illustrate the frequency of occurrence of phonation type for male and female speakers. Results from a Pearson's two-sided chi-squared test of independence show a significant association between phonation type and sex ($\chi^2$ (3, 685) = 15.559, $p = 0.001$). Males tend to favor creaky voice, while females show only a slight preference for modal voice over other phonation types. The frequency of occurrence of breathy voice and devoicing are comparable in both groups.

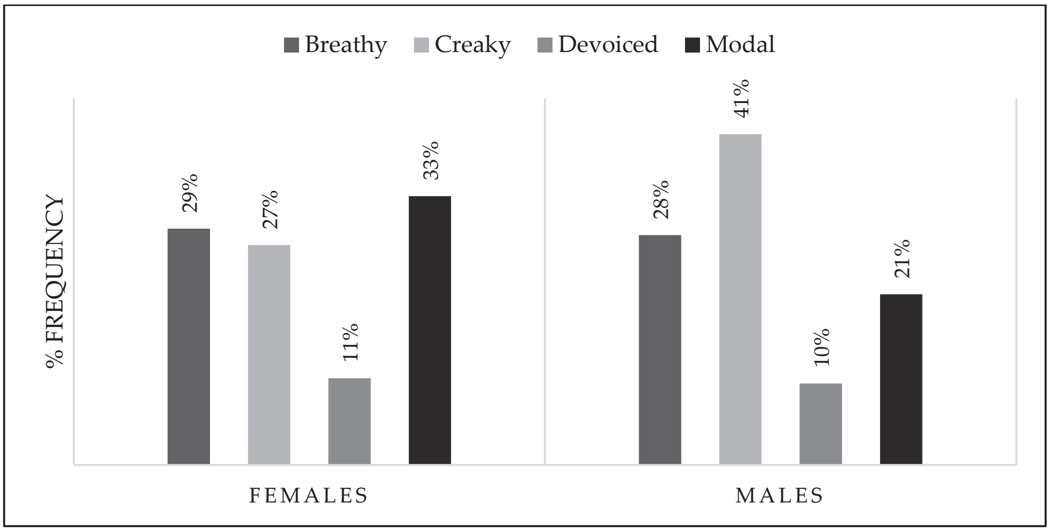



**Figure 6.** Phonation type frequency by speaker sex.

**Table 4.** Descriptive statistics for phonation type by speaker sex.

| | Females | | Males | | Total | |
|---|---|---|---|---|---|---|
| | *n* | % | *n* | % | *N* | % |
| Breathy | 135 | 29% | 66 | 28% | 201 | 29% |
| Creaky | 120 | 27% | 95 | 41% | 215 | 31% |
| Devoiced | 48 | 11% | 24 | 10% | 72 | 11% |
| Modal | 148 | 33% | 49 | 21% | 197 | 29% |
| Total | 451 | 100% | 234 | 100% | 685 | 100% |

3.2.2. Phonation Type and Prosodic Context

Figure 7 and Table 5 illustrate the frequency of occurrence of phonation type at the end of intermediate and full intonation phrases (ips and IPs, respectively). Results from a Person's two-sided chi-squared test of independence show that there is a significant association between phonation type and prosodic context ($\chi^2$ (3, 685) = 17.342, $p < 0.001$). Modal voice was more prevalent than other phonation types at the end of intermediate phrases (ips), while creaky voice and breathy voice were the most frequent phonation types attested at the end of full intonational phrases (IPs). The frequency of occurrence of devoicing was comparable in both contexts.

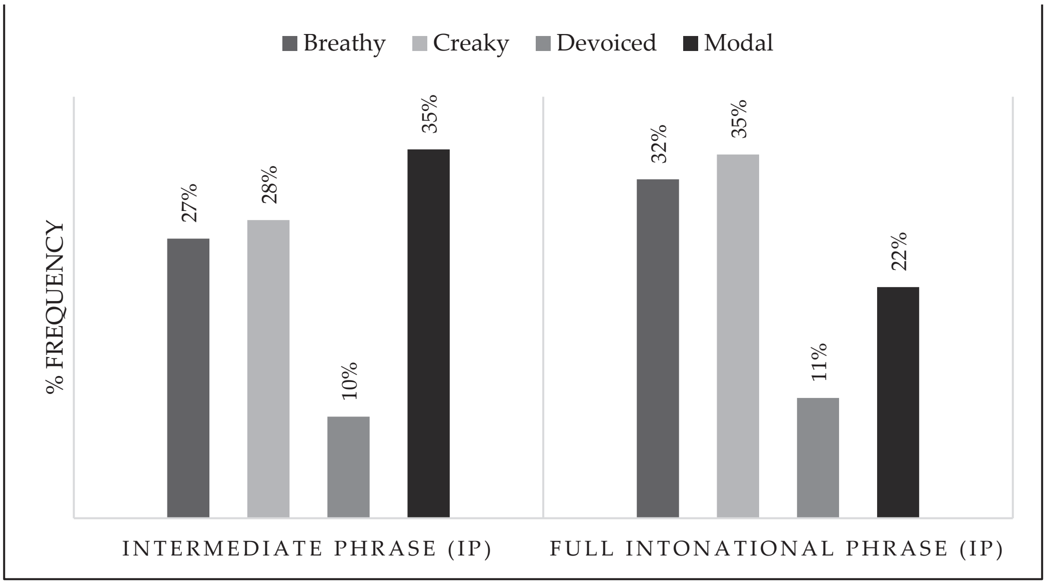



**Figure 7.** Percentages of frequency of phonation type by prosodic context.

**Table 5.** Descriptive statistics for phonation type by prosodic context.

| | ip | | IP | | Total | |
|---|---|---|---|---|---|---|
| | *n* | % | *n* | % | *N* | % |
| Breathy | 91 | 27% | 110 | 32% | 201 | 29% |
| Creaky | 97 | 28% | 118 | 35% | 215 | 31% |
| Devoiced | 33 | 10% | 39 | 11% | 72 | 11% |
| Modal | 122 | 35% | 75 | 22% | 197 | 29% |
| Total | 343 | 100% | 342 | 100% | 685 | 100% |

As shown in Table 6, the two most common phonation types attested ip-finally for female participants are modal vowels ending in breathy voice (46%) and modal vowels ending in creaky voice (13%). IP-finally, fully creaky vowels, modal/breathy vowels, and creaky/breathy vowels are the most common phonation realizations (25%, 22%, and 19%, respectively). Other phonation type combinations occur in less than 10% of cases and are for the most part comparable across prosodic contexts.

For male participants, fully creaky vowels were the most common vowel type in both prosodic contexts (Table 7). ip-finally, modal vowels, modal/creaky, and creaky/devoiced vowels were also relatively common (13%, 16% and 13%, respectively). IP-finally, the second most common phonation type was breathy voice (24%).

**Table 6.** Distribution of phonation types for females across prosodic contexts.

| Phonation Type | ip | | IP | | Total | |
|---|---|---|---|---|---|---|
| | *n* | % | *n* | % | *N* | % |
| Modal voice | 4 | 4% | 3 | 3% | 7 | 3% |
| Creaky voice | 6 | 5% | 30 | 25% | 36 | 16% |
| Breathy voice | 0 | 0% | 1 | 1% | 1 | 0% |
| Modal/Creak | 15 | 13% | 3 | 3% | 18 | 8% |
| Modal/Breathy | 51 | 46% | 26 | 22% | 77 | 33% |
| Modal/Devoiced | 9 | 8% | 9 | 8% | 18 | 8% |
| Creaky/Breathy | 6 | 5% | 22 | 19% | 28 | 12% |
| Creaky/Devoiced | 7 | 6% | 5 | 4% | 12 | 5% |
| Modal/Creaky/Devoiced | 1 | 1% | 1 | 1% | 2 | 1% |
| Modal/Breathy/Devoiced | 4 | 4% | 6 | 5% | 10 | 4% |
| Modal/Creaky/Breathy | 5 | 4% | 2 | 2% | 7 | 3% |
| Modal/Breathy/Creaky | 2 | 2% | 0 | 0% | 2 | 1% |
| Creaky/Breathy/Devoiced | 0 | 0% | 5 | 4% | 5 | 2% |
| Creaky/Modal/Creaky | 2 | 2% | 0 | 0% | 2 | 1% |
| Creaky/Modal/Devoicing | 0 | 0% | 1 | 1% | 1 | 0% |
| Creaky/Modal/Breathy | 0 | 0% | 4 | 3% | 4 | 2% |
| Total | 112 | 100% | 118 | 100% | 230 | 100% |

**Table 7.** Distribution of phonation types for males across prosodic contexts.

| Phonation Type | ip | | IP | | Total | |
|---|---|---|---|---|---|---|
| | *n* | % | *n* | % | *N* | % |
| Modal voice | 10 | 13% | 2 | 3% | 12 | 8% |
| Creaky voice | 19 | 25% | 20 | 26% | 39 | 26% |
| Breathy voice | 8 | 11% | 18 | 24% | 26 | 17% |
| Modal/Creak | 12 | 16% | 5 | 7% | 17 | 11% |
| Modal/Breathy | 5 | 7% | 5 | 7% | 10 | 7% |
| Modal/Devoiced | 1 | 1% | 0 | 0% | 1 | 1% |
| Creaky/Breathy | 8 | 11% | 6 | 8% | 14 | 9% |
| Creaky/Devoiced | 10 | 13% | 6 | 8% | 16 | 11% |
| Breathy/Devoiced | 1 | 1% | 6 | 8% | 7 | 5% |
| Modal/Creaky/Breathy | 1 | 1% | 8 | 11% | 9 | 6% |
| Total | 75 | 100% | 76 | 100% | 151 | 100% |

There was a significant interaction between sex and prosodic context for breathy voice ($\chi^2$ (1, 201) = 4.311, $p$ = 0.027) (Figure 8, Table 8). For males, breathy voice was more widespread at the end of full intonational phrases (IPs), unlike for females, where breathy voice occurred as frequently in both prosodic contexts.

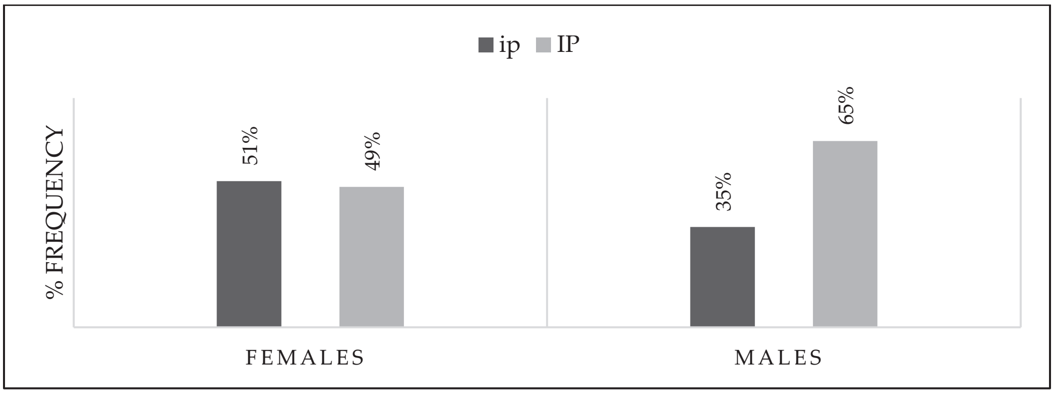



**Figure 8.** Breathy voice across prosodic contexts for both sexes.

**Table 8.** Descriptive statistics: breathy voice across prosodic contexts for both sexes.

| | **Females** | | **Males** | | **Total** | |
|---|---|---|---|---|---|---|
| | ***n*** | % | ***n*** | % | ***N*** | % |
| ip | 68 | 51% | 23 | 35% | 91 | 45% |
| IP | 67 | 49% | 43 | 65% | 110 | 55% |
| Total | 135 | 100% | 66 | 100% | 201 | 100% |

There was also a significant association between sex and prosodic context for creaky voice ($\chi^2$ (1, 215) = 3.882, $p$ = 0.033) (Figure 9, Table 9). For women, creaky voice was more widespread at the end of full intonational phrases (IPs). For men, creaky voice was only slightly more frequent at the end of intermediate phrases (ips) than full intonational phrases (IPs).

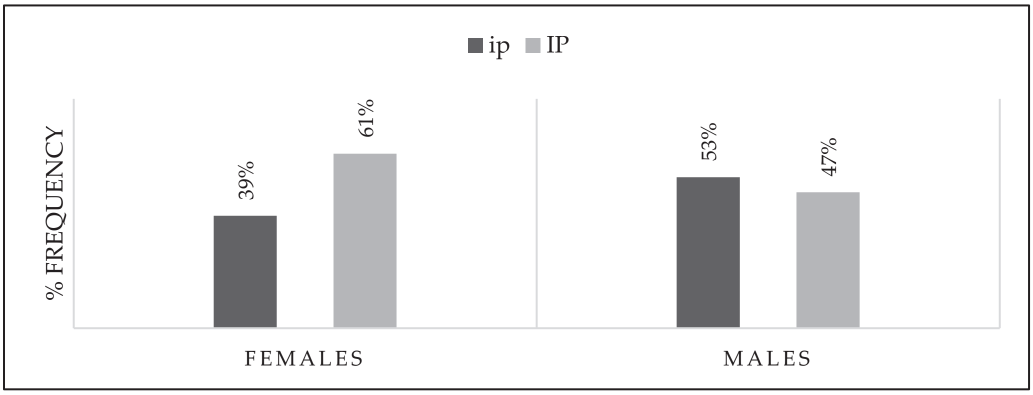



**Figure 9.** Creaky voice across prosodic contexts for both sexes.

**Table 9.** Descriptive statistics for creaky voice across prosodic contexts for both sexes.

| | **Females** | | **Males** | | **Total** | |
|---|---|---|---|---|---|---|
| | ***n*** | % | ***n*** | % | ***N*** | % |
| ip | 47 | 39% | 50 | 53% | 97 | 45% |
| IP | 73 | 61% | 45 | 47% | 118 | 55% |
| Total | 120 | 100% | 95 | 100% | 215 | 100% |

For devoicing, no significant association was found between sex and prosodic context ($\chi^2$ (1, 72) = 0.252, $p$ = 0.4 (Figure 10, Table 10). For males, devoicing was equally as prevalent at the end of ips and IPs. For females, there was slighly more devoicing at the end of full intonational phrases.

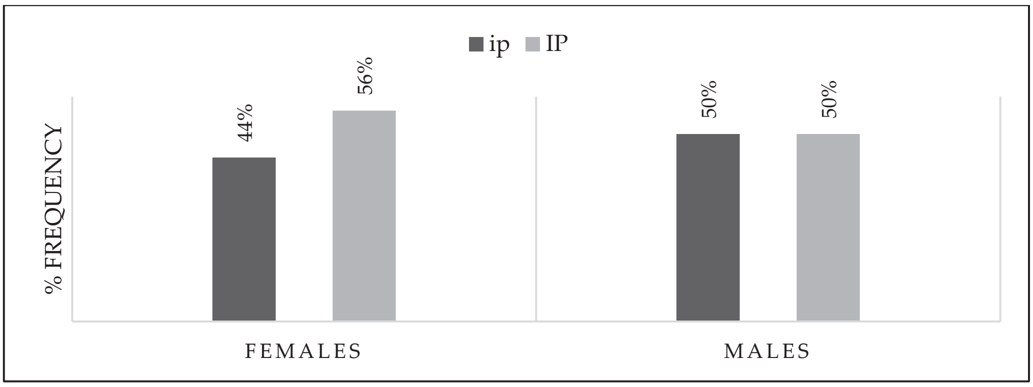



**Figure 10.** Devoicing across prosodic contexts for both sexes.

**Table 10.** Descriptive statistics for devoicing across prosodic contexts for both sexes.

| | Females | | Males | | Total | |
|---|---|---|---|---|---|---|
| | *n* | % | *n* | % | *N* | % |
| ip | 21 | 44% | 12 | 50% | 33 | 46% |
| IP | 27 | 56% | 12 | 50% | 39 | 54% |
| Total | 48 | 100% | 24 | 100% | 72 | 100% |

### 3.3. Visual Cues and Spectral Tilt

Figure 11 and Table 11 shows average H1–H2 differences in phonation types for both sexes. As expected, H1–H2 values are higher for breathy voice. For males, H1–H2 values are higher for modal voice than creaky voice. However, H1–H2 values are very similar for modal voice and creaky voice intervals for females.

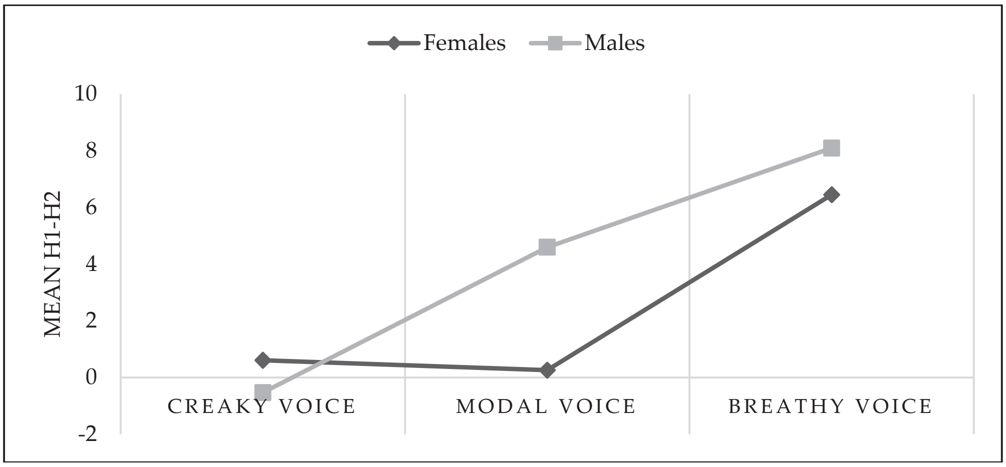



**Figure 11.** Means plots for H1–H2 in creaky, modal, and breathy voice intervals in males and females.

**Table 11.** Descriptive statistics: H1–H2 values for creaky, modal, and breathy voice.

| | Females (N = 354) | | | | Males (N = 195) | | | | Total | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *n* | % | Mean | *SD* | *n* | % | Mean | *SD* | *N* | % |
| Creaky | 103 | 29% | 0.07 | 6.23 | 87 | 45% | 0.07 | 5.67 | 190 | 35% |
| Modal | 135 | 38% | −0.84 | 8.06 | 45 | 23% | 4.56 | 4.16 | 180 | 33% |
| Breathy | 116 | 33% | 5.8 | 9.17 | 63 | 32% | 7.67 | 6.97 | 179 | 32% |
| Total | 354 | 100% | | | 195 | 100% | | | 549 | 100% |

The results from a one-way ANOVA show a significant difference for H1–H2 and phonation type for female participants (df = 2, N = 331; F = 28.776, $p < 0.001$). The results from a Bonferroni post hoc analysis with multiple comparisons shows a significant difference between breathy voice and modal voice and between breathy voice and creaky voice; the difference between modal voice and creaky voice is not significant (Table 12).

**Table 12.** Multiple comparisons, Bonferroni post hoc test of phonation and H1–H2 for females.

| | Modal Voice | | | Creaky Voice | | | Breathy Voice | | |
|---|---|---|---|---|---|---|---|---|---|
| | *p*-Value | Mean Diff. | *SE* | *p*-Value | Mean Diff. | *SE* | *p*-Value | Mean Diff. | *SE* |
| Modal voice | | | | 1 | −0.35 | 0.93 | <0.001 | −6.2 | −0.90 |
| Creaky voice | 1 | 0.35 | 0.93 | | | | <0.001 | −5.84 | 0.95 |
| Breathy voice | <0.001 | 6.2 | 0.90 | <0.001 | 5.84 | 0.95 | | | |

For males, the results from a one-way ANOVA also show a significant difference for H1–H2 values across phonation type (df = 2, N = 188; F = 52.669, $p < 0.001$). The results from Bonferroni post hoc tests with multiple comparisons show that modal voice, breathy voice, and creaky voice are all significantly different from each other (Table 13).

**Table 13.** Multiple comparisons, Bonferroni post hoc test of phonation and H1–H2 for males.

| | **Modal Voice** | | | **Creaky Voice** | | | **Breathy Voice** | | |
|---|---|---|---|---|---|---|---|---|---|
| | ***p*-Value** | **Mean Diff.** | ***SE*** | ***p*-Value** | **Mean Diff.** | ***SE*** | ***p*-Value** | **Mean Diff.** | ***SE*** |
| Modal voice | | | | <0.001 | -5.1 | 0.93 | 0.002 | 3.5 | 0.99 |
| Creaky voice | <0.001 | 5.1 | 0.92 | | | | <0.001 | 8.6 | 0.85 |
| Breathy voice | 0.002 | −3.5 | 0.99 | <0.001 | −8.6 | 0.85 | | | |

## 4. Discussion

Our study re-examines the dataset in González et al. (2022) to investigate the occurrence of non-modal voice quality beyond creaky voice at the end of prosodic constituents in Spanish, an area that is generally understudied for Spanish. Our results show that both creaky voice and breathy voice occur frequently at the end of ips and IPs; devoicing, while attested, is much less common.

Unlike previous studies, we move beyond considering one type of non-modal phonation only (as in Kim 2017 or González et al. 2022 for creaky voice; or Mendoza et al. 1996; and Trittin and de Santos y Lleó 1995, which focus on breathy voice) and provide a fine-grained description and analysis of phonation combinations in word-final vowels in Spanish. We find that vowels with double phonation are the most widespread in our data, particularly those beginning with modal voice and ending in breathy voice, followed by vowels with single phonation, especially fully creaky vowels. In addition, 11% of the vowels in our dataset involved triple phonation. The latter tend to begin with modal voice, although some instances involving initial creaky voice are also attested.

Fully devoiced vowels are unattested in the contexts examined; in addition, devoicing is positionally restricted to the end of the vowel. Other languages with positional restrictions for non-modal phonation include Santa Ana del Valle Zapotec, where non-modal phonation occurs only vowel-finally (Esposito 2005), and White Hmong, where breathy voice can only occur vowel-initially (Keating et al. 2010). In our dataset, vowels with double or triple phonation overwhelmingly tend to end in breathy voice or devoicing. As suggested by one reviewer, it is possible that this relates to vocal fold spreading in anticipation of a pause or breath intake. The fact that dynamic phonation combinations are common in Spanish vowels, and that some are more frequent than others, is not only a novel finding, but might be helpful for the segmentation of Spanish vowels in future studies, particularly those focusing on word-final contexts.

Our findings reveal a significant effect of prosodic context on phonation type. Modal voice is more widespread at the end of intermediate phrases (ips) overall, while creaky and breathy voice are more common at the end of full intonational phrases (IPs). These results are in line with prior studies showing that non-modal phonation voice is a cue to the end of higher prosodic constituents across several languages. This is the case for devoicing and breathy voice in French (Smith 1999) and for creaky voice in English, Spanish, and Italian, among other languages (Keating et al. 2015; Dilley et al. 1996; González et al. 2022; Di Napoli 2015).

In addition, our results show a significant effect of sex on phonation type, with creaky voice being more common in males, and modal and breathy voice occurring more frequently in females. Similar results for creaky and modal voice were reported in González et al. (2022). Biologically, males tend to have longer and thicker vocal folds than females, resulting in a lower pitch/F0, which may be more conducive to creaky voice overall, particularly at the end of prosodic constituents when intonation falls. On the other hand, females' vocal folds tend to not close completely when vibrating, which can often result in breathy voice (Laver 1980). We find of particular interest the fact that while males favored creaky voice, they tended to use breathy voice more often at the end of IPs. On the other hand, while females tended to prefer modal or breathy voice, they significantly had more creaky voice at the end of IPs. These findings suggest that both creaky voice and breathy voice can cue the end of higher prosodic units in Spanish, depending on the sex of the speaker. It is interesting that, as shown in Table 2, among non-modal phonation creaky voice has the longest duration, followed by breathy voice, while devoicing tends to be very short. This might indicate that creaky voice is a more salient prosodic cue at the end of constituents in Spanish than breathy voice or devoicing.

Our study also investigates the relative alignment between spectral tilt, as measured by H1–H2, and visual phonation cues in waveforms and spectrograms. As expected, spectral tilt is highest for breathy voice, both for males and females (see, for example, Klatt and Klatt 1990; Garellek 2019). For males, spectral tilt is also higher in modal voice than in creaky voice, as expected, although the average H1–H2 value for creak is not necessarily negative, as reported in some of the prior literature. For females, however, H1–H2 values for modal and creaky voice are practically the same. As Esposito and Khan (2020, p. 8) point out, H1–H2 sometimes fails to successfully measure specific voice qualities. For example, in Marathi H1–H2 is a reliable acoustic cue of breathy voice for male speakers but not so consistent for female ones (Berkson 2012).

There are several possible reasons explaining the lack of complete alignment between spectral tilt and visual phonation cues for creaky and modal voice in the female participants in our dataset. These include (i) a wide range of individual variation in spectral tilt for Spanish females; (ii) the occurrence of nasalization in stimuli where the final vowel is preceded by a nasal segment (as in *ventana* 'window'); (iii) a possible effect of vowel quality on harmonic amplitude (Klatt and Klatt 1990; Garellek et al. 2016; Garellek 2019, 2022). We consider (iii) to be the most likely explanation: our study included tokens ending in /a/, which has a high F1 that does not largely impact voice quality, but also ending in /o/, usually involving a relatively low first formant that might have influenced the amplitudes of the first and second harmonics. Further studies investigating voice quality in Spanish could include the use of additional spectral tilt measures and/or use formant correction algorithms in the analysis of H1–H2 (Iseli et al. 2007; Esposito et al. 2021). In any case, we conclude that the identification of breathy voice in Spanish vowels benefits from using H1–H2 measurements, for both males and females. For creaky voice, however, visual examination of cues such as aperiodicity and diplophonia in waveforms and spectrograms might be enough to identify it consistently.

The participants in this study were Spanish-dominant bilingual speakers with English as their L2 who used Spanish daily in personal and professional interactions. Some had just arrived in the US, and some had been living in this country for up to 40 years. Cantor-Cutiva et al. (2023) reports that Spanish-dominant bilingual speakers do not produce creaky voice as often as English-dominant bilinguals, while Kim (2017) shows that Spanish–English bilinguals often transfer creaky voice from English into Spanish (Kim 2017). González et al. (2022), who examined creaky voice in the participants of the present study, argue against English transfer of this voice quality based on the fact that (i) creaky voice was as frequent in participants who had just moved to the US as in participants who had been residing in the US for many years, and (ii) creaky voice is more frequent for females than males in American English, unlike in the Spanish dataset, which shows the opposite trend. While the present study still shows this pattern, we find it intriguing that the female participants had significantly more creak at the end of IPs, while male participants preferred breathy voice in the same content. We consider then that transfer of creaky voice from English might be a possibility at the end of higher prosodic units, at least for females, but we leave this point for future investigation.

Finally, the participants of this study were from eight different Spanish-speaking countries. Except for the case of Colombian Spanish, all Spanish dialects represented included male or female speakers, but not both. Some dialectal differences in phonation are reported for Spanish; for example, Kim (2017) did not find creaky voice utterance-finally in native speakers of Mexican Spanish (unlike Garellek and Keating 2015). We leave a more detailed investigation of dialectal differences in the voice quality of Spanish vowels for future investigations.

## 5. Conclusions

This study contributes to the investigation of phonation in Spanish vowels. It shows that creaky voice and breathy voice are pervasive word-finally, particularly at the end of full intonational phrases (IPs). While creaky voice was favored by males and modal and breathy voice by females, at the end of full intonational phrases males tended to have more breathy voice, and females more creaky voice. Our findings also show that word-final Spanish vowels often involves double or triple phonation types. In vowels realized with multiple phonation types, modal voice tends to precede other voice qualities; devoicing, when present, always occurs at the end of the vowel. These results have implications for future acoustic studies of Spanish vowels, which are often assumed to be modally voiced throughout.

Our study included the examination of phonatory visual cues in waveforms and spectrograms in addition to the measurement of H1–H2 values. The latter was calculated for each phonation interval, rather than as the average for the entire vowel, resulting in a more fine-grained investigation of voice quality. To the best of our knowledge, this is the first time this methodology is employed, at least for Spanish. Our results show that H1–H2 differences between modal and breathy voice align well with phonatory visual cues in acoustic displays. However, H1–H2 values consistently distinguish modal and creaky voice for males only. Future studies on Spanish vowels should include formant correction algorithms for H1–H2 and/or the use of additional spectral tilt measures, particularly if focusing on female speakers. It is our hope that other scholars continue to investigate phonation in vowels, since it can inform our understanding of their acoustic and perceptual characteristics both in L1 and L2 Spanish.

**Author Contributions:** Conceptualization, all authors; methodology, all authors; writing—original draft preparation, C.G.; writing—review and editing, all authors; acoustic analysis, all authors; statistics, G.R.I.; visualization, S.L.C. and G.R.I.; supervision, C.G.; funding acquisition, C.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** This project was funded in part by a COFRS grant awarded to the first author by Florida State University in 2014 (award ID: 082003 551 032945 and agency: FSU Council of Research and Creativity).

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of Florida State University (protocol code HSC No. 2011.7371) for studies involving humans.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The datasets presented in this article are not readily available because of the conditions specified in the IRB approval for the project.

**Acknowledgments:** Our sincere thanks to all participants of the study, the linguistics faculty and students at the Department of Modern Languages and Linguistics at Florida State University for their support, and the audience at the 2023 International Conference on Experimental Phonetics/Congreso Internacional de Fonética Experimental (CIFE) held at the University of Vigo in Spain for their feedback and questions. We are also immensely grateful to the three anonymous reviewers of this article for their suggestions, which have greatly improved this article. All errors, of course, remain ours. Finally, we would like to acknowledge the assistance provided by ChatGPT, an AI language model developed by OpenAI, in guiding our searches and providing valuable insights on statistical analyses and R/SPSS procedures related to the research conducted in this article.

**Conflicts of Interest:** The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

Aguilar, Lourdes, Carme De-la-Mota, and Pilar Prieto. 2009. Sp_ToBI Training Materials. Available online: http://prosodia.upf.edu/sp_tobi/ (accessed on 22 February 2024).

Baxter, Robert Patrick. 2017. Co-Constructing Turns as Interactional Competence: Collaborative Turn Sequences in L1 and L2 Spanish. Ph.D. dissertation, Indiana University, Bloomington, IN, USA.

Beckman, Mary E., Manuel Díaz-Campos, Julia Tevis McGory, and Terrell A. Morgan. 2002. Intonation across Spanish, in the Tones and Break Indices framework. *Probus* 14: 9–36. [CrossRef]

Berkson, Kelly H. 2012. Acoustic correlates of breathy sonorants in Marathi. *Journal of Phonetics* 73: 70–90. [CrossRef]

Bissiri, Maria Paola, Maria Luisa Lecumberri, Martin Cooke, and Jan Volín. 2011. The role of word-initial glottal stops in recognizing English words. Paper presented at INTERSPEECH 2011, Florence, Italy, August 27–31; pp. 165–68.

Blanch, Lope. 1987. *Estudios sobre el español de Yucatán*. Mexico City: Universidad Nacional Autónoma de México.

Boersma, Paul, and David Weenink. 2023. Praat: Doing Phonetics by Computer (3.17.24) [Computer Software]. Available online: https://www.fon.hum.uva.nl/praat/ (accessed on 3 May 2023).

Bolyanatz, Mariška. 2023. Creaky Voice in Chilean Spanish: A Tool for Organizing Discourse and Invoking Alignment. *Languages* 8: 161. [CrossRef]

Cantor-Cutiva, L. C., P. Bottalico, J. Webster, C. Nudelman, and E. Hunter. 2023. The Effect of Bilingualism on Production and Perception of Vocal Fry. *Journal of Voice* 37: 970. [CrossRef] [PubMed]

Chappell, Whitney. 2013. Social and Linguistic Factors Conditioning the Glottal Stop in Nicaraguan Spanish. Ph.D. dissertation, Ohio State University, Columbus, OH, USA.

Criado, Lorenzo. 1948. El habla de Albalá: Contribución al studio de la dialectología extremeña. *Revista de Estudios Extremeños* 4: 398–407.

Crowhurst, Megan J. 2018. The Influence of Varying Vowel Phonation and Duration on Rhythmic Grouping Biases among Spanish and English Speakers. *Journal of Phonetics* 66: 82–99. [CrossRef]

Dabkowski, Meghan F. 2018. Variable Vowel Reduction in Mexico City Spanish. Ph.D. dissertation, Ohio State University, Columbus, OH, USA.

Delforge, Ann M. 2009. The Rise and Fall of Unstressed Vowel Reduction in the Spanish of Cusco, Peru: A Sociophonetic Study. Ph.D. dissertation, University of California, Davis, CA, USA.

Delforge, Ann M. 2012. 'Nobody wants to sound like a provinciano': The recession of unstressed vowel devoicing in the Spanish of Cusco, Perú. *Journal of Sociolinguistics* 16: 311–55. [CrossRef]

Di Napoli, Jessica. 2015. Glottalization at phrase boundaries in Tuscan and Roman Italian. In *The Phonetics-Phonology Interface: Representations and Methodology*. Edited by Joaquín Romero and María Riera. Amsterdam: John Benjamins, pp. 125–48.

DiCanio, Christian. 2009. The phonetics of register in Takhian thong Chong. *Journal of the International Phonetic Association* 39: 162–88. [CrossRef]

Dilley, Laura, Stephanie Shattuck-Hufnagel, and Mari Ostendorf. 1996. Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics* 24: 423–44. [CrossRef]

Docherty, Gerard, and Paul Foulkes. 2005. Glottal varitants of (t) in the Tyneside variety of English: An acoustic profiling study. In *A Figure of Speech—A Festschrift for John Laver*. Edited by William Hardcastle and Janet Beck. London: Lawrence Erlbaum, pp. 173–99.

Esling, John H., Scott R. Moisik, Allison Benner, and Lise Crevier-Buchman. 2019. *Voice Quality: The Laryngeal Articulator Model*. Cambridge: Cambridge University Press.

Esposito, Christina M. 2005. An acoustic and electroglottographic study of phonation in Santa Ana del Valle Zapotec. Paper presented at the 79th Meeting of the Linguistic Society of America, San Francisco, CA, USA, November 18–21.

Esposito, Christina M., and Sameer ud Dowla Khan. 2020. The cross-linguistic patterns of phonation types. *Language and Linguistics Compass* 14: e12392. [CrossRef]

Esposito, Christina M., Morgan Sleeper, and Kevin Schäfer. 2021. Examining the relationship between vowel quality and voice quality. *Journal of the International Phonetic Association* 51: 361–91. [CrossRef]

Estebas-Vilaplana, Eva, and Pilar Prieto. 2009. La notación prosódica del español: Una revisión del Sp_ToBI [Prosodic notation in Spanish: A review of Sp_ToBI]. *Estudios de Fonética Experimental* 17: 265–83.

Frota, Sonia, Mariapaola d'Imperio, Gorka Elordieta, Pilar Prieto, and Marina Vigário. 2007. The phonetics and phonology of intonational phrasing in Romance. In *Segmental and Prosodic Issues in Romance Phonology*. Edited by Pilar Prieto, Joan Mascaró and Maria-Josep Solé. Amsterdam: John Benjamins, pp. 131–53.

Garellek, Marc. 2014. Voice quality strengthening and glottalization. *Journal of Phonetics* 45: 106–13. [CrossRef]

Garellek, Marc. 2019. The phonetics of voice. In *The Routledge Handbook of Phonetics*. Edited by William F. Katz and Peter F. Assmann. London: Routledge, pp. 75–106.

Garellek, Marc. 2022. Theoretical achievements of phonetics in the 21st century: Phonetics of voice quality. *Journal of Phonetics* 94: 101155. [CrossRef]

Garellek, Marc, and Patricia Keating. 2015. Phrase-final creak: Articulation, acoustics, and distribution. Paper presented at the 89th Meeting of the Linguistic Society of America, Portland, OR, USA, January 8.

Garellek, Marc, Robin Samian, Bruce R. Gerratt, and Jody Kreiman. 2016. Modeling the voice source in terms of spectral slopes. *The Journal of the Acoustical Society of America* 139: 1404–10. [CrossRef]

González, Carolina, and Christine Weissglass. 2017. Hiatus resolution in L1 and L2 Spanish: An optimality account. In *Selected Papers from the 45th Linguistic Symposium on Romance Languages*. Romance Languages and Linguistic Theory (LSRL) 12. Edited by Ruth E. V. Lopes, J. Juanito Ornelas de Avelar and Sonia M. L. Cyrino. Amsterdam: John Benjamins, pp. 79–96.

González, Carolina, Christine Weissglass, and Daniel Bates. 2022. Creaky voice and prosodic boundaries in Spanish: An acoustic study. *Studies in Hispanic and Lusophone Linguistics* 15: 33–65. [CrossRef]

Gordon, Matthew, and Peter Ladefoged. 2001. Phonation types: A cross-linguistic overview. *Journal of Phonetics* 29: 383–406. [CrossRef]

Gynan, Shaw Nicholas N., and Ernesto Luís López Almada. 2020. *The Glottal Stop in Guaraní and Paraguayan Spanish*. Issues in Hispanic and Lusophone Linguistics. Edited by Rajiv Rao. Amsterdam: John Benjamins Publishing Company, vol. 28, pp. 227–62.

Hillenbrand, James, Ronald A. Cleveland, and Robert L. Erickson. 1994. Acoustic correlates of breathy vocal quality. *Journal of Speech and Hearing Research* 37: 769–78. [CrossRef] [PubMed]

Huber, Dieter. 1988. Aspects of the Communicative Fuction of Voice in text Intonation: Constancy and Variability in Swedish Fundamental Frequency Contours. Ph.D. dissertation, University of Göteborg, Göteborg, Sweden.

IBM Corp. 2022. *IBM SPSS Statistics for Windows, Version 29.0*. Armonk: IBM Corp.

Iseli, Markus, Yen-Liang Shue, and Abeer Alwan. 2007. Age, sex, and vowel dependencies of acoustic measures related to the voice source. *Journal of the Acoustical Society of America* 121: 2283–95. [CrossRef] [PubMed]

Keating, Patricia, Christina Esposito, Marc Garellek, Sameer ud Dowla Khan, and Jianjing Kuang. 2010. Phonation contrasts across languages. *UCLA Working Paper in Phonetics* 108: 188–202.

Keating, Patricia, Marc Garellek, and Jody Kreiman. 2015. Acoustic properties of different kinds of creaky voice. Paper presented at 18th International Congress of Phonetic Sciences (ICPhS), Glasgow, UK, August 10–14.

Kim, Ji Young. 2017. Voice quality transfer in the production of Spanish heritage speakers and English L2 learners of Spanish. In *Romance Languages and Linguistic Theory*. Edited by Perpiñán Silvia, David Heap, Itziri Moreno-Villamar and Adriana Soto-Corominas. Amsterdam: John Benjamins Publishing Company, vol. 11, pp. 191–207.

Klatt, Dennis H., and Laura C. Klatt. 1990. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *The Journal of the Acoustical Society of America* 87: 820–57. [CrossRef] [PubMed]

Kreiman, Jody, and Bruce R. Gerratt. 2012. Perceptual interactions of the harmonic source and noise in voice. *Journal of the Acoustic Society of America* 131: 492–500. [CrossRef] [PubMed]

Ladefoged, Peter. 1971. *Preliminaries to Linguistic Phonetics*. Chicago: University of Chicago Press.

Ladefoged, Peter. 1983. The linguistic use of different phonation types. In *Vocal Fold Physiology: Contemporary Research and Clinical Issues*. Edited by Diane M. Bless and James Abbs. San Diego: College-Hill Press.

Ladefoged, Peter, and Ian Maddieson. 1996. *The Sounds of the World's Languages*. Oxford: Blackwell.

Laver, John. 1980. *The Phonetic Description of Voice Quality*. Cambridge: Cambridge University Press.

McKinnon, Sean. 2018. A sociophonetic analysis of word-initial vowel glottalization in monolingual and bilingual Guatemalan Spanish. Paper presented at Hispanic Linguistics Symposium, University of Texas at Austin, Austin, TX, USA, October 25.

Mendoza, Elvira, Nieves Valencia, Juana Muñoz, and Humberto Trujillo. 1996. Differences in voice quality between men and women: Use of the long-term average spectrum (LTAS). *Journal of Voice* 10: 59–66. [CrossRef] [PubMed]

Michnowicz, Jim, and Laura Kagan. 2016. *On Glottal Stops in Yucatan Spanish: Language Contact and Dialect Standardization. Edited by Sandro Sessarego and Fernando Tejedo-Herrero. Issues in Hispanic and Lusophone Linguistics*. Amsterdam: John Benjamins, vol. 8, pp. 217–40.

Mohamed, Sherez, Carolina González, and Anje Muntendam. 2019. Arabic-Spanish Language Contact in Puerto Rico: A Case of Glottal Stop Epenthesis. *Languages* 4: 93. [CrossRef]

Morrison, Geoffrey, and Paula Escudero. 2007. A cross-dialect comparison of Peninsular-and Peruvian-Spanish vowels. Paper presented at 16th International Congress of Phonetic Sciences, Saarbrücken, Germany, August 6–10; pp. 1505–8.

O'Neill, Paul. 2005. Utterance final /s/ in Andalusian Spanish. The phonetic neutralization of a phonological contrast. *Language Design* 7: 151–66.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing, Version 4.3.1*; Vienna: R Foundation for Statistical Computing. Available online: https://www.R-project.org/ (accessed on 1 May 2023).

Redi, Laura, and Stephanie Shattuck-Hufnagel. 2001. Variation in the realization of glottalization in normal speakers. *Journal of Phonetics* 29: 407–29. [CrossRef]

Sessarego, Sandro. 2012. Unstressed vowel reduction in Cochabamba, Bolivia. *Revista Internacional de Lingüística Iberoamericana* 10: 213–27.

Smith, Caroline. 1999. Marking the boundary: Utterance-final prosody in French questions and statements. Paper presented at 14th International Congress of Phonetic Science, San Francisco, CA, USA, August 1–7; pp. 1181–84.

Sosa, Juan Manuel. 2003. La notación tonal del español en el modelo Sp_ToBI. In *Teorías de la entonación [Theories of Intonation]*. Edited by Pilar Prieto. Barcelona: Ariel, pp. 155–84.

Thon, Sonia. 1989. The glottal stop in the Spanish spoken in Corrientes, Argentina. *Hispanic Linguistics* 3: 199–218.

Torreira, Francisco, and Mirjam Ernestus. 2011. Realization of voiceless stops and vowels in conversational French and Spanish. *Laboratory Phonology* 2: 331–53. [CrossRef]

Trittin, Pamela J., and Andrés de Santos y Lleó. 1995. Voice quality analysis of male and female Spanish speakers. *Speech Communication* 16: 359–68. [CrossRef]

Valentín-Márquez, Wilfredo. 2006. La Oclusión Glotal y la Construcción Lingüística de Identidades Sociales en Puerto Rico. In *Selected Proceedings of the 9th Hispanic Linguistics Symposium*. Edited by Nuria Sagarra and Almeida J. Toribio. Somerville: Cascadilla Proceedings Project, pp. 326–41.

*languages*

MDPI

*Article*

# Sound Change and Consonant Devoicing in Word-Final Sibilants: A Study of Brazilian Portuguese Plural Forms

**Wellington Mendes, Jr.**

Department of Language Studies, Universidade Federal de Uberlândia, Uberlândia 38400-902, Brazil; wellington.mendes@ufu.br

**Abstract:** This study investigates consonant devoicing in Brazilian Portuguese (BP), in order to assess whether an ongoing sound change is taking place. We examine plural forms consisting of a stop consonant followed by a word-final sibilant, such as in *redes* [hedz] ~ [heds] ~ [hets] and *sedes* [sɛdz] ~ [sɛds] ~ [sɛts], focusing on the emergence of voiceless sibilants before word-initial vowels (e.g., *redes amarelas*, 'yellow hammocks'). If sibilants remain voiceless despite a following vowel, this challenges the expected regressive voicing assimilation in BP and raises the question of the conditions under which this devoicing occurs. Data were collected through recordings of oral production from twenty Brazilian speakers, using reading and picture naming tasks. Sibilant voicing was quantified using harmonics-to-noise ratio (HNR). A linear mixed-effects model—including random intercepts and slopes for both speakers and words—reveals that sibilants are significantly more voiced before a vowel than before a pause, but this voicing is substantially reduced when the sibilant is preceded by voiceless consonants. These findings indicate an ongoing devoicing process at pre-vocalic word boundaries in BP, affecting clusters [pz, tz, kz] and [bz, dz, gz] alike. Spectrographic analyses indicate that not only the sibilants but also their preceding stop may exhibit devoicing. Moreover, minimal-pair considerations suggest that speakers potentially maintain sibilant voicing in certain lexical items to preserve intelligibility (e.g., *gra*[dz] 'grades' and *se*[dz] 'headquarters' vs. *grá*[ts] 'free' and *se*[ts] 'sets'). Drawing on Exemplar Theory, we propose a competition between the influence of the phonological environment and word-final devoicing: sibilants are sometimes voiced due to a following vowel (e.g., *botes argentinos* [bɔtz ah.ʒẽ.'tʃi.nus] 'Argentine boats'), but they often emerge as voiceless due to consonantal devoicing (e.g., [bɔts ah.ʒẽ.'tʃi.nus]), resulting in both expected and unexpected forms. We suggest that fine phonetic detail, whether associated with allophonic or emergent sound patterns, contributes to the construction of phonological representations.

**Keywords:** devoicing; sibilants; sound change; Brazilian Portuguese

## 1. Introduction

This study aims to investigate pre-vocalic consonant devoicing in Brazilian Portuguese (BP), in order to ascertain whether an ongoing sound change is taking place. Consonant devoicing (CD) is a linguistic phenomenon where voiced consonants such as /b, d, g/ are pronounced as voiceless segments such as [p, t, k]. In Brazilian Portuguese (BP), certain patterns suggest that formerly reliably voiced sibilants may now be undergoing a process of devoicing, even in contexts where voicing would be expected. For example, the plural form *redes* 'hammocks' is typically pronounced [hedz] before a word-initial vowel, but speakers may also be producing [heds] or [hets]. This variability raises questions about whether BP is experiencing a sound change affecting its word-final sibilants, particularly in environments previously associated with voicing.

In BP, /s/ and /z/ are distinct phonemes, occurring in minimal pairs of words such as *casa* /ˈkazɐ/ ('house') and *caça* /ˈkasɐ/ ('hunt'). However, the contrast between /s/ and /z/ is neutralized word-finally, such that a final /z/ does not typically occur. For example, the plural marker in BP, spelled *-s*, is realized as either [s] or a neutralized sibilant segment depending on the phonological context. When followed by a vowel or a voiced consonant, this segment undergoes regressive assimilation, as in *casas amarelas* [ˈkazɐzamaɾɛlɐs] ('yellow houses'), where the plural marker /s/ in *casas* becomes voiced [z] due to the following voiced segment [ɐ] in *amarelas* (Cristófaro-Silva & Mendes, 2022).

Traditionally, this assimilation results in a voiced realization [z], mirroring processes observed in other Romance languages. Catalan, for example, neutralizes voiced and voiceless obstruents word-finally (e.g., *amiga* [əˈmiɡə] 'female friend' vs. *amic* [əˈmik] 'male friend') but voices them before vowels or voiced consonants. In *cas* ('case'), the final /s/ is voiceless when followed by a pause [ˈkas], but it is realized as [z] when followed by a vowel-initial word, as in *cas obert* ('open case') [ˈkaz uˈβɛrt] (Carbonell & Llisterri, 2009).

However, if the sibilants in (stop + sibilant) clusters (e.g., [ts], [ps]) surface as voiceless in Brazilian Portuguese even when followed by a word-initial vowel—for instance, in phrases like *botes amarelos* ("yellow boats") or *envelopes azuis* ("blue envelopes")—this might suggest that a devoicing phenomenon is taking place. Here, the sibilant remains voiceless despite the following vowel, indicating a resistance to the expected assimilation observed elsewhere.[1] Preliminary spectrographic analyses (Mendes, 2023) provide evidence for such voiceless realizations, which contrasts with the general trend of regressive voicing in BP. This behavior warrants further investigation into the phonetic and phonological factors governing such devoicing.

Consonant Devoicing (CD), especially in word-final positions, is a common pattern in sound change phenomena confirmed by numerous cross-linguistic studies and consistently appears in both first and second language acquisition contexts (Blevins, 2006; Albuquerque, 2011; Broselow, 2018; Jatteau et al., 2019a). Various motivations for final devoicing have been proposed in the literature. These include the absence of clear transitions between consonants and vowels essential for perceiving voicing contrasts (Jatteau et al., 2019a); anticipatory glottal opening for respiratory purposes (Myers, 2012; Hutin et al., 2020); reduction in subglottal pressure toward the end of utterances leading to voicing cessation before obstruent release (Westbury & Keating, 1986); and difficulties in voicing production and perception during final lengthening (Blevins, 2006; Ohala, 1997). While these explanations account for devoicing in pre-pausal environments, the pattern observed in Brazilian Portuguese occurs in pre-vocalic contexts, raising the question of what might have led to this pattern.

When stop + sibilant clusters undergo devoicing in BP, one might initially attribute the sibilant's devoicing to the influence of a preceding voiceless segment. However, pre-vocalic sibilant devoicing also appears in contexts with a preceding voiced stop—for instance, in *bodes amarelos* 'yellow goats' ([ˈbɔds.a.ma.ˈɾɛ.lus] ~ [ˈbɔdz.a.ma.ˈɾɛ.lus]) and *jegues argentinos* 'Argentine mules' ([ˈʒɛgs.ah.ʒẽ.ˈtʃi.nʊs] ~ [ˈʒɛgz.ah.ʒẽ.ˈtʃi.nʊs]), where the devoiced sibilant follows [d] or [g].[2] Such examples indicate that devoicing cannot be explained solely by the preceding stop's voicelessness; rather, the broader preceding phonological environment—including both voiced and voiceless stops—must be evaluated to fully understand the conditions under which sibilant devoicing occurs.

CD may also interact with the following phonological environment. The study of Strycharczuk (2012) highlights how the following phonological context can either implement or inhibit word-final devoicing in West Flemish. In this dialect, word-final obstruents typically undergo devoicing in isolation or when followed by a voiceless segment. However, when a sonorant consonant or vowel follows, devoicing may be blocked, and instead, pre-sonorant voicing emerges, maintaining or introducing voicing in the obstruent. For example, in *dat mens* ('that person'), the final fricative in *mens* would usually devoice to [mɛns], but in the context of the voiced sonorant /ɪ/ in *dat mens is* ('that person is'), the fricative surfaces as voiced, yielding [mɛn.zɪs]. This shows that the following phonological environment, particularly the presence of a sonorant, can prevent word-final devoicing.

Another factor involves the influence of orthographic representations on CD. Hayes-Harb et al. (2018) explored this by exposing native English speakers to German-like words (e.g., /ʃtɑit/ and /ʃtɑid/, both pronounced [ʃtɑit]), paired with pictures and, in some cases, their written forms (e.g., *steit* and *steid*). During the test, participants who had seen the written forms were more likely to produce final voiced obstruents when naming the pictures. This suggests that visual access to the written forms interfered with their ability to acquire target-like pronunciations.

The influence of word frequency on CD has also been explored. De Schryver et al. (2008) found that Dutch low-frequency words, like *plonzen* [plɔnzən] ~ [plɔnsən] (to splash) and *omhelzen* [ɔm'ɦɛl.zən] ~ [ɔm'ɦɛlsən] (to embrace), were more susceptible to /z/ devoicing than high-frequency words. The study showed that less frequent words, which have weaker mental representations, are more likely to deviate from standard pronunciation patterns. For instance, low-frequency words were more often produced with devoiced fricatives compared to high-frequency words like *reizen* (to travel) or *blozen* (to blush), which resisted devoicing due to their frequent use.

Nevertheless, the literature presents two opposing views on the role of word frequency in lenition processes like final devoicing. On the one hand, Bybee (2010) underscores that frequent usage promotes segmental weakening, whereby words become more streamlined. For example, in English, high-frequency phrases such as "gonna" (going to) and "wanna" (want to) demonstrate how frequent words undergo phonological reduction to facilitate faster and more efficient speech. On the other hand, the aforementioned De Schryver et al. (2008) view emphasizes that high-frequency words maintain stronger lexical representations, which can shield them from undergoing devoicing.

These phonological insights underline that even in languages like contemporary Brazilian Portuguese, where certain consonantal contrasts (such as /s/-/z/) are already neutralized word-finally, variable final devoicing can still emerge. In BP, consonant devoicing might reflect an ongoing sound change. Earlier research, referred to here as 'Stage 1' (Cristófaro-Silva, 2003; Bisol, 2005; Seara et al., 2017), characterizes the expected realizations of word-final stops followed by /s/ as follows:

a. Voiced stop + /s/ before pause → [s], e.g., /bs/## → [bs]##

b. Voiced stop + /s/ before a vowel → [z], e.g., /bs/#V → [bz]#V

c. Voiceless stop + /s/ before pause → [s], e.g., /ps/## → [ps]##

d. Voiceless stop + /s/ before a vowel → [z], e.g., /ps/#V → [pz]#V

In other words, for a form like *clubes*, the expected surface pattern would be [klubs] (with no laryngeal assimilation from [b] to [s]) if followed by a pause. Under this Stage 1 description, there was no devoicing reported in the position of the prevocalic voiced sibilant (i.e., the /s/ following a voiced stop was realized as [z] before a vowel).[3]

However, the current study suggests a 'Stage 2', in which devoicing now affects precisely those environments where the literature predicted voiced realizations. Specifically, we observe that (1) instead of the expected [bz]#V in (b), speakers increasingly produce [bs]#V or even [ps]#; and (2) instead of the expected [pz]#V in (d), speakers now frequently realize [ps]#V. This shift might indicate that BP is in the midst of a sound change.

Building upon the phonetic and phonological factors discussed above, this study aims to determine which mechanisms might be driving consonant devoicing in Brazilian Portuguese. By examining the phonetic properties of CD in BP and assessing predictors such as adjacent phonological context, orthographic input (task type), and word frequency—as well as incorporating lexical item and individual speaker variation as random effects—we aim to determine whether BP is undergoing a sound change. This investigation will not only shed light on the specific dynamics of CD in BP but also contribute to the broader understanding of how phonetic tendencies interact with language-specific phonological systems.

This paper is structured as follows: the Section 2 outlines what is currently known about consonant devoicing in BP. Following that, the Section 3 outlines the methodology utilized in this study. Subsequently, the Section 4 presents and reports on the results, followed by the conclusions.

## 2. Consonant Devoicing in Brazilian Portuguese

To the best of our knowledge, prior investigations into the phenomenon of obstruent devoicing in L1 BP are lacking. However, some studies have provided insights into the devoicing of BP vowels (Meneses, 2012, 2016), as well as the devoicing of L2 English consonants by BP speakers (Albuquerque, 2011; Cristófaro-Silva & Mendes, 2022). These studies will be described below, as they may shed light on the potential occurrence of CD in BP.

Meneses (2012) examined the devoicing of the high vowels /i/ and /u/ in BP, exemplified by words such as *[kus]tódia* for 'custódia' and *[bis]coito* for 'biscoito', respectively. The author states that, in phonetic terms, devoicing can be understood as a result of coarticulation between adjacent segments during everyday speech: an increase in the overlap of articulatory gestures, influenced by structural factors, may lead to the compression of the vocalic gesture relative to the glottal gesture. Meneses (2012) found that total vowel devoicing occurred in 38% of the data, with partial devoicing observed in 23%, suggesting that the phenomenon is not categorical and may represent an ongoing sound change in BP. Furthermore, the findings imply a tendency for vowel devoicing when followed by a voiceless consonant, indicating a phonetic assimilation between the unstressed vowel and the adjacent voiceless segment.

Subsequently, Meneses (2016) also investigated devoicing in BP, focusing on the reduction of post-stressed vowels (e.g., *lance* 'bid' [lã.si] ~ [lã.si] ~ [lãs]). The study suggests that post-stressed vowels undergo a reduction process that leads to devoicing rather than outright deletion, challenging traditional views that consider this process as apocope. The findings indicate that vowel devoicing is facilitated by an extreme overlap between consonant and vowel gestures, which results in insufficient time to maintain voicing during the vowel segment. This articulatory overlap creates a narrow oral aperture, reducing the ability to sustain voicing, thus leading to devoicing. This pattern is not merely epiphenomenal but arises from specific phonetic conditions that involve adjustments in the motor program of speech.

Other studies have investigated devoicing processes in the development of L2 English consonants by BP speakers. Albuquerque's (2011) study, for instance, unveiled challenges encountered by BP speakers in perceiving contrasts such as 'cap /kæp/ vs. cab /kæb/', 'bat /bæt/ vs. bad /bæd/', and 'back [bæk] vs. bag [bæg]'. The findings indicated a common tendency among participants to misperceive voiced stops as voiceless ones, prompting inquiries into whether this challenge could be attributed to cross-linguistic interference. Results also showcased greater accuracy in discriminating the voiced–voiceless contrast for bilabials compared to alveolars and velars, regardless of the participants' proficiency level.

Still within the domain of L2, but this time concerning production, the study conducted by Cristófaro-Silva and Mendes (2022) assessed devoicing patterns of voiced alveolar fricatives among BP speakers of L2 English. Their research examined the pronunciation of word-final /z/ in English plural forms—such as in *labs*, *sides*, and *bags*—employing visual identification of the voicing bar as a method to evaluate voicing. Results showed that only 14% of words with underlying /z/ were produced as being voiced when the fricative was followed by a pause. When the word was followed by a word-initial vowel, voicing rates were higher (41%), reflecting a regressive assimilation rule from BP. The authors then ponder whether the tendency to devoice in the L2 could stem from their own L1, as opposed to being a result of mimicking English input, where word-final fricatives are only partially voiced (Maniwa et al., 2009).

The present study aims to address the gap highlighted by Cristófaro-Silva and Mendes (2022) by investigating whether BP word-final sibilants tend to devoice before a vowel and which factors influence this. To comprehend the potential devoicing of word-final obstruents, one must also consider another emerging phenomenon that interacts with CD: the weakening of high-front vowels in BP.

An integral aspect of BP phonology involves the strategic insertion of a high-front vowel to prevent the formation of illicit consonant clusters, typically represented in spelling by two consecutive consonant letters (Collischonn, 1996). This phonetic tactic applies across the native lexicon and loanwords, typically involving vowel insertion between two consonants. For instance, in native words like *dogma* [ʻdɔ.gi.mə] ʻdogma' and *afta* [ʻa.fi.tə] ʻcold sore', and in loanwords like *podcast* [pɔ.dʒi.'kɛs.tʃi],[4] the inserted vowel, most often [i] but sometimes [e], surfaces between the consonants to resolve marked clusters. In the case of #sC- clusters such as in *Skype* [is.'kaj.pi], the vowel is instead introduced at the onset of the word, effectively breaking up the initial cluster. However, recent studies indicate a gradual decline in the usage of high-front vowels within word-final post-tonic syllables in BP (Soares, 2016). Consequently, formerly illicit consonant clusters are beginning to emerge and alternate with sequences that still retain the vowel (e.g., [duks] ~ [ʻdu.kis] for *duques*, meaning "dukes"). To fully understand the phenomenon of BP devoicing, we should also acknowledge this ongoing sound alternation in BP, which will be formalized as [Cs] ~ [Cis].[5] Please refer to Table 1 for illustrative examples of this phenomenon.

**Table 1.** [Cs] ~ [Cis] alternations in Brazilian Portuguese.

| BP Singular | Transcription | BP Plural | Transcription | Gloss |
|---|---|---|---|---|
| clube | [ʻklu.bi] | clubes | [ʻklu.bs] ~ [ʻklu.bis] | clubs |
| duque | [ʻdu.ki] | duques | [ʻdu.ks] ~ [ʻdu.kis] | dukes |
| jipe | [ʻʒi.pi] | jipes | [ʻʒi.ps] ~ [ʻʒi.pis] | jeeps |
| pote | [ʻpɔ.tʃi] | potes | [ʻpɔ.ts] ~ [ʻpɔ.tʃis] | pots |

Table 1 exhibits BP nouns in their singular orthographic form in the first column, followed by their corresponding transcriptions in the second column, which include an unstressed high-front vowel word-finally. The third column displays the plural forms, appending the letter <s> to the singular form. Our primary interest, shown in the fourth column, demonstrates the alternation between [Cs] and [Cis] word-finally. As mentioned earlier, the alternation between [Cs] ~ [Cis] in BP stems from the reduction and eventual disappearance of unstressed high-front vowels when flanked between a consonant and a final sibilant (Cristófaro-Silva & Leite, 2015; Soares, 2016). The fluctuation between the presence and absence of an unstressed high vowel between consonants signifies an emerging phonetic pattern in BP undergoing gradual stabilization (Cristófaro-Silva, 2016; Mendes, 2023).[6]

Furthermore, as also outlined in Section 1, a more recent development—what we label "Stage 2"—manifests when a voiceless alveolar fricative appears even before a word-initial vowel, contrary to the more conventional expectation of a voiced fricative. For instance, [hets.ah.ʒẽ.'tʃi.nəs] is produced instead of [hedz.ah.ʒẽ.'tʃi.nəs] for "redes argentinas" (Argentine hammocks). Consider Figure 1.

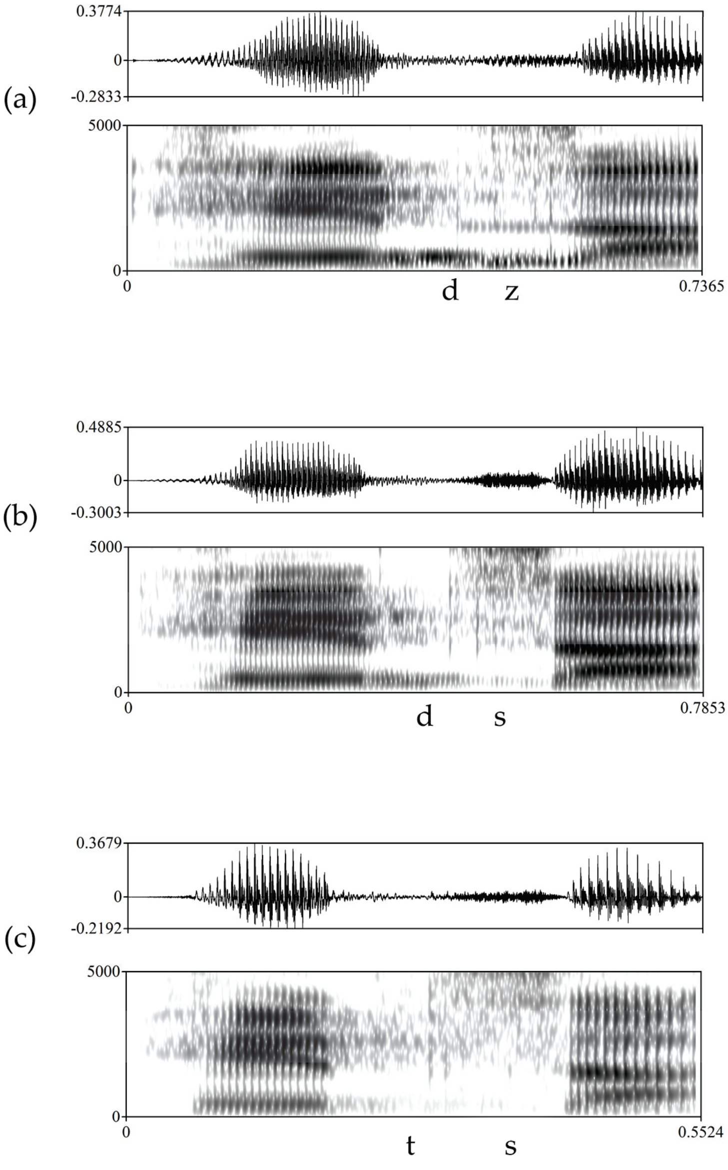



**Figure 1.** Spectrographic analysis of the phrase "redes argentinas" by three different speakers. In (**a**), speaker 1 pronounces the final /s/ as fully voiced [z]. In (**b**), speaker 2 pronounces the final /s/ with partial voicing. In (**c**), speaker 3 pronounces the final /s/ as fully voiceless [s].

Figure 1 displays spectrograms and waveforms of the string [hedza] from the phrase "redes argentinas" (Argentine hammocks) as produced by three participants. In all panels, there is no vowel between the stop and the sibilant, visible in the continuous closure and frication without intervening periodic energy. In Figure 1a, a pronounced voicing bar indicates a fully voiced cluster [dz]. Figure 1b reveals partial voicing, with a faint voicing bar present. By contrast, Figure 1c shows no clear voicing bar at all, indicating a voiceless cluster [ts] in a context where a voiced cluster might typically be expected.

The gradient nature of the spectrographic findings in Figure 1 raises the question of how best to model such variable patterns. In a traditional model of phonology, such as Generative Phonology, CD would be seen as a categorical transformational rule: an underlying voiced obstruent becomes voiceless in word-final position. However, contemporary models such as Exemplar Theory (ET) (Johnson, 1997; Pierrehumbert, 2001; Bybee, 2010) provide a more dynamic perspective by emphasizing the role of frequency in shaping speech sounds. According to ET, sounds are not governed solely by abstract rules but are stored as numerous detailed instances (exemplars) in memory. Crucially, the frequency with which specific exemplars are encountered plays a significant role in determining their prominence in shaping speech perception and production. More frequent exemplars become stronger and more easily accessed. In this sense, the process of devoicing is not a categorical rule but emerges from the accumulation of frequent instances of word-final voiceless obstruents.

ET refines the notion of phonemic categories by proposing that they emerge from clusters of stored exemplars along a continuum of variation. While traditional binary distinctions (e.g., voiced vs. voiceless consonants, tense vs. lax vowels) remain relevant as points of contrast, ET posits that phonemes are better understood as distributions of related exemplars rather than strictly discrete entities. These gradient distributions can influence both the perception and production of speech sounds, emphasizing that a single "phoneme" may encompass significant internal variability.

Furthermore, ET provides a framework for understanding how phonetic detail can be influenced by a wide range of linguistic and non-linguistic factors. Factors such as word frequency, orthography, lexical robustness, speech rate, and speaker identity can all shape the distribution of exemplars and, consequently, affect the realization of speech sounds (Bybee, 2001). This holistic approach to phonetic variation and change highlights the dynamic nature of language use and challenges the notion of a fixed set of phonological rules governing speech production. Such premises will serve as the focal point of our analysis.

## 3. Methodology

### *3.1. Participants*

This study included twenty Brazilian speakers, comprising an equal split of ten males and ten females. The participants' ages ranged from 15 to 17 years old during the data collection phase.[7] We selected individuals who met the following conditions: (a) resided in Belo Horizonte (state of Minas Gerais) or the surrounding region for at least four years;[8] (b) did not report any hearing or speech problems; (c) did not wear orthodontic appliances during the experiment; and (d) were non-smokers. During data collection, all participants were enrolled in one of the technical courses offered by the Federal Center for Technological Education of Minas Gerais, located in the city of Belo Horizonte.

### *3.2. Cluster Types and Words*

Since this study assesses the voicing property of (stop + fricative) sequences, we selected the following cluster types as targets: /ps/, /ts/, /ks/, /bs/, /ds/, and /gs/.[9] When followed by a vowel, these clusters are expected to surface as [pz], [tz], [kz], [bz], [dz] and [gz], respectively. After outlining the cluster types, we defined the criteria used for selecting the target words, as detailed in Table 2.

Table 2 displays the distribution of chosen words in BP, categorized by cluster type and alphabetically organized. We opted to control for grammatical category, number of syllables, and lexical stress. Consequently, we primarily selected disyllabic and paroxytone nouns. We also prioritized lexical items that were visually straightforward to identify in the picture-naming task. However, a limitation arose during the selection process due to the difficulty of finding disyllabic words that both fit the required phonological clusters and could be easily depicted in the picture-naming task. As a result, five non-disyllabic words (underlined items) were included, as it was challenging to represent other disyllabic words with the necessary cluster types in a visually comprehensible manner (e.g., *torpes*, *sebes*, *brigues*, and *naipes*).

**Table 2.** Target words.

| ps | ts | ks | bs | ds | gs |
|---|---|---|---|---|---|
| alpes | artes | cheques | árabes | baldes | açougues |
| chopes | botes | cliques | Caribes | cidades | bumerangues |
| crepes | chutes | duques | clubes | grades | jegues |
| envelopes | cortes | leques | orbes | redes | mangues |
| jipes | dentes | toques | plebes | sedes | ringues |
| xaropes | potes | truques | robes | tardes | sangues |

### *3.3. Experimental Protocol*

This section presents the experimental protocol adopted in the study. Due to the COVID-19 pandemic, all interactions with participants were conducted remotely. Initially, participants were briefed on the research's nature and ethics standards, and they digitally signed the Informed Consent Form (ICF). Following this, they completed the Experimental Group Data Collection Form via the Google Forms Platform. Subsequently, each participant joined an online meeting on Google Meet with the researcher, which lasted approximately 40 min. Uniform instructions and timeframes were provided to ensure consistency across participants.

The data collection process encompassed three distinct stages: a familiarization task, picture naming, and sentence reading. All these stages included one distractor per target word[10]. During the familiarization task, participants were asked to identify lexical items corresponding to displayed images. This initial phase served to acquaint participants with the stimuli and was not subjected to analysis.

For the picture-naming task, participants were prompted to count and name the objects shown, with the instruction: "Conte e diga o nome de cada elemento que você vê", meaning "Count and say the name of each object you see". This allowed participants to enumerate and pronounce the target words in isolation (see Table 3). In this study, the experimental design contrasts two following phonological contexts: pause and vowel.[11] Due to this reason, participants repeated the picture-naming task with the target words embedded in a carrier phrase. The carrier phrase "[number] [target word] [adjective] são vistos" ("[number] [target word] [adjective] are seen"), was used to assess whether a following word-initial vowel would influence voicing properties in BP, as well as to ensure natural and consistent usage of the target words within a controlled context. Table 3 exemplifies this task.

The second experiment involved reading controlled sentences, ensuring a fixed syllable count, along with consistent morphological classes and intonation patterns. When the target words were followed by a pause, participants read a number followed by a plural noun, with the total syllable count restricted to 3 or 4, depending on the presence or absence of the post-tonic vowel [i] in the target word (e.g., *Cinco alpes*, meaning "five Alps"). When the target words were followed by a vowel, the syllable count was restricted to 12 or 13, depending on the presence or absence of [i] (e.g., *Os alpes italianos são belíssimos*, meaning "the Italian Alps are stunning"). In this latter case, the sentences were structured as follows: [determiner] [target word] [adjective] [linking verb] [adjective or participle]. Criteria such as avoiding rhymes, alliterations, and cacophonies were also considered. By the end, participants produced each target word four times, considering two different tasks and two phonological contexts. Data containing misidentified target words or background noise were excluded from the analysis. Table 4 summarizes our data.

**Table 3.** Examples of stimuli and expected answers.

| Stimuli | Expected Answers |
| --- | --- |
|  | quatro crepes |
|  | quatro crepes argentinos são vistos |

**Table 4.** Data summary.

| | |
| --- | --- |
| Words | 36 |
| Tasks | 2 |
| Phonological contexts | 2 |
| Participants | 20 |
| Total of expected tokens | 2880 |
| Total of tokens after filtering | 2833 |

Data were recorded using Open Broadcaster Software Studio (version 28.1) and processed using Adobe Premiere 2020 (version 14.0) and Praat software (version 6.3; Boersma & Weenink, 2023).

### *3.4. Measurement and Variables*

To determine whether a consonant is voiced or voiceless, acoustic analysis often involves examining spectrograms using software like Praat (Boersma & Weenink, 2023). In a spectrogram, voicing is indicated by periodic striations and the presence of low-frequency energy resulting from vocal fold vibrations. However, visual inspection alone can be insufficient due to variations in recording quality and potential background noise that may obscure these features. Therefore, quantitative acoustic measures are essential for accurately assessing voicing. Several parameters can be used for this purpose, including pulse-based methods, low frequency-to-total intensity ratio, segment duration, and spectral properties like center of gravity. For this study, we employed the Harmonics-to-Noise Ratio (HNR), as validated by Gradoville (2011), because it offers reliable quantification of voicing, particularly in noisy recordings where other methods may be less effective.

According to Bjorndahl (2022), the Harmonics-to-Noise Ratio is defined as the ratio between the harmonic force of periodic components and the noise in an acoustic signal, expressed in decibels (dB). Lower HNR values, close to 0 dB, are typically linked to voiceless sounds, whereas values between 5 and 20 dB suggest a higher degree of voicing. Although commonly used in detecting speech pathologies, HNR has also proven useful in discriminating between various sounds, notably fricative consonants (Hamann & Sennema, 2005; Maniwa et al., 2009; Gradoville, 2011).

Of particular relevance to this study, Maniwa et al. (2009) noted that voiceless English fricatives (e.g., [s] and [ʃ]) in intervocalic contexts did not surpass the 5 dB threshold, whereas voiced intervocalic fricatives (e.g., [z] and [ʒ]) typically ranged between 5 and 10 dB. These findings provide a reference for assessing the degrees of voicing observed in the fricatives of BP in our study. Thus, we determined the voicing of the final sibilant as our dependent variable, expressed by continuous levels of Harmonics-to-Noise Ratio (HNR) values. HNR values were automatically extracted using Silva's (2022) script in Praat (Boersma & Weenink, 2023).

We pose the overarching hypothesis that devoicing of final sibilants is a systematic phenomenon in BP. The predictions regarding the independent variables are outlined as follows:

1. Preceding phonological context: We predict higher sibilant voicing rates when the final sibilant is preceded by a voiced stop or a high-front vowel, as part of a potential progressive assimilation. As previously noted, pre-vocalic devoicing of sibilants may occur not only when the preceding segment is a voiceless stop (e.g., *botes amarelos* ['bɔts.a.ma.'ɾɛ.lus] instead of ['bɔtz.a.ma.'ɾɛ.lus] 'yellow boats') but also when it is a voiced stop (e.g., *sedes amarelas* ['sɛds.a.ma.'ɾɛ.lɐs] instead of ['sɛdz.a.ma.'ɾɛ.lɐs] 'yellow headquarters'). Moreover, although the high-front vowel is inserted only infrequently in this dialect (see Note 7), it remains a potential conditioning factor. The precise role of these preceding segments in Brazilian Portuguese, therefore, remains to be fully explored.

2. Following phonological context: It is anticipated that sibilant voicing will be higher when the final sibilant is followed by a vowel, as opposed to when it is followed by a pause. In this latter environment, it is expected that consonant devoicing will be favored. This projection is based on phonological studies of BP, which propose that the following environment serves as a categorical trigger for the voicing of the final sibilant within word boundaries (Bisol, 2005; Seara et al., 2017). Moreover, studies that assessed consonant devoicing in other Romance languages, such as French (Jatteau et al., 2019a, 2019b) and Catalan (Carbonell & Llisterri, 2009) found an effect of this variable.

3. Task type: We predict that the reading task will elicit lower rates of sibilant voicing compared to the picture-naming task. This expectation is based on the premise that visual exposure to the grapheme <s> encourages participants to associate it with the [s] sound, which typically has lower voicing rates than [z]. Consistent with Hayes-Harb et al. (2018), who demonstrated that written forms can influence learners' productions, the orthographic cues in the reading task are expected to reduce voicing rates. Conversely, the picture-naming task lacks explicit written prompts, potentially resulting in higher rates of sibilant voicing.

4. Word Frequency: We expect that high-frequency words will exhibit a greater degree of devoicing, reflected in a lower harmonics-to-noise ratio, as compared to their low-frequency counterparts. According to Bybee (2010), high-frequency lexical items tend to exhibit greater phonetic reduction and segmental changes over time, as they are stored and retrieved more frequently, reinforcing lenition processes like devoicing. Conversely, low-frequency items, having fewer stored exemplars, may resist such phonological shifts. Word frequency values were retrieved using the search tool provided by *Corpus do Português: Web/Dialects* (Brazilian Corpus: Web/Dialects) (Davies & Ferreira, 2016).

### *3.5. Data Analysis*

An automatic segmentation of the audio signals was performed using a native function of Praat (Boersma & Weenink, 2023). In the Praat Objects window, the option *Annotate → To TextGrid (Silences)* was selected. To delimit the boundaries of the words analyzed in this study, the following parameters were applied: *minimum pitch*: 70 Hz; *time step*: 0 (auto);

*silence threshold*: −35 dB; *minimum silent interval*: −35 s; *minimum sounding interval*: 0.1 s; *sounding interval label*: 0. Following that, we applied an automatic segmentation Python script (version 1.0) adapted from Silva (2022) to identify potential stops, sibilants, and [i]-occurrence based on acoustic features like silence intervals, release bursts, and frication noise. To focus on significant vowel productions, we established a duration threshold: only high-front vowels longer than a specific, low millisecond value—such as 10 milliseconds—were segmented and included in the analysis. This approach helped differentiate genuine [i]-occurrence from brief transitional sounds. We then manually reviewed and adjusted the automatically segmented boundaries for accuracy.

A linear mixed-effects model was specified to assess how the predictors influence HNR values. The model included an interaction between preceding phonological context (voiced stops, voiceless stops, or high-front vowels) and following context (pause or vowel), allowing direct assessment of whether and how sibilant voicing is jointly modulated by these environments. Additional fixed effects were included for task type (reading vs. picture-naming) and word frequency (rescaled).

To account for individual variability, random intercepts and slopes were specified for both participants and words. Each participant and word had a varying overall HNR baseline (intercept) as well as a varying HNR change magnitude between a following pause and a following vowel (slope). The decision to include the following context as a random slope reflected its significance as a categorical phonological factor (Cristófaro-Silva, 2003; Bisol, 2005; Seara et al., 2017).

All analyses were performed in R (R Core Team, 2022) using the *lmer*() function from the *lme4* package (Bates et al., 2015). The model was fit via restricted maximum likelihood (REML) to obtain stable estimates of variance components. Instead of stepwise selection or other model comparison procedures, this single, full model was retained with all hypothesized fixed and random effects. Standard diagnostic checks—including residual plots and convergence assessments—were conducted to confirm appropriate model fit.

## 4. Results

This section explores the voicing rates of word-final sibilants in BP, in order to account for a possible ongoing devoicing process. The data include target words produced with and without the presence of orthographic input, and in two contexts: plural nouns with a final sibilant followed by a pause (e.g., *cidades* [sidad**s**#] 'cities') and plural nouns with a final sibilant followed by a vowel (e.g., *cidades azuis* [sidad**z**azuɪs] 'blue cities').

Given that the production of final sibilants in BP is variable, we investigated the impact of fixed effects—namely, preceding and following phonological contexts, task type, and word frequency—on HNR values. We also treated individual behavior and lexical item as random effects to capture speaker-specific variation and idiosyncrasies tied to particular words. These factors will be addressed in the following subsections.

### *4.1. Following Phonological Context*

Table 5 displays the number of tokens, mean, median, and standard deviation rates of word-final sibilants in BP, utilizing the harmonics-to-noise ratio (HNR) measurement, expressed in dB.

As described in the Methodology section, higher HNR values are associated with the production of segments with a higher degree of voicing. According to Table 5, sibilants before a pause (e.g., *cheques* [ʃɛks#], 'cheques'), which are expected to be produced as [s], have a mean value of 2.8 dB and a median of 2.6 dB. On the other hand, sibilants before a vowel (e.g., *cheques amarelos* [ʃɛkz a.ma.'ɾɛ.lʊs], 'yellow cheques'), which are expected to be produced as [z], have a mean value of 7.9 dB and a median of 7.4 dB.

**Table 5.** Voicing rates of word-final sibilants in BP.

| | N | Mean (HNR) | Median (HNR) | Standard Deviation |
|---|---|---|---|---|
| before pause | 1409 | 2.8 dB | 2.6 dB | 1.7 |
| before vowel | 1424 | 7.9 dB | 7.4 dB | 4.5 |
| Total | 2833 | 5.4 | 4.0 | |

Analyzing the standard deviation values in Table 5, it can be seen that there is greater dispersion around the mean values of [z] than of [s]. In terms of ET, we can assume that currently, exemplars that incorporate the phonetic detail of the voiceless sibilant are more robust (i.e., consistent) in BP. On the other hand, exemplars associated with the production of the voiced sibilant show greater phonetic gradience, thus being more variable. This alternation contradicts the premises of traditional phonological literature, where there would be the categorical production of either voiced or voiceless sibilants.

As established in Section 1 (Bisol, 2005; Seara et al., 2017), the alveolar fricative /s/ in BP is typically voiceless in word-final or syllable-final position but becomes voiced before a vowel or voiced consonant. Consequently, we predict higher HNR indices when sibilants are followed by vowels. Refer to Figure 2.

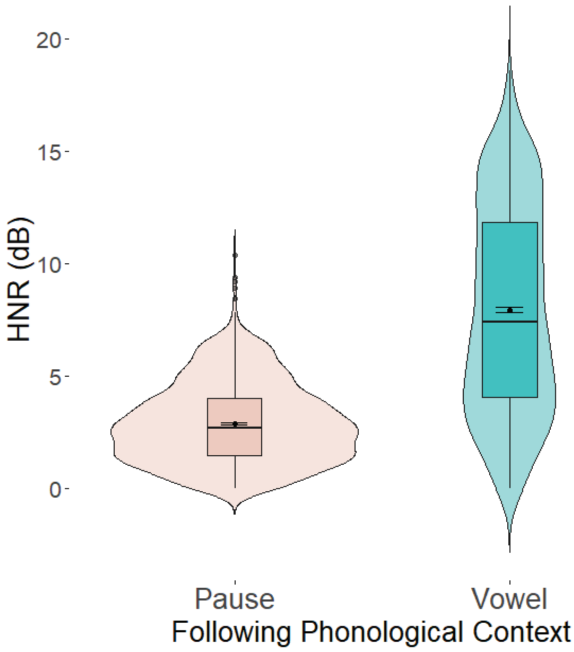



**Figure 2.** Voicing rates of final sibilants per following phonological context.

Figure 2 provides a visual complement to the data reported in Table 5 by displaying how final sibilant voicing varies according to the following phonological context. The results support the earlier observation that voicing levels increase in pre-vocalic environments (e.g., potes [pɔts#] 'pots' vs. potes azuis [pɔtzazuɪs] 'blue pots'). Analysis from the linear mixed-effects model confirmed this factor as statistically significant, showing a substantial positive effect of 5.04 dB for sibilants preceding vowels ($t \approx 28.76$, $p < 0.0001$) (see Table A1 in Appendix A for the entire regression output).

This pattern parallels observations by Cristófaro-Silva and Mendes (2022), who reported that L2 English voiced sibilants produced by Brazilian speakers exhibit significant variability when followed by vowels, being produced both with and without voicing. In order to understand the variability observed in the production of these sibilants in BP, we revisited our spectrographic analysis. Refer to Figure 3.

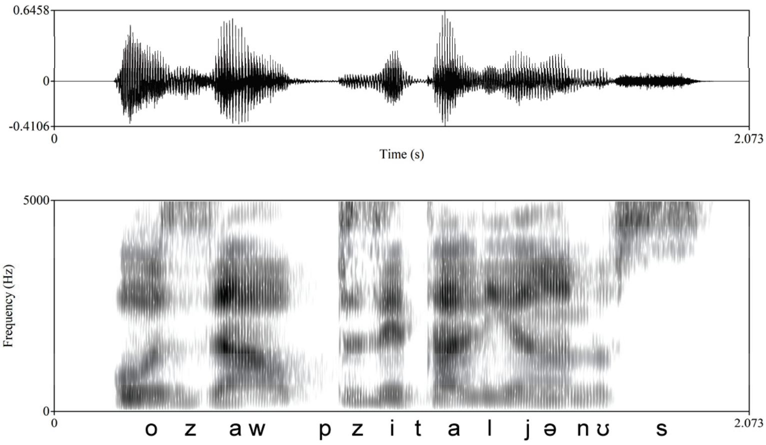



**Figure 3.** Spectrographic analysis of the phrase "os alpes italianos".

Figure 3 displays the spectrogram and waveform of the phrase "os alpes italianos [...]" (the Italian alps), produced by a male participant during the reading task. Notice the presence of the voicing bar during the production of the sibilant in the target word. That is, the form [awp**z**italjənʊs] was produced due to the influence of the following vowel, as predicted by the principle of regressive assimilation in BP. However, the results reported in Figure 2 indicate that many sibilants remain voiceless at word boundaries even when followed by a vowel, as a significant amount of tokens concentrate in the bottommost part of the violin plot. In order to illustrate such tokens, refer to Figure 4.

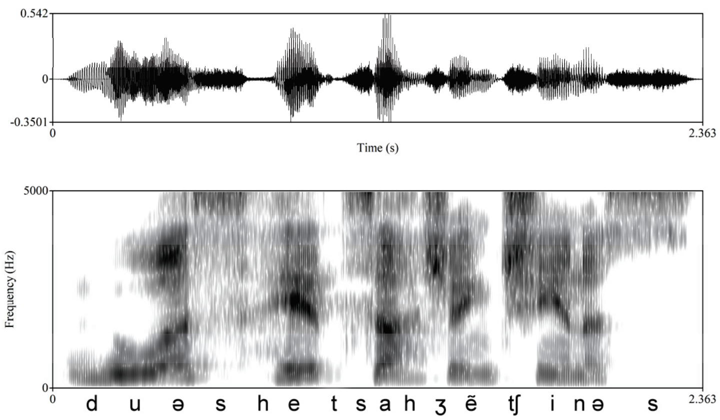



**Figure 4.** Spectrographic analysis of the phrase "duas redes argentinas".

Figure 4 displays the spectrogram and waveform of the phrase "duas redes argentinas [...]" (two argentine hammocks), produced by a male participant. Notice the absence of the voicing bar during the production of the sibilant in the target word. Also note the absence of the dark voicing bar during the production of the stop consonant. This indicates that not only did the sibilant remain voiceless, but the preceding stop also assimilated the voicelessness property.[12] Thus, there was the production of the form [ˈhetsah.ʒẽ.ˈtʃi.nəs] instead of the traditionally expected [ˈhedizah.ʒẽ.ˈtʃi.nəs]. It is worth noting that such behavior was observed in all cluster types followed by vowels evaluated in this paper: [pz#V], [tz#V], [kz#V], [bz#V], [dz#V], [gz#V] became [ps#V], [ts#V], [ks#V] or [bs#V], [ds#V], [gs#V]. Although infrequently, this behavior was observed even in words where consonantal devoicing results in a loss of phonemic contrast (cf. Section 4.5). For instance, "grades" (fences) and "grátis" (free) were sometimes pronounced the same ([ˈgɾats]). Similar cases included pairs like "sedes" (headquarters) and "setes" (sets) [ˈsɛts], "ringues" (boxing rings) and "rinques" (ice rinks) [ˈhĩks], and "tardes" (afternoons) and "tartes" (pies) [ˈtahts].

Hence, the data reported above suggest that the devoicing of (stop + sibilant) sequences in BP may represent a gradually unfolding sound change. Notably, devoicing processes have also been documented in other Romance languages, including European Portuguese (Jesus & Shadle, 2002), Catalan (Carbonell & Llisterri, 2009), French (Jatteau et al., 2019a, 2019b), and Romanian (Hutin et al., 2020).

*4.2. Preceding Phonological Context*

This section investigates whether the preceding phonological environments influence sibilant (de)voicing in BP. We begin by examining the presence or absence of a high-front vowel [i] before /s/.

From the total of 2833 tokens collected after filtering, 64% represent stop + sibilant sequences without any production of an intrusive high-front vowel, as in *cheques* pronounced [ʃɛks]. Conversely, 36% of the tokens include gradient productions of an intrusive [i], as in *cheques* pronounced [ˈʃɛ.kis]. This distribution indicates that while [i]-occurrence remains, a significant portion of speakers are omitting it, reflecting the reported gradual decline of this vowel insertion in unstressed positions in BP (see Soares, 2016; Cristófaro-Silva & Mendes, 2022). Consider Figure 5.

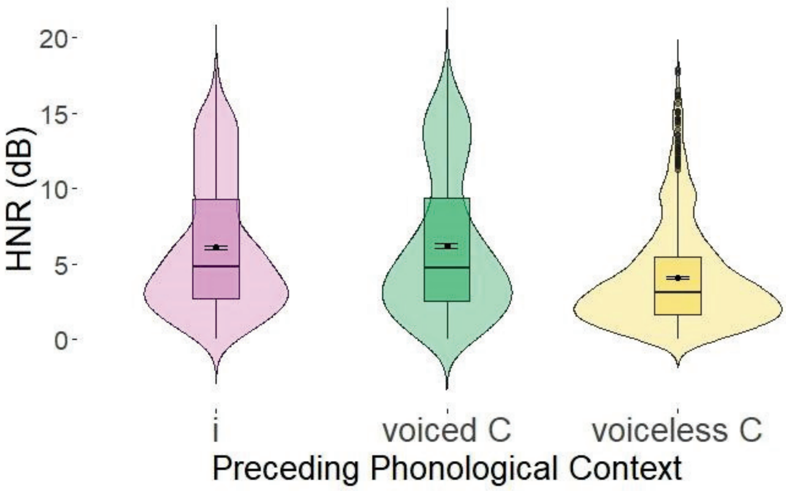



**Figure 5.** Voicing rates of final sibilants per preceding phonological context.

Figure 5 illustrates the voicing rates of final sibilants per preceding phonological context. Sibilants preceded by a high-front vowel [i] (left violin plot) exhibit an average HNR of 6 dB. Sibilants preceded by voiced consonants (middle violin plot) show an average of 6.2 dB. In contrast, sibilants preceded by voiceless consonants (right violin plot) display a lower average HNR of 4 dB.

Although these descriptive means suggest that sibilants are somewhat more voiced when preceded by a vowel or a voiced consonant, the model does not reveal a statistically significant main effect of the preceding environment in isolation (voiced consonant: Estimate = $-0.1538$, $t = -0.648$, $p = 0.5170$; voiceless consonant: Estimate = $-0.2364$, $t = -1.007$, $p = 0.3143$) (cf. Appendix A, Table A1).

However, when we re-leveled the factor so that 'voiced consonant' became the reference rather than 'vowel', the model did uncover a significant difference between 'voiceless consonant' and 'voiced consonant' (Estimate = $-0.506$, $t = -2.537$, $p < 0.05$). This direct comparison aligns with the descriptive means in Figure 5, where voiceless consonants show lower HNR values. The effect also emerges more clearly when we take into account the following phonological context. Let us now focus on the results regarding the interaction between the preceding and following phonological environments.

*4.3. Interaction Between Adjacent Contexts*

To capture whether sibilant voicing differs depending on both the preceding and following phonological environments, the analysis was fit to include an interaction between the adjacent contexts. The model reveals a highly significant effect ($\beta \approx -2.15$, $t \approx -7.69$, $p < 0.0001$) indicating that, in pre-vocalic contexts, sibilants preceded by voiceless stops are realized with notably lower voicing levels than those preceded by voiced stops (see Table A1 in Appendix A for the entire regression output).

Because the LME model selected [i] as the reference level for the preceding context variable, the estimates for voiceless and voiced consonants are compared to the effect of the [i] vowel. Results indicate that when the following context is a pause (baseline), neither a preceding voiced consonant (Estimate = −0.1538, $t = -0.648$) nor a preceding voiceless consonant (Estimate = −0.2364, $t = -1.007$) significantly affects HNR relative to the [i] baseline. In contrast, in a pre-vocalic environment, the interaction term for voiceless consonants is highly significant (Estimate = −2.1465, $t = -7.694$, $p < 0.001$), indicating a substantial decrease in sibilant voicing compared to [i].

Post-hoc pairwise comparisons (see Table A2 in Appendix A) confirm this pattern in two ways. First, when the following context is a vowel (i.e., in pre-vocalic position), sibilants preceded by voiceless stops have significantly lower HNR than those preceded by the [i]-vowel baseline, and those preceded by voiced stops (difference ≈ 2.71 dB, $p < 0.0001$). By contrast, when the following context is a pause, neither a preceding voiceless stop nor a preceding voiced stop leads to a significant HNR difference relative to the [i] baseline (both $p > 0.05$).

These results regarding the interaction between preceding and following environments also allow us to compare two situations originally labeled (b) and (d) in Section 1, where we discussed four possible outcomes for stop + /s/ sequences. Specifically, (b) is the case of a voiced stop + /s/ before a vowel, hypothesized to devoice from [bz] to [bs], and (d) is the case of a voiceless stop + /s/ before a vowel, realized as [ps] instead of [pz]. The interactions (cf. Appendix A, Table A1) show that sibilants following voiceless consonants remain significantly less voiced in pre-vocalic contexts. Crucially, however, the data do not provide strong evidence that sibilants after voiced stops are devoicing before a word-initial vowel: the average HNR for /voiced C + s/ #V remains relatively high (i.e., closer to [z]-like voicing). This asymmetry suggests that the situation described in (d)—the shift toward voiceless realizations (e.g., /ps/#V → [ps]#V)—is robustly supported by the current findings, whereas the situation proposed in (b)—devoicing /bs/#V → [bs]#V—does not receive equivalent statistical support. In other words, progressive voiceless assimilation from a preceding voiceless stop (as in (d)) appears to override the vowel's regressive voicing effect on the adjacent sibilant, whereas a preceding voiced stop still preserves the expected voicing (as in [bz]#V). Notably, the post-hoc comparisons further indicate that HNR after a voiceless stop in a pre-vocalic context remains significantly lower than in the pause condition (see Table A2), reinforcing that progressive voiceless assimilation is robust across these contexts.

*4.4. Task Type*

Regarding task type, we predicted that the reading task would elicit lower rates of sibilant voicing compared to the picture-naming task. This assumption stemmed from the notion that visual exposure to the grapheme <s> might prompt participants to associate it with the phoneme /s/, which has lower voicing rates compared to [z], and, by extension, potentially link it to patterns of word-final devoicing. Consider Figure 6, which displays sibilant voicing rates per task type.

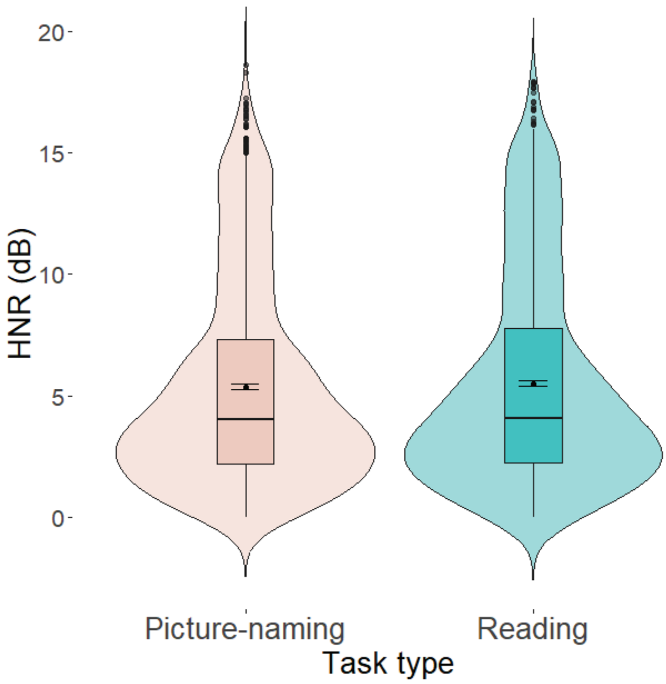



**Figure 6.** Voicing rates per task type.

Figure 6 displays the voicing rates of final sibilants by task type. Sibilants produced in the picture-naming task (with no orthographic input) showed an average HNR of 5.4 dB, whereas those produced in the reading task (with orthographic input) averaged 5.5 dB. Statistical analysis confirmed that this difference was not significant (Estimate = 0.13, SE = 0.12, $t = 1.03$, $p = 0.30$).

This indicates that additional variables, such as phonetic context, lexical properties, and individual linguistic behavior, may exert greater influence on the manifestation of consonant devoicing patterns than orthographic input. Having considered the potential impact of phonetic and orthographic factors on sibilant voicing, the focus now shifts towards examining the influence of lexical properties, particularly the characteristics of individual words.

### *4.5. Word Frequency, Lexical Item and Homophony Avoidance*

In ET, word frequency is regarded as an important factor shaping phonetic variation. Consider Figure 7.

Figure 7 presents a polar plot illustrating the relationship between Log Word Frequency and Mean HNR for different lexical items. Each word is positioned radially according to its log-transformed frequency—the farther from the center, the more frequent the word. The color gradient, ranging from blue (lower HNR values) to red (higher HNR values), represents the mean HNR for each word. Although some frequent words (e.g., "cidades", "redes") appear toward the red end of the spectrum and some infrequent words (e.g., "chopes", "leques") appear in the blue, the overall distribution shows no consistent pattern. Indeed, the LME model indicates that word frequency does not significantly affect HNR values ($\beta = -0.11$, $t = -0.96$, $p = 0.34$), confirming that frequency exerts negligible influence on voicing in these data.

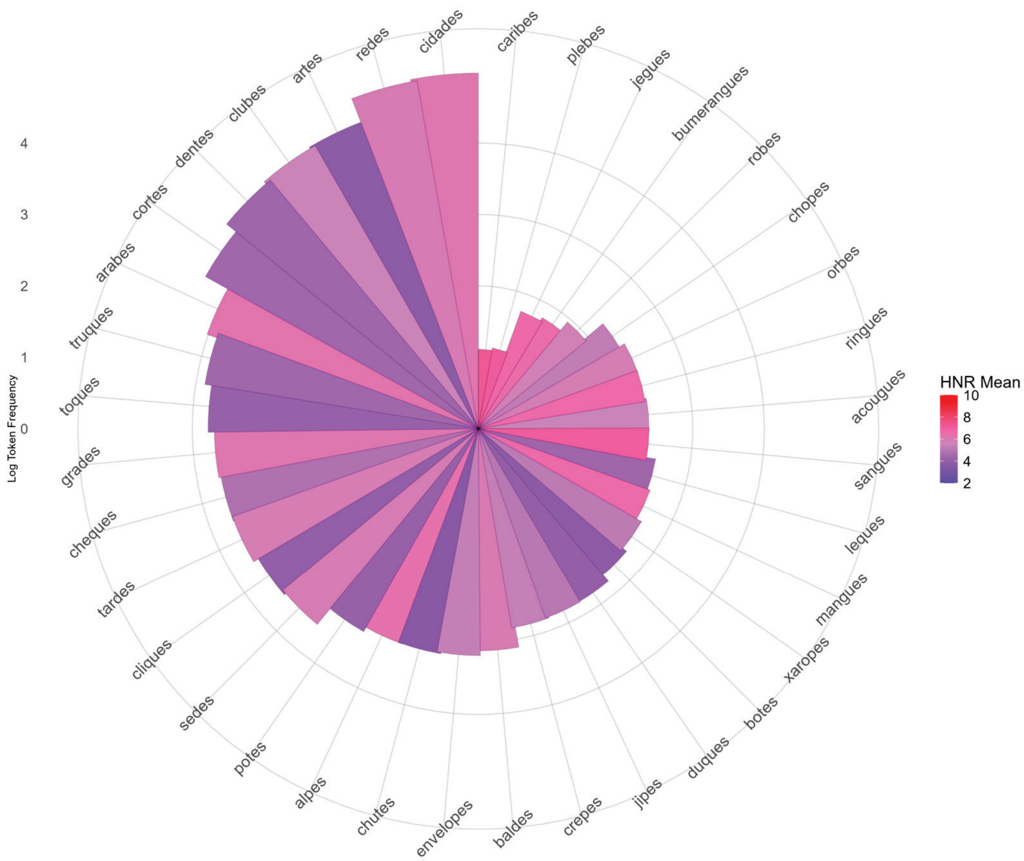



**Figure 7.** Voicing rates per word and lexical frequency.

Given that ET posits that individual words serve as the primary locus of representation, it was also anticipated that each word would exhibit varying levels of sibilant voicing. Consider Figure 8.

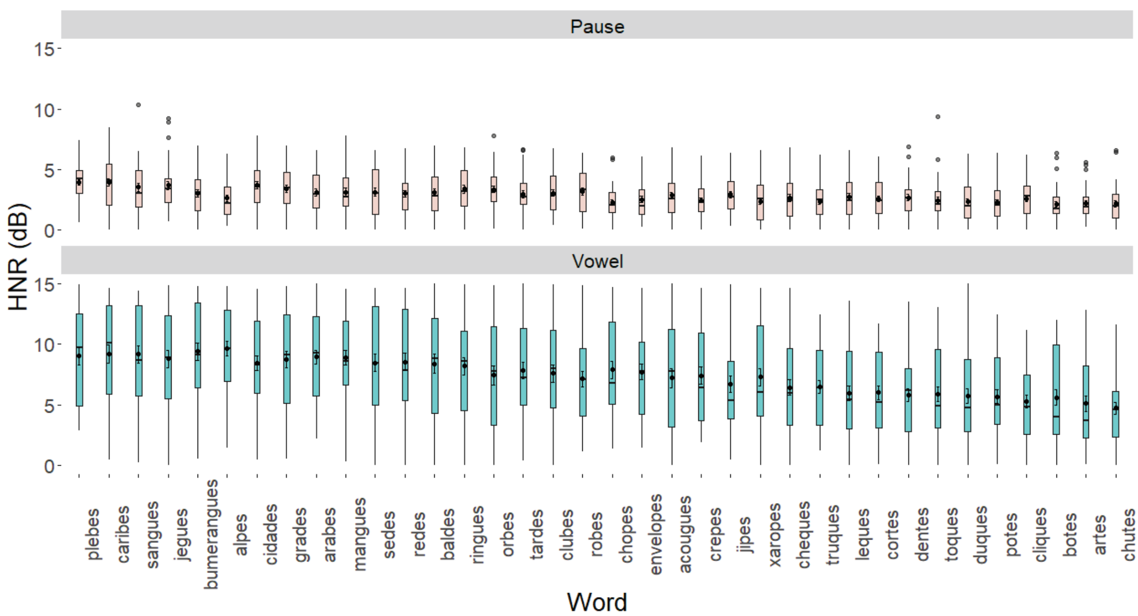



**Figure 8.** Voicing rates per word.

Figure 8 exhibits the voicing rates of final sibilants in BP per lexical item. The results are grouped as follows: nouns with a final sibilant followed by a pause are located in the upper part and are represented by yellow boxplots; nouns with a final sibilant followed by a vowel are located in the lower part and are represented by green boxplots. The distribution of lexical items is displayed in descending order considering the average values of HNR in both production contexts (i.e., sibilant followed by pause or vowel).

A closer inspection of Figure 8 shows that the words with the highest HNR were ‘plebes’ [plɛbz], ‘caribes’ [ka.’ɾibz], ‘sangues’ [s@̃gz], and ‘jegues’ [ʒɛgz]. We highlight the fact that these words are constituted of a voiced stop preceding the sibilant. Conversely, the words with the lowest HNR values—“cliques” [kliks], “botes” [bɔts], “artes” [ahts], and “chutes” [ʃuts]—contain a voiceless stop preceding the sibilant, and cluster on the right side of the graph. These findings corroborate the earlier observation that sibilants tend to exhibit a higher degree of voicing when preceded by a voiced stop. Furthermore, they also confirm that HNR values are highest in intervocalic contexts: as shown in the lower part of the graph, sibilants preceded by a voiced stop and followed by a vowel display the most elevated HNR values. This pattern underscores how the surrounding phonological environment—particularly the presence of adjacent vowels or voiced segments—fosters greater sibilant voicing at word boundaries.

An examination of the random effects from the LME model indicates that the random intercept for words (i.e., (1 | word)) has a variance of approximately 0.038, suggesting that different lexical items contribute relatively little variability to baseline HNR values. However, the random slope of the following phonological environment by word (i.e., (1 + following environment | word)) shows a larger variance of about 0.754, indicating that some words exhibit a bigger difference in voicing (HNR) depending on whether the sibilant is followed by a pause or a vowel. The correlation between the random intercept and slope is 1.00, suggesting an overlap in how those random effects are being estimated—possibly due to specific lexical properties or data limitations.

In practical terms, these findings imply that while the overall contribution of individual words to HNR baseline (the intercept) is relatively small, how much an item’s final sibilant voicing increases in vowel environments versus pause environments can differ substantially across lexical items. This variability could stem from word-level factors such as phonotactic constraints, word similarity, or lexical frequency.

In light of Exemplar Theory (ET), the results suggest that learners store multiple phonetic representations for words, encompassing varying degrees of voicing in (stop + sibilant) clusters. ET posits that these exemplars reflect different voicing patterns shaped by prior exposure, which may explain the observed variability in voicing rates across words. Although frequency does not significantly affect HNR values, individual lexical items still exhibit distinct voicing tendencies, likely due to the strength and recurrence of particular exemplar-based representations. Indeed, some exemplars capture fully devoiced clusters preceding vowels, illustrating how phonetic details can spread among related words. In the next section, we will explore how these exemplar-driven patterns are further shaped by individual experience, leading to variation across speakers.

### *4.6. Individual Behavior*

It is expected that different individuals exhibit varying rates of sibilant voicing. This is because, according to the ET, interindividual variation in a linguistic system is dynamic and typically unpredictable, given that speakers have individual learning experiences with the language (Bybee, 2010). Consider Figure 9.

From the mixed-effects model, the random intercept variance for participants is approximately 0.59, indicating that some speakers consistently produce higher or lower HNR overall, possibly due to individual voice quality or habitual articulatory settings. More importantly, the random slope variance for the following context (pause vs. vowel) is about 3.21, which underscores that speakers differ substantially in how strongly they voice sibilants in pre-vocalic environments. In other words, some participants exhibit a large boost in HNR before vowels, while others show a smaller or even negligible difference between the two contexts.

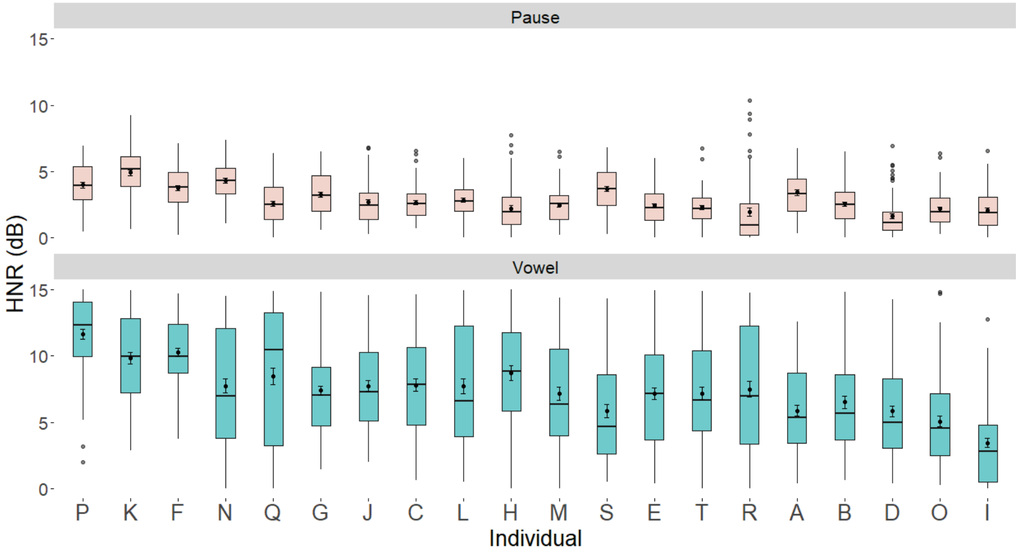



**Figure 9.** Voicing rates per individual.

Looking at the specific HNR ranges in Figure 9 helps illustrate these findings. When the sibilant precedes a pause, HNR values across individuals cluster in a relatively narrow band (approximately 1.6 to 4.9 dB), reflecting the robust stability of the voiceless [s] in BP before a pause. By contrast, when a sibilant is expected to be voiced [z] before a vowel, HNR values span a much wider range (about 3.4 to 11.7 dB). This substantial inter-speaker variability is captured statistically by the large random slope variance mentioned above: some participants produce near-categorical voicing, whereas others barely increase their HNR from the pause baseline, resulting in a rather dynamic pattern of voicing.

For instance, participant P, a 16-year-old from Belo Horizonte, demonstrates one of the highest mean HNR values (11.7 dB) in the vowel context, whereas participant I—also 16 and from Belo Horizonte—shows a much lower mean value (3.4 dB). These extremes highlight how individuals can diverge in their phonetic realization of sibilants, even within the same dialect community.

Taken together, these observations reinforce two key points. First, the voiceless sibilant [s] remains categorically stable before a pause for most speakers. Second, the voiced sibilant [z] before a vowel shows extensive phonetic variability, driven both by the phonological environment and by ongoing processes of devoicing. Within ET (Bybee, 2010), such variability reflects a competition between established patterns (fully voiced clusters before vowels) and emerging realizations (devoiced sibilants)—ultimately underscoring how sound changes may unfold differently across speakers.

## 5. Discussion and Conclusions

This study sheds light on the occurrence of consonant devoicing in Brazilian Portuguese. Our study introduces methodological innovations by assessing sibilant voicing in a gradient manner, utilizing the harmonics-to-noise ratio measurement. This approach allows for the observation of fine phonetic detail, contributing significantly to our understanding of sibilant variation in BP.

Our results indicate that the [z] sibilant, which typically occurs in BP when followed by vowels and voiced consonants (e.g., *cheques amarelos* [ʃɛkz a.ma.ˈɾɛ.lʊs] 'yellow checks'), now exhibits significant rates of pre-vocalic devoicing. These findings suggest that a sound change could be taking place in the Belo Horizonte dialect of Brazilian Portuguese. Sibilant devoicing appears to be more frequent than previously reported in earlier studies (e.g., Cristófaro-Silva, 2003; Bisol, 2005; Seara et al., 2017), indicating a possible shift in this particular dialect. However, our current data do not allow for broader generalizations regarding Brazilian Portuguese as a whole. Additional research is needed to determine whether this devoicing pattern is limited to a synchronic variation in the Belo Horizonte dialect or reflects a more widespread phonological development in BP.

By employing a Linear Mixed-Effects model, we investigated multiple predictors of sibilant voicing, including adjacent phonological contexts, task type, and word frequency. The model incorporated random slopes for speaker and word, capturing variability in how individual participants and lexical items responded to the presence or absence of a following vowel.

Results of the LME model indicate that sibilant voicing depends on both the preceding and following phonological environments, as shown by a significant interaction in the statistical model. In particular, when the following context is a vowel, sibilants preceded by voiceless stops exhibit substantially lower voicing levels than those preceded by voiced stops. Post-hoc pairwise comparisons confirm that these pre-vocalic sibilants maintain a robust voiceless realization when following a voiceless stop, suggesting strong progressive devoicing (e.g., /ps/ $\rightarrow$ [ps] before a vowel).

In contrast, pre-vocalic sibilants following a voiced stop generally exhibit relatively high HNR values, indicating a tendency toward [z]-like voicing—for example, /bs/#V often surfaces as [bz]#V. However, it should be noted that there is still some overlap in HNR values across voiced- and voiceless-stop contexts, so these sibilants are not categorically fully voiced in every token. Although progressive assimilation leading to voicelessness is clearly observed after voiceless stops, there is no comparably strong evidence for devoicing when the sibilant follows a voiced stop; nonetheless, some degree of partial devoicing may occur in certain cases. This discrepancy highlights an asymmetry: a preceding voiceless stop consistently overrides the vowel's regressive voicing effect, whereas a preceding voiced stop tends to preserve higher levels of sibilant voicing before a word-initial vowel—albeit with variability in individual tokens.

Our analysis of task type revealed that visual exposure to the grapheme <s> did not significantly influence participants' production of the sibilant, as no statistically significant differences were found between the reading and picture-naming tasks. Similarly, lexical frequency showed no significant effect on voicing rates.

In analyzing different lexical items, we found that while the overall influence of each word on baseline voicing remained relatively modest, some words displayed substantial shifts in sibilant voicing across pause and vowel environments. This variation likely reflects word-specific properties—such as phonotactic constraints or segmental composition—that can either enhance or suppress voicing in final sibilants.

At the speaker level, participants showed little variability in producing expected pre-pausal voiceless [s] sibilants, but significant variability occurred with expected pre-vocalic voiced [z], as many tokens were devoiced. Speakers also varied considerably in how strongly they voice sibilants before vowels versus pauses. Some showed a large difference (high HNR in vowel contexts), whereas others were more uniform across contexts.

Our findings highlight the competitive dynamics between phonetic variants within the mental lexicon, showing that Exemplar Theory offers an insightful framework for investigating the kind of phonetic variation under study (Johnson, 1997; Pierrehumbert, 2001; Bybee, 2010). Specifically, exemplars associated with traditionally categorical sound patterns—such as the production of voiced clusters followed by vowels—are competing with an emerging sound pattern characterized by the devoicing of word-final clusters. The competition between progressive voiceless assimilation from the preceding stop and regressive voicing assimilation from the following vowel necessitates a dynamic, stochastic model of sound change, as outlined in Exemplar Theory, where phonological representations are continually reshaped by language use and exposure to variable exemplars.

According to Exemplar Theory, fine phonetic detail plays a crucial role in shaping phonological representations. The complex relationship between sibilant voicing and the voicing of their preceding stops in BP suggests an ongoing sound change influenced by multiple phonological and aerodynamic factors, consistent with cross-linguistic findings (De Schryver et al., 2008; Strycharczuk, 2012; Hayes-Harb et al., 2018; Hutin et al., 2020). Moreover, the variability we observed—both across speakers and across lexical items—fits well with the notion that each individual's mental lexicon stores a multitude of exemplars, and that phonological categories are continually updated as speakers encounter both traditional and emerging patterns.

In conclusion, our gradient, variable, and mixed-effects approach provides a richer understanding of consonant devoicing in BP than previously available. By examining sibilant voicing in detail with HNR measures, we demonstrate that devoicing patterns are not limited to a single context but arise from the interplay of multiple phonological and aerodynamic factors. The statistical modeling incorporating random slopes shows that (1) contextual (preceding and following) environments remain the strongest predictors of devoicing, and (2) individual and lexical variation are substantial, suggesting that phonological change may be spreading unevenly through the speech community and the lexicon.

Future research could examine whether devoicing extends beyond word-final sibilants to other segments, including the occasional devoicing of preceding stops observed in this paper through spectrographic analyses (e.g., *redes argentinas* produced as [ˈhets ah.ʒẽ.ˈtʃi.nəs] rather than [ˈhedʒiz ah.ʒẽ.ˈtʃi.nəs]). Additionally, it would be important to investigate how the presence of a following voiced consonant might influence such patterns, in order to provide a more complete picture of devoicing in Brazilian Portuguese. Investigating additional cluster types (e.g., fricative + stop, nasal + stop, or liquids + sibilants) and dialects of BP could further clarify whether these patterns reflect a broader phenomenon, and shed light on the role of fine phonetic detail in shaping phonological representations.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** This research was approved by the Ethics Committee of the Federal University of Minas Gerais (Universidade Federal de Minas Gerais) under reference number 15116119.9.0000.5149, on 20 August 2019.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author (restrictions due to ethical considerations apply).

**Acknowledgments:** I am grateful to the anonymous reviewers of *Languages*, the Editor of this special issue, and the journal's editorial team for their valuable comments and constructive criticism. My heartfelt thanks to Thaïs Cristófaro Silva, whose insights significantly enriched the theoretical discussions in this paper. I also extend my appreciation to the organizers of the 75th StuTS Conference at the University of Graz, Austria, where I presented an earlier version of this article in May 2024. Additionally, I acknowledge the contributions of my team members from the Laboratory of Applied Linguistics at the University of Groningen and the *Laboratório de Fonologia* at the Federal University of Minas Gerais for their indispensable support in this research.

**Conflicts of Interest:** The author declares no conflict of interest.

## Appendix A

Table A1 summarizes fixed and random effects for the linear mixed-effects model predicting HNR, with "picture-naming" as the baseline task, "pause" as the following context, and "vowel" as the preceding context. *p*-values, calculated with *lmerTest* degrees of freedom. The following significance codes are used: *** ($p < 0.001$), "n.s." (not significant), and "marg." (marginal).

**Table A1.** LME model summary.

| Fixed Effect | Estimate | Std. Error | t-Value | *p*-Value |
|---|---|---|---|---|
| (Intercept) | 2.9461 | 0.3329 | 8.850 | <0.0001 *** |
| Task type: reading | 0.1287 | 0.1245 | 1.034 | 0.3015 (n.s.) |
| Following context: vowel | 5.6177 | 0.1953 | 28.764 | <0.0001 *** |
| Preceding context: voiced consonant | −0.1538 | 0.2375 | −0.648 | 0.5170 (n.s.) |
| Preceding context: voiceless consonant | −0.2364 | 0.2347 | −1.007 | 0.3143 (n.s.) |
| Word frequency | −0.1087 | 0.1127 | −0.964 | 0.3350 (n.s.) |
| Following context: vowel: Preceding context: voiced consonant | 0.4839 | 0.2935 | 1.649 | 0.0994 (marg.) |
| Following context: vowel: Preceding context: voiceless consonant | −2.1465 | 0.2790 | −7.694 | <0.0001 *** |
| Random Effect (By word) | **Variance** | **Std. Dev.** | **Corr.** | |
| (Intercept) | 0.03793 | 0.1948 | – | |
| Following context: vowel | 0.75373 | 0.8682 | +1.00 | |
| Random Effect (By speaker) | **Variance** | **Std. Dev.** | **Corr.** | |
| (Intercept) | 0.59196 | 0.7694 | – | |
| Following context: vowel | 3.21057 | 1.7918 | +0.10 | |

Table A2 summarizes Bonferroni-adjusted pairwise contrasts, comparing preceding and following context levels. Negative estimates indicate lower HNR (less voicing) in the first condition. The following significance codes are used: *** ($p < 0.001$), "n.s." (not significant), and "marg." (marginal).

**Table A2.** Post-hoc pairwise comparisons (Bonferroni-adjusted).

| Contrast | Estimate | SE | df | t-Ratio | *p*-Value |
|---|---|---|---|---|---|
| Pause ([i] vowel baseline)–Vowel ([i] vowel baseline) | −5.6177 | 0.195 | 2783 | −28.739 | <0.0001 *** |
| Pause [i]–Pause, voiced C | 0.1538 | 0.239 | 2665 | 0.645 | 1.0000 (n.s.) |
| Pause [i]–Vowel, voiced C | −5.9479 | 0.226 | 2639 | −26.374 | <0.0001 *** |
| Pause [i]–Pause, voiceless C | 0.2364 | 0.237 | 1614 | 0.998 | 1.0000 (n.s.) |
| Pause [i]–Vowel, voiceless C | −3.2348 | 0.234 | 1611 | −13.826 | <0.0001 *** |
| Vowel [i]–Pause, voiced C | 5.7715 | 0.231 | 2509 | 24.939 | <0.0001 *** |
| Vowel [i]–Vowel, voiced C | −0.3301 | 0.224 | 2441 | −1.471 | 1.0000 (n.s.) |
| Vowel [i]–Pause, voiceless C | 5.8542 | 0.225 | 1737 | 26.036 | <0.0001 *** |
| Vowel [i]–Vowel, voiceless C | 2.3829 | 0.225 | 1759 | 10.569 | <0.0001 *** |
| Pause voiced C–Vowel, voiced C | −6.1016 | 0.213 | 2771 | −28.658 | <0.0001 *** |
| Pause voiced C–Pause, voiceless C | 0.0827 | 0.251 | 595 | 0.329 | 1.0000 (n.s.) |
| Pause voiced C–Vowel, voiceless C | −3.3886 | 0.253 | 631 | −13.372 | <0.0001 *** |
| Vowel voiced C–Pause, voiceless C | 6.1843 | 0.250 | 577 | 24.769 | <0.0001 *** |
| Vowel voiced C–Vowel, voiceless C | 2.7131 | 0.250 | 594 | 10.844 | <0.0001 *** |
| Pause voiceless C–Vowel, voiceless C | −3.4713 | 0.195 | 2751 | −17.782 | <0.0001 *** |

## Notes

1 Our analysis assumes that the final alveolar fricative in plural forms is underlyingly voiceless (/s/). In contexts where regressive voicing assimilation is expected—before a vowel or a voiced consonant—traditional accounts predict a fully voiced output ([z]). However, in these pre-vocalic environments, voicing may fail to occur, resulting in a voiceless realization. While this can be framed as the absence of expected assimilation rather than the active devoicing of a previously voiced segment, we view these as complementary perspectives of the same emergent phenomenon.

2 In these contexts, the historically expected epenthesis of [i] frequently fails to apply (See Section 2).

3 Cristófaro-Silva (2003) and Bisol (2005) discuss sibilant voicing patterns in BP without instrumental or acoustic measurements, relying on phonetically trained observation. Seara et al. (2017) similarly do not quantify the extent of voicing, but rather note that the default realization of the word-final sibilant before a vowel is voiced. While we acknowledge that these earlier works do not offer numerical evidence to confirm a near-100% voicing rate, their categorical impressionistic accounts strongly suggest that any devoicing—if present—was negligible or unreported.

4 Alveopalatal affricates [tʃ, dʒ] may occur in BP when followed by a high-front vowel, reflecting a palatalization process: *tia* [tia] ~ [tʃia] (aunt), *dia* [dia] ~ [dʒia] (day). When the high-front vowel is deleted, a stop + sibilant cluster is produced instead, reflecting the situation addressed in this paper.

5 Acknowledging the [Cs] ~ [Cis] alternation is relevant in this context, as the absence of the high-front vowel in [Cs] results in a consonantal cluster that promotes /s/ devoicing, whereas vowel retention in [Cis] potentially disrupts the cluster formation, thereby weakening the conditions for devoicing.

6 Linking Meneses' (2012, 2016) insights on the devoicing and eventual deletion of high-front vowels to sibilant devoicing highlights a potential shared phonetic tendency in BP: the weakening of segmental content in post-tonic environments and in highly coarticulated speech contexts.

7 The 15–17 age group was chosen to capture potential linguistic changes in progress, as younger speakers are often at the forefront of sound shifts (Labov, 1994).

8 Participants were recruited from the same metropolitan area to ensure dialect consistency and minimize regional linguistic variation.

9 In the Portuguese dialect examined in this study (Belo Horizonte), the insertion of a high-front vowel between a stop consonant and a sibilant (e.g., [pis], [tis], [kis], [bis], [dis], and [gis]) occurs, albeit infrequently (cf. Mendes, 2023). This infrequency makes the dialect ideal for studying vowelless realizations. While the primary focus of this research is consonant devoicing rather than vowel insertion (henceforth [i]-occurrence), it is relevant to acknowledge the variability in BP vowel production.

10 Distractors consisted of 36 singular nouns that did not have a consonant cluster in word-final position: *abelha* (bee), *avenida* (avenue), *bambu* (bamboo), *banana* (banana), *batata* (potato), *bingo* (bingo), *bolo* (cake), *brinquedo* (toy), *cadeira* (chair), *caminho* (path), *caneta* (pen), *carteira* (wallet), *cobra* (snake), *copo* (cup), *corvo* (crow), *estátua* (statue), *família* (family), *festa* (party), *flecha* (arrow), *foto* (photo), *gato* (cat), *gravata* (tie), *lago* (lake), *lenço* (handkerchief), *logotipo* (logotype), *menino* (boy), *mesa* (table), *metrô* (metro), *mochila* (backpack), *pizza* (pizza), *sapo* (frog), *sapato* (shoe), *sofá* (sofa), *tornado* (tornado), *torta* (pie) and *vulcão* (volcano).

11 While the literature on BP reports that both voiced and voiceless consonants as following environments can influence the voicing of final sibilants (Bisol, 2005; Seara et al., 2017), these contexts were not included in the present experiment. Future studies could extend this analysis by examining the effects of voiceless and voiced consonants.

12 It should be noted that stop devoicing—apparently triggered by the following voiceless sibilant—constitutes a separate phenomenon of regressive voiceless assimilation. Since this process was not part of our original research questions and was not included in our statistical model, we present it here only as an anecdotal observation that warrants further investigation.

## References

Albuquerque, J. (2011). Dessonorização terminal: Dados de informantes paranaenses falantes de inglês como L2. *Revista l@ el em (dis-) curso*, *3*(1), 54–69.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. [CrossRef]

Bisol, L. (Ed.). (2005). *Introdução a estudos de fonologia do português brasileiro*. EdiPUCRS.

Bjorndahl, C. (2022). Voicing and frication at the phonetics-phonology interface: An acoustic study of Greek, Serbian, Russian, and English. *Journal of Phonetics*, *92*, 101136. [CrossRef]

Blevins, J. (2006). *Phonetically-based sound patterns: Typological tendencies or phonological universals*. Available online: https://www.degruyter.com/document/doi/10.1515/9783110224917.2.201/pdf?licenseType=restricted (accessed on 3 March 2025).

Boersma, P., & Weenink, D. (2023). *Praat: Doing phonetics by computer* [Computer software]. Phonetic Sciences, University of Amsterdam. Available online: http://www.praat.org/ (accessed on 3 March 2025).

Broselow, E. (2018). Laryngeal contrasts in second language phonology. In *Phonological typology* (pp. 312–340). De Gruyter Mouton.

Bybee, J. (2001). *Phonology and language use*. Cambridge University Press.

Bybee, J. (2010). *Language, use and cognition*. Cambridge University Press.

Carbonell, J. F., & Llisterri, J. (2009). Catalan. *Journal of the International Phonetic Association*, *22*(1–2), 53–56. [CrossRef]

Collischonn, G. (1996). Um estudo da epêntese à luz da teoria da sílaba de Junko Ito (1986). *Letras de hoje*, *104*(1), 149–158.

Cristófaro-Silva, T. (2003). *Fonética e fonologia do português: Roteiro de estudos e guia de exercícios*. Cortez.

Cristófaro-Silva, T. (2016). *Trajetórias fonológicas: Evolução e complexidade* (pp. 215–229). Revista LinguíStica—Revista do Programa de Pós-Graduação em Linguística, Special Volume. Available online: http://www.letras.ufmg.br/padrao_cms/documentos/profs/thaiscristofaro/2016_ART_Trajetorias_fonologicas_evolucao_e_%20complexidade.pdf (accessed on 25 February 2025).

Cristófaro-Silva, T., & Leite, C. (2015). Padrões sonoros emergentes: (oclusiva alveolar + sibilante) no português brasileiro. *Caderno de Letras*, *24*, 15–36. [CrossRef]

Cristófaro-Silva, T., & Mendes, W., Jr. (2022). Plural formation in English: A Brazilian Portuguese case study. In *Second language pronunciation: Different approaches to teaching and training*. De Gruyter Mouton.

Davies, M., & Ferreira, M. (2016). *Corpus do português: Web/dialects*. Available online: http://www.corpusdoportugues.org/web-dial (accessed on 3 March 2025).

De Schryver, J., Neijt, A., Ghesquière, P., & Ernestus, M. (2008). Analogy, frequency, and sound change. The case of Dutch devoicing. *Journal of Germanic Linguistics*, *20*(2), 159–195. [CrossRef]

Gradoville, M. (2011). Voiceless sibilant devoicing in spontaneous American English. *Journal of Phonetics*, *39*(4), 550–564.

Hamann, S., & Sennema, A. (2005). Analysis and resynthesis of sibilant fricatives: Perceptual evaluation of devoicing. *Journal of the Acoustical Society of America*, *118*(3), 1668–1677.

Hayes-Harb, R., Brown, K., & Smith, B. (2018). Orthographic input and the acquisition of German final devoicing by native speakers of English. *Language and Speech*, *61*, 547–564. [CrossRef] [PubMed]

Hutin, M., Niculescu, O., Vasilescu, I., Lamel, L., & Adda-Decker, M. (2020, May 11–12). *Lenition and fortition of stop codas in Romanian*. 1st Joint SLTU and CCURL Workshop (SLTU-CCURL 2020)-LREC, Marseille, France.

Jatteau, A., Vasilescu, I., Lamel, L., & Adda-Decker, M. (2019a, August 5–9). *Final devoicing of fricatives in French: Studying variation in large-scale corpora with automatic alignment*. 19th International Congress of Phonetic Sciences (pp. 295–299), Melbourne, Australia.

Jatteau, A., Vasilescu, I., Lamel, L., Adda-Decker, M., & Audibert, N. (2019b). *"Gra [f] e!" Word-final devoicing of obstruents in standard French: An acoustic study based on large corpora*. ISCA.

Jesus, L. M., & Shadle, C. H. (2002). A parametric study of the spectral characteristics of European Portuguese fricatives. *Journal of Phonetics*, *30*(3), 437–464. [CrossRef]

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In J. L. Morgan, & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 121–134). Erlbaum.

Labov, W. (1994). The study of nonstandard English. In *Language* (pp. 555–562). Macmillan Education UK.

Maniwa, K., Jongman, A., & Wade, T. (2009). The role of consonant voicing in distinguishing English word-final sibilants: An acoustic and perceptual study. *Journal of the Acoustical Society of America*, *125*(3), 1694–1704.

Mendes, W., Jr. (2023). *Variabilidade fonética na produção de formas de plural do Inglês como Segunda Língua = Phonetic Variability in the production of plural forms of English as a Second Language* [Doctoral dissertation, Universidade Federal de Minas Gerais].

Meneses, F. (2012). *As vogais desvozeadas no Português Brasileiro: Investigação acústico-articulatória = Devoiced vowels in Brazilian portuguese: An acoustic-articulatory investigation* [Doctoral dissertation, Universidade de Campinas].

Meneses, F. (2016). *Uma visão dinâmica dos processos de apagamento de vogais no português brasileiro = A dynamic account of vowel deletion processes in Brazilian Portuguese* [Doctoral dissertation, Universidade de Campinas].

Myers, S. (2012). Final devoicing: Production and perception studies. In *Prosody matters: Essays in honor of Elisabeth Selkirk* (pp. 148–180). University of Toronto Press.

Ohala, J. J. (1997, August 12). *Aerodynamics of phonology*. The Seoul International Conference on Linguistics (vol. 92, p. 97), Seoul, Republic of Korea.

Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In J. Bybee, & P. Hopper (Eds.), *Frequency effects and the emergence of linguistic structure* (pp. 137–157). John Benjamins.

R Core Team. (2022). *R: A language and environment for statistical computing* [Software]. R Foundation for Statistical Computing. Available online: https://www.R-project.org/ (accessed on 3 March 2025).

Seara, I. C., Nunes, V. G., & Lazzarotto-Volcão, C. (2017). *Para conhecer fonética e fonologia do português brasileiro*. Editora Contexto.

Silva, A. (2022). *Compute fricative features: A script for automating tasks* [Python script]. GitHub. Available online: https://github.com/adelinocpp/Compute_Fricative_Features (accessed on 3 March 2025).

Soares, V. H. (2016). *Encontros consonantais em final de palavra no português brasileiro* [Master's thesis, Universidade Federal de Minas Gerais].
Strycharczuk, P. (2012). *Phonetics-phonology interactions in pre-sonorant voicing* [Doctoral Thesis, The University of Manchester].
Westbury, J. R., & Keating, P. A. (1986). On the naturalness of stop consonant voicing. *Journal of Linguistics*, *22*(1), 145–166. [CrossRef]

languages

MDPI

*Article*

# Buenas no[tʃ]es y mu[ts]isimas gracias: A Sociophonetic Study of the Alveolar Affricate in Peninsular Spanish Political Speech

**Matthew Pollock**

Department of English and Foreign Languages, Louisiana State University Shreveport, Shreveport, LA 71115, USA; matthew.pollock@lsus.edu

**Abstract:** While variation in the southern Peninsular Spanish affricate /tʃ/ has been considered in the context of deaffrication to [ʃ], this study examines an emergent variant [ts] in the context of sociolinguistic identity and style in political speech. Based on a corpus of public speech from Madrid and Andalusia, Spain, this study examines the phonetic and sociolinguistic characteristics of the affricate, finding variation in the quality of the frication portion of the segment through an analysis of segment duration (ms), the center of gravity (Hz), and a categorical identification of realization type. The results suggest that both linguistic variables, like phonetic environment, stress, lexical frequency, and following vowel formant height, as well as extralinguistic variables, like speaker city, gender, political affiliation, and speech context, condition use. Based on these findings, it appears that production of the alveolar affricate [ts] is an incipient sociolinguistic marker in the process of acquiring social meaning. It is particularly associated with female speech and prestige norms that transcend regional identification. This alveolar variant serves as an additional sociolinguistic resource accessible for identity development among politicians and offers insight into ongoing change in the affricate inventory of southern and northern-central Peninsular Spanish.

**Keywords:** Peninsular Spanish; sociophonetics; affricate; variable use; political speech; fronting; Andalusia; center of gravity

## 1. Introduction

Political speech offers insight into the agency and intentions of politicians, while at the same time highlighting resources that speakers can employ as part of their process of identity construction. The indexicality vested in certain terms is reflected in the persona that these individuals develop on a national stage. Given the public nature of their language use, the ways in which politicians express solidarity, converge with certain groups and diverge from others is often subject to close scrutiny, offering insight into the meaning of these differing forms. The current study selects a particular sociolinguistic phenomenon that undergoes a process of phonetic differentiation in the speech of certain politicians, namely the Spanish affricate /tʃ/, in the productions of politicians from both Madrid and Andalusia, Spain. This sociophonetic analysis does not only focus on the linguistic and extralinguistic factors conditioning variable use: it also performs an acoustic analysis on an emergent variant, the alveolar [ts], that has received little mention in descriptions of Peninsular Spanish (e.g., Samper-Padilla 2011) and only recent attention in perceptual research.[1]

Sociolinguistic style develops not just from the words speakers use, or the phonetic forms that they select, but also from the cloud of culturally charged items with which they surround themselves. All of these tools for self-representation, termed "bricolage" by Eckert (2008), serve to distinguish individuals in an indexical web of meaning, as described by Silverstein (2003). Specifically, in the realm of political speech, this can mean aligning oneself with an ideological stance or a certain subset of voters by using regional variants associated with the working class or rural speakers to show solidarity (e.g., Hernández-Campoy and Cutillas-Espinosa 2013; Pollock and Wheeler 2022; Holliday 2017), or by contrasting with one's opponents through divergence in form usage (Cruz-Ortiz 2022; Pollock 2023a; Pollock forthcoming). At the same time, politicians navigate personal motivations and party-based norms to develop a coherent identity in the face of political opposition (e.g., Hall-Lew et al. 2012; Hall-Lew et al. 2017; Pollock and Wheeler forthcoming).

Sibilant fricatives and affricates have been studied broadly within Romance (Recasens and Espinosa 2007) and other typologically distinct languages (Gordon et al. 2002). Within Peninsular Spanish, several variable phenomena have been identified that are relevant to the Spanish post-alveolar affricate, although focus is usually placed upon deaffrication in the speech of elderly, rural speakers (e.g., Henríquez-Barahona and Fuentes-Grandón 2018; Herrero de Haro 2017b; Samper-Padilla 2011). Alongside production of /tʃ/ (e.g., in *hecho* [hetʃo] 'made/did') as the deaffricated Andalusian [ʃ] (e.g., [heʃo]; Villena-Ponsoda 2013), researchers have also identified a voiced post-palatal [d̠ʒ] in the Canary Islands (e.g., [hed̠ʒo]; Almeida 2019) and a voiced pre-palatal [ʤ] in Murcia (e.g., [heʤo]; Torrano Moreno 2017). For the most part, however, the prototypical production of middle-aged urban speakers in central and southern Spain has been described as unvaryingly post-alveolar [tʃ] (Hualde 2005). The current study, which examines the use of a voiceless alveolar affricate [ts], also considers the distinction between a sociolinguistic "indicator", which Labov (1972, p. 237) describes as having no social value, and a "marker", which has acquired social meaning and occurs in a stratified way.

In this study, politicians with similar positions were selected in the two largest national political parties in Spain: the left-leaning *Partido Socialista Obrero Español* (PSOE) 'Spanish Socialist Workers' Party' and the right-leaning *Partido Popular* (PP) 'People's Party'. Of the 32 individuals under consideration—eight each from Madrid, Malaga, Cordoba, and Seville—all productions of the affricate /tʃ/ were collected from a 35.8 h corpus of political speech. Sibilant affricates were examined based on their center of gravity (COG, in Hz), following vowel formant locus, auditory identification, and duration (as a percentage of the frication period in comparison with overall segment length), as well as extralinguistic factors including political affiliation, speech context, gender, and region.

This article is organized in the following way. Section 2 discusses existing literature on political speech, particularly in Peninsular Spanish, as well as the state of research into Spanish affricates and fricative-like segments cross-linguistically. This section concludes by posing three phonetic, linguistic, and socially motivated questions. Following that, Section 3 describes the speakers under consideration, the method of acoustic analysis, relevant dependent and independent variables, and the phonetic and statistical analyses employed to examine variation. Next, Section 4 describes the results of the study, including mixed-effects models to determine the factors influencing variable use, as well as the acoustic norms of the Spanish affricate. Section 5 addresses the phonetic and sociolinguistic ramifications of the results, referencing previous research and returning to the questions posed in Section 2. Finally, Section 6 concludes, offering some thoughts for future research and means of expanding the study of both Spanish affricates and sociolinguistic style in political speech.

## 2. Identity, Political Speech, and the Spanish Affricate

### *2.1. Sociophonetics and Political Speech*

Early descriptions of sociolinguistic style by researchers like Labov (1972) cast the concept as attention paid to speech, with naturalistic interviews yielding more vernacular productions, while reading tasks such as word lists showed a higher rate of normative, non-vernacular variants. More recent studies, such as the classifications of identity described by Bucholtz and Hall (2005) and Coupland (2001), have problematized the representation of style as a binary distinction between vernacular and non-vernacular speech. Instead, they cast social meaning as an association of certain forms and speech decisions, part of a web of indexicality that allows speakers to take on the identity of others who are known to use these forms, following the description of Silverstein (2003). Through this process, linguistic style becomes just one aspect of the broader bricolage of a speaker's social persona, as Eckert (2008) describes, contributing to a coherent vision of a speaker as having a consistent social identity and belonging to certain social subgroups.

This understanding of speaker agency emerges in part from earlier descriptions of speaker behavior relating back to audience expectations (e.g., Bell's (1984) audience design theory) and accommodative behavior toward favorable and non-favorable social meanings (e.g., Giles et al.'s (1991) accommodation theory), and connects to a third-wave sociolinguistic approach associated with speaker design theory (e.g., Schilling-Estes 2013). This approach views speakers as having a multitude of linguistic and social goals that are carried out through variable uses of stylistic resources, according to the context in which they find themselves and the social associations pertinent to a specific speech encounter. Politicians, whose audiences and linguistic environments can vary greatly, from close supporters to international listeners and from one-on-one interviews to broadcast audiences of millions, use language to present themselves in a way that builds solidarity and crafts a consistent ideological persona. It is worth mentioning that, when accommodating to listeners, politicians base linguistic decisions not necessarily on the productions of specific individuals, but rather on the idealized norms associated with certain groups of speakers (e.g., Eckert 2008; Coupland 2001).

Numerous English-language researchers have focused on public and political speech as a means of accessing linguistic variation, from Bell's (1984) work with New Zealand news anchors and Coupland's (2001) discussion of a Cardiff disc jockey to more recent work looking at British politicians (e.g., Kirkham and Moore 2016), Scottish politicians (Hall-Lew et al. 2017), American politicians (Hall-Lew et al. 2012), Queen Elizabeth (Harrington 2007), Barack and Michelle Obama (Holliday 2017), Indian-American journalist Fareed Zakaria (Sharma 2018), and even Oprah Winfrey (Hay et al. 2010). However, there has also been work in recent years to probe some of these same sociolinguistic and even sociophonetic questions in Peninsular Spanish. Hernández-Campoy and Cutillas-Espinosa (2010, 2013) analyzed the speech of a former female president of Murcia, Spain, finding that she produced certain regional variants at a rate higher than even working-class rural male speakers, who were most closely associated with vernacular, regional identity in her community. The authors described this as a means of reinforcing her rural, working-class roots despite holding the highest office in her province, showing solidarity as a socialist politician toward her constituents. Hernández-Campoy and Cutillas-Espinosa (2010) argued that left-leaning Spanish politicians tended to use vernacular variants more than conservatives, a finding that Pollock and Wheeler (2022) also identified for a former female president in Andalusia, with respect to the elision of intervocalic /d/ in past participles.

However, other Peninsular Spanish research has offered less definitive evidence for these tendencies than Hernández-Campoy and Cutillas-Espinosa's work. For example, in a study of Andalusian television presenters on social media, Fernández de Molina Ortés (2020) tracked the way that users' identity work was interpreted by commenters. Of the four presenters studied, those with consistent identities and regional variant use received the most positive perceptual classifications, while dialect-switching was critiqued as a means of showing disloyalty to one's region. In an examination of political speech production and perceptual norms in Andalusia and Madrid, Pollock (2023b) described an overall set of tendencies centered more around individual speaker motivations and goals than stable, party-based behaviors. However, Pollock (2023a) did find a tendency for regional variants to be produced by and perceived as associated with conservative political actors, which may be part of an ongoing shift in Peninsular politics toward populist and alt-right voices using novel linguistic patterns to appeal to rural and working-class voters in the south of Spain. At the level of forms of address and politeness, Pollock (forthcoming) further supports these findings by showing the extent to which impoliteness behavior increased on Spanish social media leading up to and during the COVID-19 pandemic. These studies suggest the importance of regional variants in political speech, due at least in part to the type of identity associated with these forms, as well as a complicated political reality in which linguistic and behavioral norms are in flux.

There is strong evidence to suggest that political actors are governed by the same linguistic expectations and norms as members of the broader speech community (e.g., Hernández-Campoy and Jiménez-Cano 2003). However, as Pollock (2023a) found in a perceptual analysis of regional Andalusian forms, potential voters are more likely to rate voices negatively once they are aware that they come from politicians. These speakers are at times subjected to unique norms of interpretation, which can also be reflected in the way they opt to speak. As Cruz-Ortiz (2022) described, there are certain norms in the Andalusian community associated with political speech, most particularly the prevalence of intervocalic /d/ deletion. She determined that politicians are much more likely to elide /d/ than even the most traditionally vernacular speakers in a given community. This tendency is also supported by the findings of Pollock and Wheeler (2022) for the former Andalusian president, Pollock (2023a) for a larger cohort of Madrid and Andalusian politicians, and Pollock and Wheeler (forthcoming) for a cohort of politicians in Galicia, Spain. These studies identify norms associated with an identity as a political actor in Spain that, at times, transcend both regional and normative speech expectations.

### *2.2. The Spanish Affricate and Acoustic Measures of Frication*

Since the fifteenth century, there has been a marked reduction in sibilants within the Spanish inventory. While early Spanish had a robust eight-part sibilant system with six fricative and as many as two affricate components (i.e., /ts/, /dz/, /s/, /z/, /ʃ/, /ʒ/, /ʤ/, and /tʃ/), this had reduced to six components by the 16th century (i.e., s̺, z̺, /ʃ/, /ʒ/, /ʤ/, and /tʃ/) with the collapse of the alveolar/dento-alveolar fricative distinction. By the end of the seventeenth century, this had further reduced to two components (i.e., s̺ and /tʃ/) as voicing distinctions were lost and the palatal fricative merged with /h/ (Penny 2002, p. 130; Bradley and Lozano 2022). As a result, modern Spanish has no phonemic voicing contrasts among sibilant fricatives and affricates (and, with respect to the occlusive portion of the affricate segment, word-initial voiced stops are produced with negative voice onset time (VOT) and voiceless obstruents have minimal VOT without aspiration, as described by Abramson and Lisker (1972), among others).

This historical reduction in systematicity in the sibilant range may offer the opportunity for increased variability in productions of the Spanish affricate. Speakers who alter their production of the affricate may have a reduced risk of being misinterpreted for producing an unexpected affricate or sibilant. This, in turn, offers an opportunity for the development of social meaning. As Labov (1972) described, linguistic features begin as indicators used in the same way across speakers. However, as they experience stratification (and thus an acquisition of indexical meaning), they become markers that can point to a speaker's social class, gender, or even style, as speakers make use of the connection between the marker and its social associations.

Generally, the Spanish affricate is described in phonetic and phonological texts as /tʃ/, a segment with a prototypically voiceless, post-alveolar production (although note that Hualde and Colina (2014, p. 30) reference the possibility of production as a fricative [ʃ] or alveolar affricate [ts]). Almeida (2019) also described the length of the occlusive and fricative sub-components of the segment as normally being comparable. There is some discussion of variable production across Spanish-speaking communities. Deaffrication, well-documented in southern Spain, seems to be on the decline, particularly given its association with an older generation of speakers (e.g., Villena-Ponsoda 2013; Regan 2020). While focus in the current article centers on Peninsular Spanish, other work has also shown that reduction and deaffrication exists across the Spanish-speaking world (e.g., Díaz-Campos et al. 2023; Mazzaro 2022). In the Canary Islands, studies have shown that voiced post-palatal variants with increased tongue–palate contact can occur (Almeida 2019), whereas, in the southwest of Spain, a voiced pre-palatal [ʤ] is produced by mainly older Murcian speakers (Torrano Moreno 2017). Additionally, Pollock (2023a, 2023b) identified the presence of [ts] as a realization of the Spanish affricate in Andalusian and Madrid varieties that deserves further examination, and work by Del Saz, Vida-Castro and others identified overlap between fronted variants of the alveopalatal affricate and the production of a [$t^s$] affricate in contexts of /st/ cluster reduction (see Footnote 1).

Outside of the peninsular context, there are also references to the production of an alveolar affricate. Flores (2018) is one of several authors to describe this variant in Chile. In her examination of public radio speech, she identifies the variant as being associated with higher-prestige contexts and occurring in more frequent lexical items than the normative post-alveolar variant. Notably, these two allophones show not just audible differences, but also clear acoustic ones. When Flores (2014) compares the two, the segment of frication in alveolar variants has an average starting frequency concentration around 7000 Hz, while frication in normative post-alveolar productions starts closer to 5000 Hz. The presence of an alveolar affricate, attested in other dialectological accounts of Chilean Spanish, is often described as the result of contact with Indigenous languages like Huilliche in the region, which have a contrastive distinction between /tʃ/ and /t̪s̪/ (Henríquez-Barahona and Fuentes-Grandón 2018).

Much discussion of the Peninsular Spanish affricate has focused on deaffrication, particularly that found in the speech of elderly, rural speakers in parts of Seville, Granada, and Malaga (Henríquez-Barahona and Fuentes-Grandón 2018; Herrero de Haro 2017b; Samper-Padilla 2011). Additional linguistic and extralinguistic factors influence the occurrence of this variant. For example, Melguizo-Moreno (2006) found that, among older, working-class men with little education in Granada, the deaffricated variant [ʃ] mostly occurred word-initially. Villena-Ponsoda (2008, p. 148) described a regional divide, with speakers from southwestern parts of Andalusia like Jerez de la Frontera still producing high rates of deaffrication (83% of cases), while stigma and association with elderly rural speech had caused the variant to decline precipitously in northern and eastern parts of the region. According to his findings, fricative production makes up less than a quarter of tokens from speakers in Granada and Malaga.

From a methodological perspective, it is also important to point out how research has offered new means of approaching affricates using a continuous classification system. Phonetic work by researchers like Jongman (1989) and Jongman et al. (2000) showed how certain cues, including spectral, amplitudinal, and temporal ones, permit differentiation between fricative productions. More recently in Spanish, Díaz-Campos et al. (2023) considered affricates from a diachronic corpus of Caracas, Venezuela, dividing the segments into two sub-components, comparing the period of occlusion and that of frication through an analysis of temporal duration. They described the changes they identified over the two-decade span of the corpus as involving a process of retiming. While they did not find complete deaffrication in Venezuela, they did suggest that there is ongoing change moving in that direction. This study offers a way of conceptualizing affricates not through categorical means, but rather through a continuous analysis of their component parts, seeing the extent to which the balanced production of occlusion and frication described by Almeida (2019) in affricates is actually borne out in naturalistic speech.

In order to analyze a potentially alveolar production of the Spanish affricate, the current study turned its focus toward acoustic phonetic examinations of frication to identify the measures most often used to distinguish production norms. For example, in two varieties of Catalan, Recasens and Espinosa (2007) considered fricative and affricate production in an electropalatographic and acoustic analysis. They found that production of the alveolar affricate [ts] differed with respect to segment duration from the alveopalatal [tʃ], being generally the longer of the two segments. Meanwhile, in their study of seven typologically distinct languages, Gordon et al. (2002) used several measures to distinguish alveolar and post-alveolar frication, including segment duration, a spectral slice analysis, and the center of gravity (Table 1).

**Table 1.** Sibilant fricative duration and COG norms (Gordon et al. 2002).

| Language | /s/ Duration (ms) | COG (Hz) | /ʃ/ Duration (ms) | COG (Hz) |
|---|---|---|---|---|
| 1. Aleut | 361.8 | 5219 | - | - |
| 2. Apache | 172.3 | 5461 | 175.9 | 4859 |
| 3. Chickasaw | 123.6 | 5163 | 112.8 | 4679 |
| 4. Gaelic | 130.4 | 4884 | 110.7 | 4396 |
| 5. Hupa | 276.4 | 4797 | 217.7 | 4440 |
| 6. Montana Salish | 171.8 | 4601 | 178.9 | 4134 |
| 7. Toda | 198.3 | 4529 | 239.6 | 4704 |
| **Average** | **204.9** | **4950.6** | **172.6** | **4535.3** |
| standard deviation | 79.4 | 317.5 | 48.3 | 239.2 |

Gordon et al. (2002) also tracked the trajectory of the formant locus in the following vowel to determine what coarticulatory effect might be present as a result of the preceding fricative. The hypothesis was that coarticulatory effects of fricative tongue placement would also appear in the left portion of the following vowel, with a corresponding difference in formant height for the same vowels produced after a more alveolar or palato-alveolar affricate. Other acoustic correlates were also tracked by these authors. For example, across all seven languages, the average duration was found to differ by only 32 milliseconds (ms), with /s/ being generally longer than /ʃ/ (similar to the findings for Catalan mentioned previously). Neither the duration measure nor the spectral slice were able to offer a clear and consistent distinction between production types. Center of gravity (COG), on the other hand, which averages the frequency of a spectral slice and tends to correlate with the place of articulation, was reliable when distinguishing between the fricative segments, with an average 400 Hz distinction between the two.

Other authors have identified it to be useful in the analysis and differentiation of coronal affricate productions, including Vida-Castro and Villena-Ponsoda (2016) for Spanish and Li and Li (2019) for Chinese. For the affricate [ts], Vida-Castro and Villena-Ponsoda found that a lower COG (i.e., 4619 Hz) was more likely to be perceived as the alveolar affricate /tʃ/, while productions with a higher COG (i.e., 6768 Hz and 4995 Hz) were more often perceived as associated with reduction in a word-internal /st/ cluster. This could have resulted from the process of post-aspiration in Spanish, where /st/ is produced as [$t^h$] and is reinterpreted as [ts], following from the pattern of aspiration where coda /s/ spirantizes to [$^h$]. Meanwhile, Li and Li (2019) found consistent differences between gendered productions of /$ts^h$/ (F = 9076 hz, M = 7989 hz) and /tɕ/ (F = 6596 hz, M = 5703 hz) for Mandarin Chinese. Across all of these examples, an /s/-like segment consistently tends to have a higher COG than a further backed post-alveolar or palatal segment. It is for this reason that, while the current study also examines duration and formant locus, COG is the main measure used to distinguish between affricate production norms in Spanish.

*2.3. Research Questions*

Three key questions guided the current analysis of Peninsular Spanish /tʃ/ and its allophonic production as either alveolar [ts] or normative post-alveolar [tʃ]: the first focused on phonetic variation, the second on statistical variation in linguistic and extralinguistic factors, and the third on implications for the sociolinguistic value of the affricate.

First, what variation exists at the phonetic level for the peninsular affricate? Based on acoustic measures, how do realization types differ, and to what extent is an alveolar affricate being produced by political actors?

Second, what variation exists among politicians, and to what extent is it influenced by linguistic and extralinguistic factors? While these are speakers who make performative use of language as a public means of identity construction, it is unclear the extent to which this variation occurs consistently across this population.

Finally, does the alveolar affricate function more as a sociolinguistic indicator or as a full sociolinguistic marker? If there are signs that the alveolar variant has become socially stratified, this suggests that there is a degree of community awareness of this phenomenon, at least insofar as it is associated with the broader speech patterns of a specific social group. If this is the case, it also welcomes future examination of stylistic variation with respect to this phenomenon beyond political speech to determine how widespread this sociophonetic variable is in Andalusian and northern-central Peninsular Spanish.

## 3. Methodology

### *3.1. Speaker Selection*

For this study, 32 politicians were identified who formed a community of practice across southern and central Spain, sharing high-ranking positions in both regional and national governing bodies. Speakers included presidents, senators, and mayors who took part in interviews and speeches broadcast at the regional and national levels. Speakers were balanced by gender, city of origin, and political party. They were born in or near one of four cities—Malaga, Seville, Cordoba, and Madrid—and were affiliated with one of two major national parties—either the left-leaning PSOE or the right-leaning PP. See Supplementary Table S1 for a full list of speakers, including information on their sociolinguistic background and speech context.

Three types of speech situations were recorded in order to examine differences in language use based on the linguistic environments in which politicians found themselves. This included scripted speeches, often delivered before parliamentary bodies, as well as unscripted interviews with male interlocutors and with female interlocutors, both of which often came from broadcasts of major television news shows. Scripted speech was expected to reflect more carefully chosen decisions than those found in extemporaneous interviews, while gendered differences were expected to arise in the interviews based on the interlocutor, as Flores (2014) found for radio speech in Chile. Following previous studies that suggested that the earliest moments in interviews are the least vernacular and the most governed by normative expectations, coding began four minutes into each audio file (Díaz-Campos et al. 2018). Audio clips were 22 min in length on average, yielding 35.8 total hours.

Audio dated between the years 2011 and 2019 was collected, which roughly correlates to the span of time when the PSOE led the regional Andalusian government, and the PP was in charge of national governance and the regional government of Madrid. To be specific, the audio data span two periods of transition: the national PP government came into power in December of 2011 and then lost its control of the national government in June of 2018 following a vote of no confidence. By the end of the timespan considered in these audio data, a major shift had occurred at both the Andalusian and national levels that caused the parties in charge of national and regional governance to swap. As this was the first time that the PSOE lost control of the Andalusian government in over three decades, and as the alt-right Vox party experienced a marked growth in support, these speech data reflect a moment of substantial political change within Spain (Rama et al. 2021).

### *3.2. Acoustic Analysis of the Affricate*

In total, 3175 tokens of the affricate were analyzed based on their production as either the alveolar affricate [ts] or the post-alveolar [tʃ]. All instances of the affricate were coded across each file following the four-minute mark, with the only (rare) exceptions being in cases where an interviewer's speech or an audience's applause overlapped with the speaker's production. Outside of these contexts, affricates were produced clearly, without any cases of deaffrication or ambiguity. For normalization purposes, a comparable number of tokens of /s/ were also coded in various contexts throughout each file. In addition to the center of gravity, spectral examination was included as a means of better differentiating between tokens. On the whole, because broadcast-quality recording equipment was used, background noise was negligible and audio quality was high.

The first step in identifying applicable affricate contexts within the corpus involved the use of the speech-to-text transcriber available through Watson (IBM 2019). This software produced a simple transcript for each file, allowing the identification of all tokens of <ch> and a timestamp. Based on these data, Praat TextGrid files were created for audios that included intervals for each of the 3175 words identified with affricate tokens (Boersma and Weenink 2023). From there, intervals were manually created and affricate segments were identified.

For each token, six acoustic measures were collected. This included (1) the duration of the segment of occlusion and (2) the duration of the segment of frication in each affricate, based on visible wave form and spectrographic patterns. Additionally, both (3) F1 and (4) F2 were collected in vowels directly following the affricate, on the left edge of the vowel segment adjacent to the frication portion of the affricate to determine if there was a coarticulatory influence on this vowel (as examined by Gordon et al. (2002)). This measure was treated as an independent variable, to determine whether the differences Gordon and colleagues identified were also present in southern and northern-central varieties of Peninsular Spanish. This variable was collected in addition to the preceding and following environment, a categorical factor. Next, (5) the COG of the frication portion of each affricate token was measured. Finally, (6) a classification of the affricate was created based on the audible presentation of each segment by the researcher, a native-like speaker of the Peninsular variety of Spanish.

A script was developed in Praat to collect one COG value for the central 50% of aperiodicity in each fricative, as well as the F1 and F2 from the following vowel (collected at a single point, 20% from the leftward boundary to avoid acoustic noise at the edge of the segment), and the duration of tokens of /tʃ/. An example of a segment coded as post-alveolar is shown in Figure 1, while that of a segment coded as alveolar is given in Figure 2—in both cases, the COG and minimum frequency of aperiodic noise are provided (in Hz), showing how a more alveolar production (as found in typologically distinct languages) tends to have a higher COG than more backed ones.

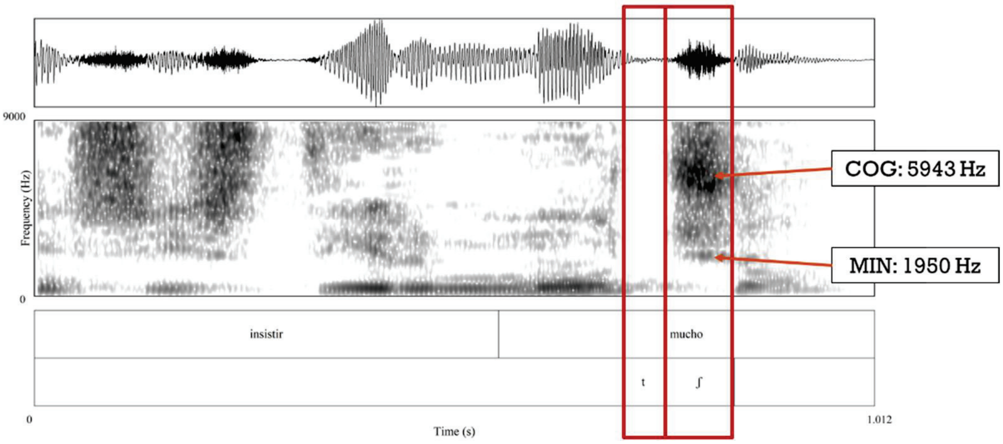



**Figure 1.** Production of *insistir mucho* 'insist a lot' with affricate production as the post-alveolar [tʃ], and a COG of 5943 Hz, with the red boxes showing the occlusion and frication portions of the segment.

Duration was determined, as shown in the red boxes in Figures 1 and 2, by identifying the space between the conclusion of voicing in the preceding segment and the beginning of voicing in the following segment. Both the waveform and spectrogram were consulted to identify the end of the occlusion period (i.e., onset of high-frequency energy in the spectrogram and aperiodicity in the waveform) to begin the segment, and the onset of periodicity of the following vowel to end the segment. COG, on the other hand, was measured by creating an interval around only the densest aperiodicity of the fricative portion of the segment, then capturing the spectral slice of the middle 50% of this segment and identifying a single COG value for each token's frication (i.e., the red box in Figure 1 and the blue box in Figure 2).

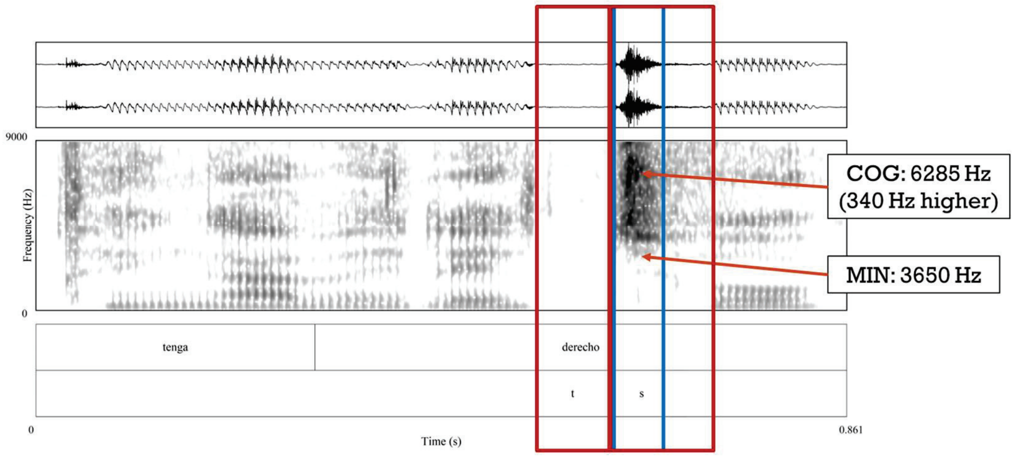



**Figure 2.** Production of *tenga derecho* 'would have the right' with affricate production as the alveolar [ts], and a COG of 6285 Hz, with the red boxes showing the occlusion and frication portions of the segment (duration), and the blue box showing the analysis of COG (Hz) in the frication portion.

### *3.3. Dependent and Independent Variables*

With respect to the dependent variables, two continuous and one categorical measure were used to describe variation in mixed-effects regression models. The continuous measures included the "percent of frication duration", which divided the duration of the frication period of each affricate segment by the overall duration of the segment in an aim to account for speech rate. Additionally, COG (Hz) was examined as a second dependent variable, and was normalized based on the average COG of a speaker's /s/ production (clarified in Section 3.4). Finally, the categorical measure represented the coding of affricate productions as alveolar [ts] or post-alveolar [tʃ], depending on acoustic examination and auditory identification. As no cases of deaffrication were identified, a categorical binary analysis was used.

Five linguistic and four extralinguistic independent variables were included in the analyses. The linguistic environment was examined through the inclusion of both preceding and following context, categorical measures that differed from the continuous F1 and F2 measures of the following vowel. The hypothesis in including these independent variables was that surrounding front vowels may have a coarticulatory effect that would correlate with the production of a more alveolar affricate token. Next, lexical accent was coded based on the tonicity of the syllable with the affricate. Pollock (2023a) identified a role of stylistic variation in the production of the Andalusian affricate—it is hypothesized that productions in tonic position have increased social salience, leading to higher rates of the fronted variant in the categorical analysis of [ts] and [tʃ] as a means of providing emphasis and indexing femininity, youth, and political affiliation. If this is an incipient change, more fronted productions of [tʃ] in tonic position would also be expected.

Next, two continuous linguistic variables were included in the analysis: a measure tracking segment duration and lexical frequency. First, in the models where percent of frication duration was not treated as a dependent variable, it was included as an independent variable to identify influences of frication length. This measure was expected to reflect the findings of Gordon et al. (2002), in that the alveolar segment is typically longer than the post-alveolar one. Subsequently, lexical frequency, coded as the number of times a word occurred in the overall corpus, was expected to identify often-used items that might be subject to higher rates of stylistic variation.

Extralinguistic variables included items coded in supplementary Table S1: a speaker's city of origin, gender, and political affiliation, as well as the speech context of the audio file. Two additional variables were included as random effects to account for assumptions of the mixed-effect model related to the independence of tokens: the speaker and the unique lexeme. The breakdown of variables and possible individual factors are provided in Table 2.

**Table 2.** List of dependent and independent variables, as well as random effects in the analysis.

| Variable Type | Variable Name | Factors | | | |
|---|---|---|---|---|---|
| Dependent variables | Frication period COG (Hz) | *Continuous* | | | |
| | Percent frication duration | *Continuous* | | | |
| | Place of production | *Alveolar [ts]* | *Post-alveolar [tʃ]* | | |
| Independent variables | Preceding context | *Front vowel* | *Central vowel* | *Back vowel* | *Alveolar consonant* |
| | Following vowel type | *Front* | *Central* | *Back* | |
| | Following vowel F1 (Hz) | *Continuous* | | | |
| | Tonicity | *Tonic* | *Atonic* | | |
| | Lexical frequency | *Continuous* | | | |
| | City | *Seville* | *Cordoba* | *Malaga* | *Madrid* |
| | Gender | *Male* | *Female* | | |
| | Party | *Left (PSOE)* | *Right (PP)* | | |
| | Context | *Scripted* | *Unscripted male* | *Unscripted female* | |
| Random effects | Speaker | n = 32 | | | |
| | Lexeme | n = 228 | | | |

*3.4. Data Analysis*

In an effort to account for differences between speakers' articulatory tracts, a means of normalization was necessary for affricate productions. Toda (2007) applied one such technique to reduce interspeaker variation among French and Japanese speakers for fricatives. This author found, through a quantitative analysis, that normalization preserves acoustic differences while reducing individual variation, leading to localized improvements (particularly in French). However, Toda also identified divergent tendencies between vowel categories and sibilant productions that may have resulted from coarticulatory effects. In order to avoid this possible influence, the current study similarly carries out normalization, albeit through the Lobanov (1971) method by comparing the COG of affricate frication with that of speakers' productions of /s/.

This normalization method compares a given COG value ($COG_{n[V]}$) to the average COG of analyzed sibilants for a given audio file ($Mean_N$), then divides by the standard deviation ($S_N$) to acquire a normalized value centered around zero. The formula is provided in (1), as follows:

$$COG_{n[V]}{}^{N} = (COG_{n[V]} - Mean_N)/S_N \quad (1)$$

By including both fricative and affricate productions in the normalization model, the resulting difference in the apico-alveolar frication and post-alveolar frication period of the affricate are identified, while individual speaker differences are reduced. This, in turn, offers a basis for comparison within the articulatory tract of a single speaker. The resulting figures were then rescaled, following an altered version of the NORM standard found in the R vowels package formula (Kendall and Thomas 2010). This employed the formula in (2), where $COG_N$ is the normalized value derived from (1) for a token's COG, and $COG_N$MAX/MIN are the maximum and minimum normalized values for the entire group of speakers, as follows:

$$F'1 = 3000 + 1770\ (COG_N - COG_N MIN)/(COG_N MAX - COG_N MIN) \quad (2)$$

To account for the sibilant COG, 3000 Hz was set as the baseline value, and variance was established as the standard deviation of /tʃ/ and /s/ in the current data (i.e., 1770 Hz). This effectively establishes a maximum scaled COG at 4770 Hz and a minimum at 3000 Hz.

This analysis used both the acoustic measurement of COG in the central 50% of the frication portion of affricates, as well as the binary categorical classification of tokens based on auditory coding with confirmation through visible acoustic measures in the waveform and spectrogram. The resulting data were analyzed using two mixed-effects regression models in the Rbrul software for R (version 4.0.2, Johnson 2009). Stemming from the variable rule approach to sociolinguistics, this type of model includes random effects, which help reduce imbalances across naturalistic data by accounting for dependencies (Johnson 2014).

Regression models offer several important statistics to describe how linguistic and extralinguistic factors influence variation (e.g., Tagliamonte 2012). The *p*-value of a variable indicates the strength of evidence against the null hypothesis, which would suggest that there is no difference between groupings of data, with *p*-values under 0.05 being treated as significant. The Rbrul statistical software offers a *p*-value for each variable. In logistic regressions, directionality can be attributed to an effect based on the factor weight. In linear regressions, this is shown in the coefficient. Factor weights range between 0 and 1, with values above 0.5 favoring use of the application value and those below 0.5 disfavoring it, while coefficients are centered around 0 but follow the same pattern. Finally, the range in factor weights, which reflects a difference in log-odds between factors, serves to indicate the magnitude of effect in logistic regressions. Greater differences between the highest and lowest weight for a single factor, also known as higher factor ranges, suggest that there are greater differences across the factors of a given significant variable. The models produced for each of the two dependent variables identify linguistic and extralinguistic factors that play a role in conditioning variable use. Model selection was carried out through comparisons of the Akaike Information Criterion (AIC) and log-likelihoods, favoring models with fewer variables when possible.

## 4. Results

In total, 3175 allophones of the Spanish affricate were collected for analysis, in addition to 3466 tokens of /s/ to permit normalization by speaker and audio file. On average, this meant that each of the 32 speakers produced around 33 affricates in each of the three speech contexts under consideration (i.e., scripted, unscripted with male interlocutor, and unscripted with female interlocutor). As shown in Table 3, differences were identified between productions of /s/ and the affricate by duration (ms) and COG (Hz). Note that COG values of frication segments of each affricate were subjected to the normalization process described in Section 3.4. Additionally, formant loci for F1 and F2 were collected for following vowels as a continuous measure. A Lobanov normalization was performed on these F1 and F2 values using the NORM standard found in the R vowels package formula (Kendall and Thomas 2010) by first applying the formula in (1) used for COG, and then rescaling following the formula in (2) with the constants adjusted to the values expected for vowels (i.e., F1: 250, 500; F2: 850, 1400; Kendall and Thomas (2010)).

**Table 3.** Average of acoustic measures for the affricate allophones and fricative tokens.

| Production | Count | COG | Scaled, Normalized COG | Duration | Percent Frication | Scaled, Normalized F1 (next vowel) | Scaled, Normalized F2 (next vowel) |
|---|---|---|---|---|---|---|---|
| [tʃ] | 2307 | 3979 Hz | 4012 Hz | 108.9 ms | 50.3% | 418.9 Hz | 1422.5 Hz |
| [ts] | 868 | 4783 Hz | 4081 Hz | 119.9 ms | 51.6% | 467.8 Hz | 1491.2 Hz |
| **All /tʃ/ tokens** | **3175 (33/file)** | **4199 Hz** | **4030 Hz** | **111.9 ms** | **50.7%** | **432.3 Hz** | **1441.3 Hz** |
| All /s/ tokens | 3466 (35/file) | 4610 Hz | - | 86.9 ms | - | - | - |

The average COG across the 3466 tokens of the sibilant fricative was 4609.6 Hz, and the duration was 86.9 ms. Of the 3175 productions of the affricate, 868 tokens were coded as alveolar [ts], which had an average COG of 4783 Hz, a duration of 119.9 ms, and a frication period taking up 51.6% of the segment. The remaining 2307 tokens were classified as post-alveolar with an average COG of 3979 Hz, a duration of 108.9 ms, and a frication period taking up 50.3% of the segment. While the normalization process for the affricate reduced some difference, there was still an over 65 Hz difference in tokens identified as alveolar and post-alveolar. Finally, formant values were only calculated in vowels following the affricate: the average normalized, scaled F1 was 418.9 Hz for vowels following [tʃ] and 467.8 Hz for those following [ts], while the average normalized, scaled F2 was 1422.5 Hz for vowels following [tʃ] and 1491.2 Hz for those following [ts].

With relation to the coarticulatory effects of the frication period of the affricate in the transition period to the following vowel, described by Gordon et al. (2002), the formant values in Table 3 suggest a relatively minimal average difference in vowel quality following the two allophones. Studies of differences in vowel height in English have suggested a perceptual threshold of at least 60 Hz necessary to distinguish between a low and mid vowel (Labov et al. 2013). While there is not an established threshold for vocalic differences in Spanish, Herrero de Herrero de Haro's (2017a) perceptual study of allophones of /e/ before contexts of elided /s/, /r/, and /θ/ suggests that an average difference of at least 60 Hz (and often closer to 100 Hz) is noticeable to listeners. Thus, while the differences in F1 and F2 suggest that preceding alveolar affricates are followed by lower and (very slightly) backed vowels, this difference should be taken with caution until further study can identify the extent to which listeners perceive these differences.

Meanwhile, an examination of the raw COG of /s/ and affricates suggests a trend between types of production of Spanish sibilants, represented in Figure 3. Politicians who produced tokens of /s/ with higher COG frequencies tended to produce both [tʃ] and [ts] with higher COGs as well. For both genders, the normative post-alveolar allophone tends to have a lower COG across most speakers. However, gendered differences also play a role in distinguishing productions, with most female tokens coded as [tʃ] having an average COG frequency similar to that of the male-produced /s/ or [ts]. This underscores the necessity of normalizing COG values to account for these anatomical differences across speakers.

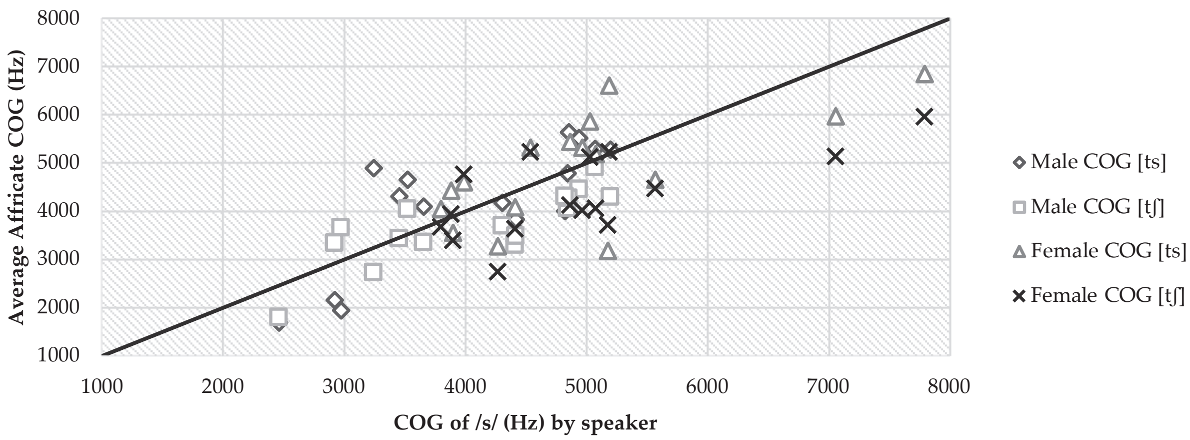



**Figure 3.** Affricate frication period COG by COG of /s/, subdivided by speaker gender and realization type of [ts] or [tʃ], with reference line to compare affricate productions to /s/ values.

A second measure referenced by Gordon et al. (2002) as a means of classifying productions of sibilant fricatives was segment duration. In their comparison, the authors found that alveolar /s/ tended to be longer than post-alveolar /ʃ/. As the current study focuses on affricates rather than fricatives, there may be a reduction in the magnitude of durational differences, as the two manners of production are not identical. However, as shown in Figure 4, the correlation identified by Gordon et al. is borne out in the current data. While the alveolar affricate had a greater overall variance in duration, average productions of this allophone tended to be longer than post-alveolar tokens, with men producing frication in tokens of [ts] on average 18 ms longer than in [tʃ], and women producing frication 9.5 ms longer.

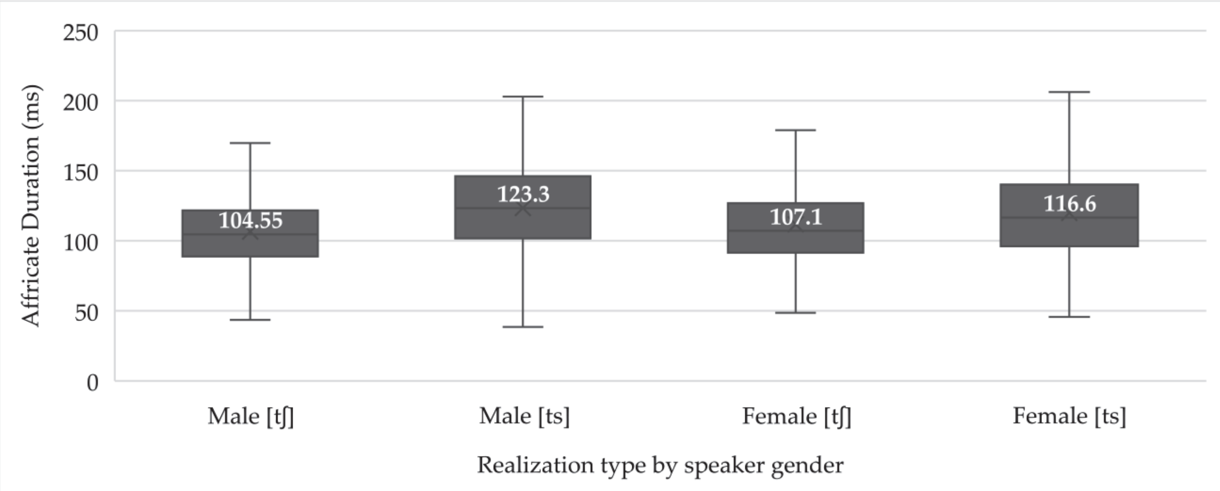



**Figure 4.** Affricate allophone duration (ms) by speaker gender.

Before developing a logistic regression model, descriptive comparisons of social factors in the data were made to identify general tendencies. The first comparison involved the differences in affricate realization type by speaker city and gender (Figure 5). Overall, the figure shows a reduced rate of alveolar affricate production among men, particularly in Malaga, both overall and in comparison to female speakers. While female politicians in Madrid and Cordoba produce around half of their affricates as alveolar [ts], those in Malaga and Seville produced around a quarter in that way, while men had near or below half as frequent a production of the alveolar variant.

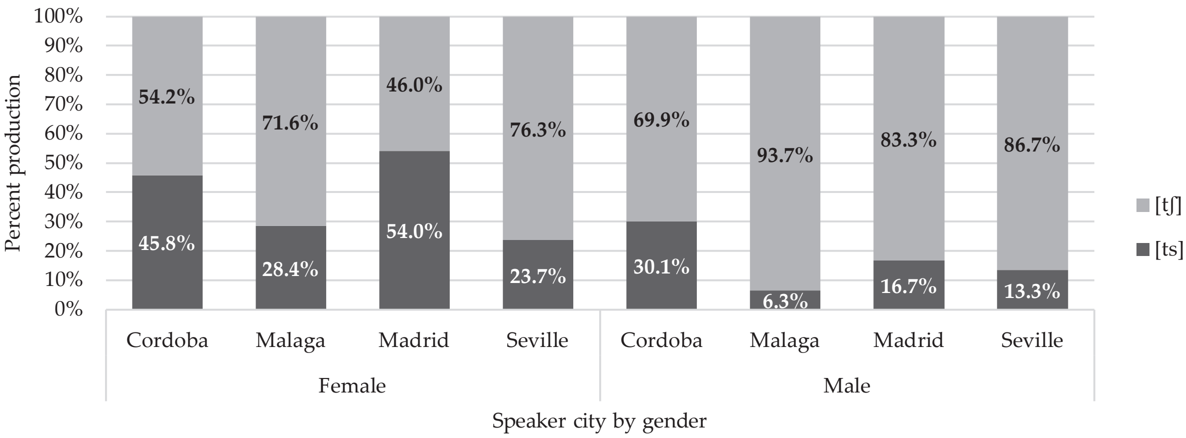



**Figure 5.** Affricate realization by speaker city and gender.

The final comparison involved speakers' political party and city of origin, allowing a consideration of differences based on political affiliation, as referenced in the discussion of Hernández-Campoy and Cutillas-Espinosa (2013), among others (Figure 6). Unlike many previous features discussed in contexts of political speech, both Figures 5 and 6 suggest a minimal effect of regional variation, with speakers from Madrid and the northern Andalusian city of Cordoba grouping together. To be specific, for political affiliation and gender, conservatives and women from these two cities have a particularly high rate of alveolar production. The variant is infrequent in Malaga across genders and political parties, and relatively infrequent in Seville, as well as among most socialist and male politicians.

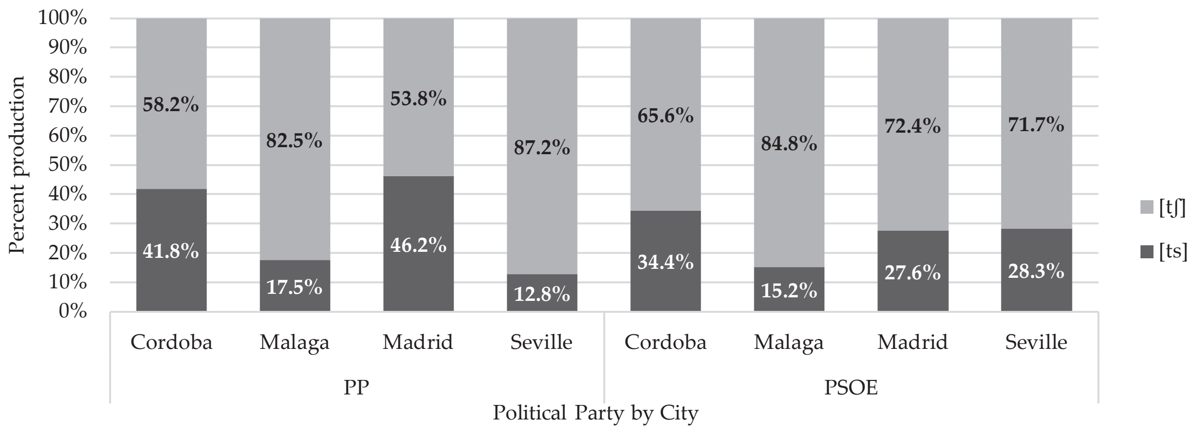



**Figure 6.** Affricate realization by speaker city and political affiliation.

Based on the dependent and independent variables identified in Section 3.3, three models were developed to track variation in the data through three different means: continuously in a mixed-effects linear regression using the percent frication duration in each segment as a method of normalizing duration, followed by the normalized, rescaled COG (Hz) of the frication period of affricates, and then using a binary independent variable in a mixed-effects logistic regression, with segments classified as alveolar frication being treated as the application value. These models considered all linguistic and extralinguistic variables to determine which influenced COG values and affricate selection among this cohort of political actors. The two random effects were also included to account for speaker and lexical variation, as well as interactions between social factors that were theoretically motivated (i.e., the interaction between speaker and interlocutor gender, or those between political affiliation, speaker gender, and city).

In the first linear regression model, presented in Table 4, seven variables were selected as significantly conditioning frication duration in the segment, five of which were linguistic and two extralinguistic. First, the impressionistic realization type followed the trend identified by Gordon et al. (2002), in that [ts] had a longer frication period than the post-alveolar affricate. Keeping this in mind, there was an inverse relationship between duration percentage and preceding and following context—both non-front preceding vowels and consonants favored [ts], while, in following position, front vowels favored it. This suggests a process of movement, with longer frication segments reflecting a shift in the tongue from the back of the mouth or consonant production into a front vowel. While fricative sounds are traditionally seen as resistant to coarticulatory effects (Recasens 2018), duration may reflect the degree of emphasis placed upon production and speak to the indexical social meaning being attached to production norms. More research into the effects of coarticulation in this context is merited. The fourth significant variable, lexical frequency, was continuous, showing that more frequent items tended to favor longer productions. Lastly, the scaled F1 of the following vowel also showed a relationship with duration, in that lower vowel height overall favored longer frication durations.

Among the social factors selected in this model were political party and speech context. In the case of the former, socialists tended to produce affricates with longer frication periods, while, for the latter, scripted speeches were the only context to disfavor longer frication periods. No interactions were selected as significant predictors of variation in this model.

**Table 4.** Mixed-effects linear regression with percent frication duration in the affricate as the continuous dependent variable and speaker and lexeme as random effects.

| **Variable** | **Factor** | **Coefficient** | **Tokens** | **Frication as a % of Segment Duration** |
|---|---|---|---|---|
| Political party ($p = 0.003$) | Socialist (PSOE) | 0.024 | 1428 | 53.5% |
| | Conservative (PP) | −0.024 | 1747 | 48.4% |
| *Range* (%) | | | | *5.1%* |
| Preceding context ($p < 0.001$) | Consonant | 0.014 | 235 | 52.8% |
| | Back vowel | 0.002 | 1783 | 51.5% |
| | Central vowel | 0.001 | 79 | 50.8% |
| | Front vowel | −0.016 | 1078 | 48.9% |
| *Range* (%) | | | | *3.9%* |
| Following vowel ($p < 0.001$) | Front vowel | 0.014 | 502 | 52.3% |
| | Central vowel | −0.006 | 1036 | 50.9% |
| | Back vowel | −0.008 | 1637 | 50.1% |
| *Range* (%) | | | | *2.2%* |
| Realization type ($p < 0.001$) | [ts] | 0.008 | 868 | 51.6% |
| | [tʃ] | −0.008 | 2307 | 50.3% |
| *Range* (%) | | | | *1.3%* |
| Speech context ($p = 0.001$) | Unscripted male | 0.006 | 1047 | 51.1% |
| | Unscripted female | 0.001 | 988 | 51.1% |
| | Scripted speech | −0.006 | 1140 | 50.0% |
| *Range* (%) | | | | *1.1%* |
| Lexical frequency ($p < 0.001$) | Continuous | Coefficient | | |
| | +1 | 0.001 | | |
| Scaled following vowel F1 ($p = 0.010$) | Continuous | Coefficient | | |
| | +1 | 0.001 | | |
| n = 3175; df = 15; Log-likelihood = 3784; AIC = −7538; $R^2$ fixed = 0.221; $R^2$ total = 0.323 | | | | |

In the second linear regression model, displayed in Table 5, eight variables were selected as significantly conditioning increased COG (Hz), including five that were linguistic, three that were extralinguistic, and one interaction. Based on the findings of Gordon et al. (2002), a higher frequency of the COG of a segment was expected to correlate with a more alveolar production. The eight variables in this model are organized in order of the range in COG between their factors, with city coming first. Speakers from Madrid produced the highest rescaled COG in affricates overall, while those in Seville produced the lowest—however, the model described all speakers from Andalusia as producing an affricate that was less likely to have a higher frequency. Next, the model selected the preceding sound, with central and front vowels and alveolar consonants favoring a higher COG, and back vowels being more likely to pair with reduced COG. Following that, the realization coding was found to significantly predict frication segment COG frequency, with coding as an alveolar [ts] favoring a higher COG. After this, the following vowel was selected, with both front and central vowels again favoring a higher COG, while back vowels disfavored it.

The fifth variable selected in the model was tonicity, with tonic contexts favoring a higher COG, showing a potential correlation between the alveolar production and tonicity. Next, the speech context was selected, with scripted and female unscripted speech favoring a higher COG, and unscripted speech with a male interlocutor disfavoring it. Following that, the final extralinguistic factor selected in the model was speaker gender, with men having a higher normalized COG overall than women. The last linguistic factor was the rescaled, normalized F2 of the following vowel. More front vowels (i.e., those with increased F2 values) favored production after the affricate with a higher COG, associating the alveolar variant with coarticulatory effects. Finally, there was an interaction between speaker gender and speech context, with female politicians disfavoring COG raising in speeches with male interviewers, and male politicians disfavoring raising in scripted speeches and when talking with female interviewers, as demonstrated in Figure 7.

**Table 5.** Mixed-effects linear regression with the normalized, rescaled COG value as the continuous dependent variable and speaker and lexeme as random effects.

| Variable | Factor | Coefficient | Tokens | Rescaled COG (Hz) |
|---|---|---|---|---|
| **City** ($p = 0.001$) | Madrid | 63.838 | 793 | 4101.1 |
| | Malaga | −5.442 | 695 | 4032.8 |
| | Cordoba | −14.479 | 717 | 4024.1 |
| | Seville | −43.917 | 970 | 3976.8 |
| *Range* (Hz) | | | | *124.3* |
| **Preceding context** ($p < 0.001$) | Central vowel | 35.486 | 79 | 4110.5 |
| | Alveolar | 8.322 | 235 | 4106.0 |
| | Front vowel | 1.517 | 1078 | 4043.9 |
| | Back vowel | −45.326 | 1783 | 4009.4 |
| *Range* (Hz) | | | | *101.1* |
| **Realization coding** ($p < 0.001$) | [ts] | 26.8 | 868 | 4081.4 |
| | [tʃ] | −26.8 | 2307 | 4011.7 |
| *Range* (Hz) | | | | *69.7* |
| **Following vowel** ($p = 0.010$) | Front | 10.21 | 502 | 4073.0 |
| | Central | 8.4 | 1036 | 4050.1 |
| | Back | −18.61 | 1637 | 4005.7 |
| *Range* (Hz) | | | | *67.3* |
| **Tonicity** ($p = 0.001$) | Tonic | 19.406 | 543 | 4077.4 |
| | Atonic | −19.406 | 2632 | 4021.2 |
| *Range* (Hz) | | | | *56.2* |
| **Speech context** ($p < 0.001$) | Scripted | 16.916 | 1140 | 4047.9 |
| | Female Unscripted | −0.303 | 988 | 4037.7 |
| | Male Unscripted | −16.612 | 1047 | 4005.6 |
| *Range* (Hz) | | | | *42.3* |
| **Gender** ($p = 0.021$) | Men | 20.921 | 1521 | 4041.7 |
| | Women | -20.921 | 1654 | 4020.8 |
| *Range* (Hz) | | | | *21* |
| **Interaction: Speaker Gender * Speech context** ($p = 0.001$) | Male:Scripted | 18.081 | 593 | 4050.0 |
| | Female:Scripted | 13.735 | 547 | 4045.6 |
| | Female:Female Interloc | 4.345 | 542 | 4042.0 |
| | Male:Male Interloc | −4.345 | 482 | 4039.9 |
| | Male:Female Interloc | −13.735 | 446 | 4032.5 |
| | Female:Male Interloc | −18.081 | 565 | 3976.4 |
| *Range* (Hz) | | | | *73.6* |
| **Scaled normalized following vowel F2** ($p = 0.015$) | Continuous | Coefficient | | |
| | 1 | 0.026 | | |
| n = 3175; df = 20; Log-likelihood = −21,322; AIC = 42,683; $R^2$ Fixed = 0.105; $R^2$ Total = 0.148 | | | | |

The final model in Table 6, a logistic regression of realization types, compared productions based on auditory coding reinforced with visible acoustic measurements taken in the waveform and spectrogram. This model included eight factors, including five linguistic and three extralinguistic. As opposed to the model in Table 5, this one had rescaled, normalized COG of frication as an independent variable and the coded realization type as the dependent variable. Variables in this model are organized by their range in factor weights, reflecting the overall magnitude of effect. Speaker gender is first; while male politicians produced higher normalized COGs in affricate frication periods overall, this model shows that female politicians produced more affricates coded as the alveolar [ts]. This disparity merits perceptual analysis to determine if the trend holds among native speakers of Andalusian varieties of Spanish as well—it is unclear whether this is the result of gendered perceptions of language or perhaps the presence of an additional acoustic cue not under consideration in this study that may further inform classification.

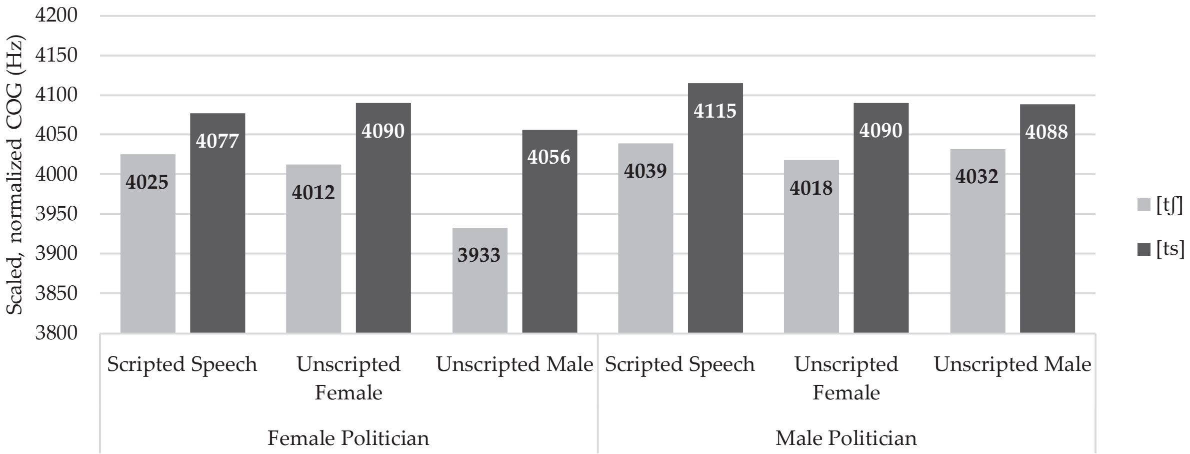



**Figure 7.** Normalized COG of the frication segment in the affricate by speech context and speaker gender.

Several of the remaining variables have similar trends to those identified in the linear regression model in Table 5. For the variables with the second- and third-highest magnitudes of effect, the preceding and following environment, back vowels and central vowels disfavored [ts] in the preceding context, while back vowels alone disfavored it in the following context. Two non-significant main effects included due to their participation in a significant interaction were city and speech context, the latter approaching significance with a *p*-value of 0.085. We see that speakers from southwestern and southeastern Andalusia produced less [ts], while those in Cordoba and Madrid, as well as those in unscripted interviews with women and in scripted speeches, favored [ts]. For the remaining continuous independent variables, affricates with a higher COG, a greater percentage of the segment occupied by frication, and a raised F1 in the following vowel all favored [ts] production.

**Table 6.** Mixed-effects logistic regression of affricate realization type with the alveolar production [ts] as the application value and the speaker and lexeme as random effects.

| Variable | Factor | Coefficient | Tokens | % [ts] | Factor Weight |
|---|---|---|---|---|---|
| **Gender** ($p = 0.001$) | Female | 0.827 | 1654 | 37.7% | 0.696 |
| | Male | −0.827 | 1521 | 16.1% | 0.304 |
| *Range* | | | | | *39.2* |
| **Preceding context** ($p < 0.001$) | Consonant | 0.663 | 235 | 49.4% | 0.660 |
| | Front vowel | 0.329 | 1078 | 30.5% | 0.581 |
| | Central vowel | −0.34 | 79 | 29.1% | 0.416 |
| | Back vowel | −0.652 | 1783 | 22.4% | 0.343 |
| *Range* | | | | | *31.7* |
| **Following vowel** ($p < 0.001$) | Central | 0.335 | 1036 | 35.5% | 0.583 |
| | Front | 0.097 | 502 | 29.3% | 0.524 |
| | Back | −0.432 | 1637 | 21.6% | 0.394 |
| *Range* | | | | | *18.9* |
| ***City*** ($p = 0.270$) † | *Cordoba* | *0.515* | *717* | *38.4%* | *0.626* |
| | *Madrid* | *0.452* | *793* | *37.3%* | *0.611* |
| | *Seville* | *−0.362* | *970* | *18.9%* | *0.411* |
| | *Malaga* | *−0.605* | *695* | *16.4%* | *0.353* |
| *Range* | | | | | *27.3* |
| *Speech context* ($p = 0.085$) † | *Unscripted female* | *0.146* | *988* | *30.2%* | *0.536* |
| | *Scripted speech* | *0.036* | *1140* | *26.8%* | *0.509* |
| | *Unscripted male* | *−0.182* | *1047* | *25.3%* | *0.455* |
| *Range* | | | | | *8.1* |

**Table 6.** *Cont.*

| Variable | Factor | Coefficient | Tokens | % [ts] | Factor Weight |
|---|---|---|---|---|---|
| **Speaker Gender * Speech context** | Female:Scripted | 0.172 | 547 | 0.397 | 0.543 |
| ($p$ = 0.037) | Female:Female Interloc | 0.14 | 542 | 0.382 | 0.535 |
| | Female:Male Interloc | 0.031 | 565 | 0.352 | 0.508 |
| | Male:Female Interloc | −0.031 | 446 | 0.204 | 0.492 |
| | Male:Scripted | −0.14 | 593 | 0.148 | 0.465 |
| | Male:Male Interloc | −0.172 | 482 | 0.137 | 0.457 |
| *Range* (Hz) | | | | | *8.6* |
| **Scaled COG of frication** ($p$ < 0.001) | Continuous | | | | |
| | +1 | 0.001 | | | |
| **Percent frication** ($p$ < 0.001) | Continuous | | | | |
| | +1 | 3.362 | | | |
| **Normalized F1 of following vowel** ($p$ = 0.016) | Continuous | | | | |
| | +1 | 0.017 | | | |
| **City * Following Context** ($p$ = 0.007) | | | | | |
| n = 3175; df = 25; Log-likelihood = −1391; AIC = 2832; $R^2$ fixed = 0.262; $R^2$ = 0.485 | | | | | |

† Note that the two main effects, city and speech context, are included in the table in italics, as they do not have a significant *p*-value, because they form part of a significant interaction (listed at the bottom of the table).

Two additional interactions were identified as significant predictors of variation. The first paired politician gender with the speech context. The correlation is demonstrated in Figure 8. There is a general tendency for unscripted speech with male interviewers to have the lowest rates of [ts], while unscripted speech with female interlocutors has a consistently higher rate, and female politicians are most likely to use [ts] in all contexts. The second interaction involved politicians' city and the following context, provided in Figure 9. While it is the case that a following /i/ and /e/ favor [ts] production in Madrid, the pattern in the Andalusian cities is for a following /a/ to favor [ts], even in Cordoba, where rates of the fronted variant are comparable to those in Madrid. This phenomenon is generally more frequent in the two cities closest to central Spain, even as there appears to be different fronting patterns based on linguistic context that are differentiated depending on whether the city is part of Andalusia.

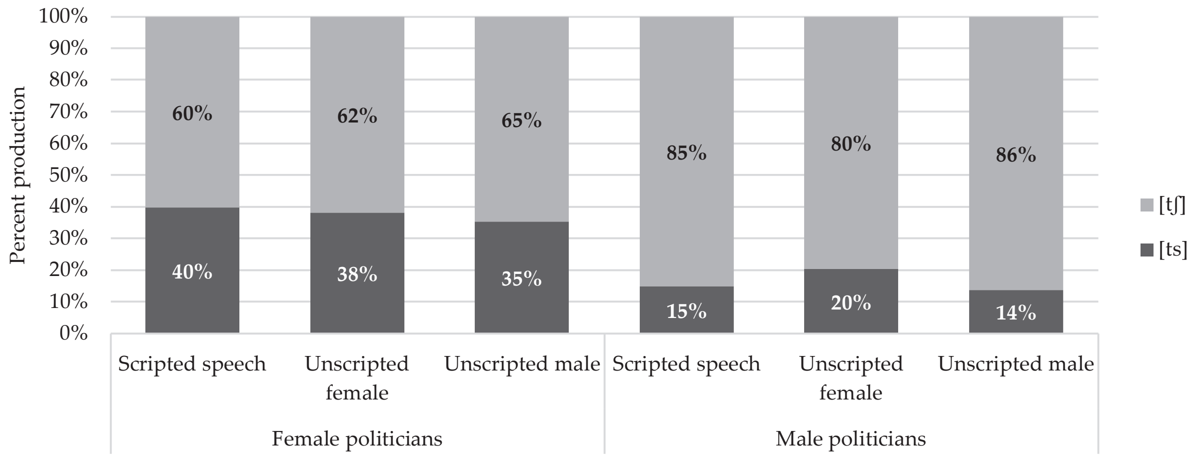



**Figure 8.** Affricate realization type by speech context and speaker gender.

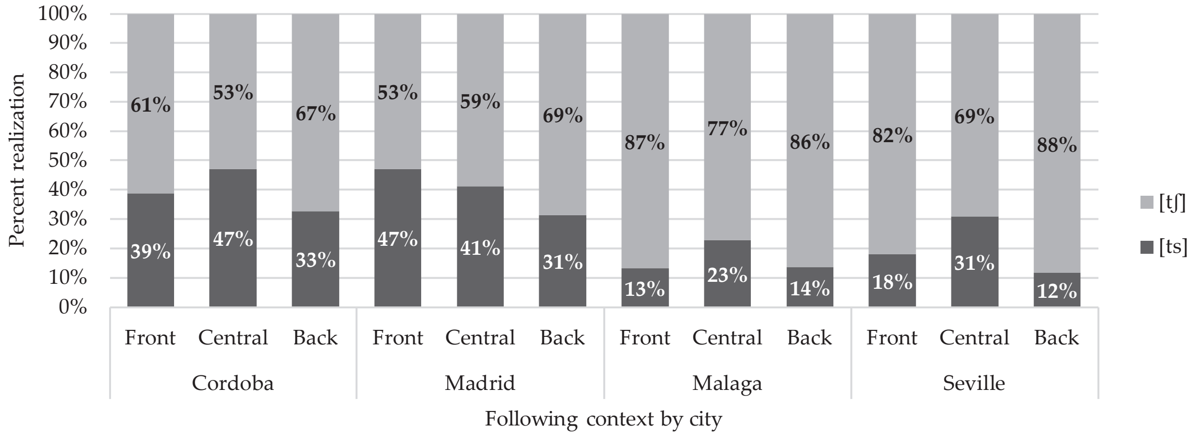



**Figure 9.** Affricate realization type by following context and city.

## 5. Discussion

### *5.1. Phonetic Variation in the Affricate*

This study shows variable production of the Spanish affricate based on quantitative differences that surpass the level of perceptibility described by other research on segment frequency measures (e.g., Labov et al. 2013; Herrero de Haro 2017a; Gordon et al. 2002). While the values of F1 and F2 at the 20% mark of vowels following the affricate did not show marked differences based on realization types, both the center of gravity of the frication period of the affricate and the percent of frication in the segment followed trends identified in previous studies for phonetically perceptual and even phonemic distinctions between alveolar and post-alveolar fricative segments. This includes Recasens and Espinosa's (2007) findings for Catalan, where the duration of both the frication period and whole segment of affricate [ts] were found to be longer than [tʃ] in both Majorcan and Valencian. It also includes Gordon et al. (2002), who found the alveolar fricative to generally be the longer of the two segments across an array of typologically distinct languages. A comparison of Gordon and colleagues' findings with those for the frication period of affricates in the current study is shown in Table 7. As they do not normalize sibilant fricatives, the raw COG values for the current study are used in this table.

**Table 7.** Comparison of sibilant fricative results from Gordon et al. (2002) and the current ones for allophones of the Peninsular Spanish affricate.

| **Language** | **/s/ Duration** | **COG** | **/ʃ/ Duration** | **COG** |
|---|---|---|---|---|
| 1. Aleut | 361.8 | 5219 | - | - |
| 2. Apache | 172.3 | 5461 | 175.9 | 4859 |
| 3. Chickasaw | 123.6 | 5163 | 112.8 | 4679 |
| 4. Gaelic | 130.4 | 4884 | 110.7 | 4396 |
| 5. Hupa | 276.4 | 4797 | 217.7 | 4440 |
| 6. Montana Salish | 171.8 | 4601 | 178.9 | 4134 |
| 7. Toda | 198.3 | 4529 | 239.6 | 4704 |
| **Average** | **204.9** | **4950.6** | **172.6** | **4535.3** |
| st.dev. | 79.4 | 317.5 | 48.3 | 239.2 |
| **Peninsular Spanish** | **[ts] duration** | **COG** | **[tʃ] duration** | **COG** |
| 8. Frication portion | 61.8 | 4783.4 | 54.5 | 3978.4 |
| 8.1. Total for affricate | 140.3 | | 127.4 | |
| **Overall average** | **187.1** | **4929.7** | **155.7** | **4455.8** |
| st.dev. | 94.1 | 323.0 | 65.7 | 318.7 |

The ordering for both duration and COG in Spanish is consistent with the language tendencies identified by Gordon and colleagues. For the duration of the frication period of the Spanish affricate, it is important to note that this represents only around half of the overall segment length (i.e., in addition to the occlusion period of the affricate, Almeida (2019)), and it should not be taken for granted that affricate frication necessarily maps onto trends in fully fricative segments. However, by examining the duration, as well as Table 4 with its frication percentage model, similarities can be observed between these frication portions and the fricative segments examined in other languages. Additionally, while the COG of frication in the realization coded as the alveolar affricate [ts] is well within the range of other sibilant fricative productions of /s/, the COG for the post-alveolar production is lower than any other value found by Gordon and colleagues. In part, as these are non-normalized results, this could stem from the difference between data recorded in a laboratory and the naturalistic televised audio used in the present study. However, it could also point to a difference in sibilant quality of the normative Peninsular post-alveolar affricate compared to the prototypical production of /ʃ/ found in these other languages. In general, though, the average Spanish alveolar production of [ts] is longer than the post-alveolar production [tʃ] and has an elevated COG of its frication period, similar to the distinctions made in other languages where this difference is phonemic in the sibilant fricative system.

These results offer insight into this novel phenomenon, while at the same time providing baseline numbers for typical affricate productions in naturalistic speech in Peninsular Spanish. While the affricate is often treated as unvaryingly post-alveolar among middle-aged and younger speakers, these results suggest that this is not fully true at the phonetic level, at least among these political actors from both Andalusia and Madrid. Although this phenomenon does not directly affect the phonemic system of Spanish (although see the discussion by Vida-Castro and Villena-Ponsoda (2016) about the overlap in perception between [ts] as /tʃ/ and aspiration in the /st/ cluster—also discussed below at the end of Section 5.3), it is nevertheless expected to reach a level of perceptibility for listeners, offering a context for social stratification between the alveolar and post-alveolar forms. Further research is needed, particularly among members outside the political speech community,[2] to determine the extent to which this is an ongoing shift in the phonetics of Peninsular Spanish.

### *5.2. Affricate Variation in Peninsular Political Speech*

The next focus in this study was to identify variables influencing allophonic selection in the political speech of Madrid and Andalusia. As shown in Figure 3, speakers' articulatory tracts tended to scale their affricate productions alongside those of /s/, such that individuals with productions of the fricative with higher COG tended to also produce affricates with higher COG. Normalization reduced these differences, particularly those related to gender and individual differences in articulatory tract size. Additionally, descriptive differences across realization type showed some degree of variation across all four social variables: city, political affiliation, gender, and speech context. However, through the mixed-effects models, specific linguistic and extralinguistic variables were found to significantly explain the differences in the quality and type of affricate that speakers produced, as laid out in Table 8. In the following paragraphs, significant variables are discussed, and contrasts are made across the three models to describe what trends duration, COG, and the categorical realization coding identify for the analysis.

Two linguistic variables with consistently high magnitudes of effect were the surrounding context, both preceding and following. In general, the Spanish affricate /tʃ/ most often occurs intervocalically (e.g., *mucho* 'a lot'), although there are also cases with a typically alveolar preceding consonant (e.g., *ancho* 'wide'). With respect to the following context, there was a general trend across all three models in which front vowels favor production of an affricate with a longer frication portion, a higher COG, and coding as [ts]. Additionally, for COG and realization type, the central vowel [a] favored these productions.

Given the consistency and direction of effect across models, this suggests a coarticulatory effect of the following linguistic environment, where following [u] or [o] is less likely to permit tongue advancement to the alveolar ridge for the affricate.

**Table 8.** Significant variables and tendencies from the mixed-effects regression models.

| # | Variable Name | Frication Percent (Favor Higher % Frication) | Center of Gravity (Favor Higher COG) | Realization Coding (Favor [ts]) |
|---|---|---|---|---|
| **Linguistic Variables** | | | | |
| 1 | Frication Percent | | - | Greater percent |
| 2 | Center of Gravity | - | | Higher COG |
| 3 | Realization Type | [ts] | [ts] | |
| 4 | Preceding Context | Consonants, back and central vowels | Consonants, front and central vowels | Consonants and front vowels |
| 5 | Following Context | Front vowels | Front and central vowels | Front and central vowels |
| 6 | Lexical Frequency | More frequent | - | - |
| 7 | Following Vowel F2 (normalized and scaled) | - | Higher F2 | - |
| 8 | Following Vowel F1 (normalized and scaled) | Higher F1 | - | Higher F1 |
| 9 | Tonicity | - | Tonic syllables | - |
| **Extralinguistic Variables** | | | | |
| 10 | City | - | Madrid | Cordoba and Madrid |
| 11 | Political Party | PSOE politicians | - | - |
| 12 | Speech Context | Male and female unscripted | Scripted and female unscripted | Scripted and female unscripted |
| 13 | Gender | - | Men | Women |

Meanwhile, the preceding context offers a slightly more complicated picture. While, for the duration-adjacent measure of frication percent in the segment, greater rates of frication correlated with consonants, back, and central vowels, both increased COG and realization coding of [ts] were favored in the context of preceding consonants and front vowels (as well as central vowels for COG). As Gordon et al. (2002) discussed, duration-related measures were the least reliable in distinguishing between the place of frication production. In only three of the six typologically distinct languages for which the duration of /ʃ/ in comparison with /s/ was considered (i.e., Apache, Montana Salish, and Toda) did /s/ have a shorter duration of the two segments, even though the overall average showed it to have a longer one. The difference identified in the current study may reflect the more uncertain role of duration in distinguishing between sibilant production location. For the most part, it seems that preceding consonants and central vowels favor a more alveolar production of the affricate, while increased COG and realization coding as [ts] additionally favored preceding front vowels, which may point to a similar effect as seen for the following context.

The next pair of linguistic variables significant across all three models was the formant height of the following vowel. The first formant was significant for both frication percent and realization coding, with higher F1 values favoring coding as [ts] and a longer period of frication. This change in formant value would correlate with a general lowering in vowel height following a more alveolar production, possibly reflecting the prototypically concave apico-dental production of the Spanish alveolar sibilant, which lowers the dorsum of the tongue and turns the tip upward (Núñez-Méndez 2022). This tongue movement may result in a co-articulatory lowering of the following vowel, in part due to the pressure being imposed by the coarticulatory resistance of the adjacent /s/. Alternately, the second formant was only found to be significant for the COG model, with higher values of F2 (i.e., more fronted productions) being favored alongside higher values of frication COG. This finding would make articulatory sense, as a raised tongue gesture in the affricate would be expected to have an effect on the following sound.

Context alone does not influence production: tonicity was also found to play a role in the continuous COG model. Instances of the affricate in tonic syllables tended to have a higher COG than those in atonic syllables, with an average difference of 56.2 Hz. This suggests that more fronted productions would favor those contexts with greater emphasis, aligning with previous discussion of this phenomenon as a possible stylistic marker growing in salience (Pollock 2023a). It is important to note, however, that the average difference in normalized COG between tonic and atonic contexts does not quite reach the perceptibility threshold of 60 Hz described in previous vowel studies (e.g., Labov et al. 2013; Herrero de Haro 2017a), which may suggest incipient change still underway surrounding this form.

Lexical frequency was also found to play a role in variable production based on the frication period. For this variable, more frequent lexical items within the corpus favored production with longer duration periods, correlated to some extent with [ts] production. There were slightly over 200 unique lexemes across the 3175 tokens, with only 4 having more than 100 tokens (i.e., *mucho/a(s)* 'a lot', *hecho* 'done', *dicho* 'said', and *marcha* 'proceeding'), and 19 having 10 or more occurrences. In many cases, it has been determined that language change spreads first from high-frequency items to those that are less frequent, as argued by Bybee (2002, 2010), particularly in the type of retiming phenomena being tracked by the percent frication measure (e.g., Bybee 1998, p. 422; Díaz-Campos et al. 2023)[3]. When combined with tonicity, the results for lexical frequency indicate that the variation targeted in this study may in fact represent the first stages of a change in progress.

The final three linguistic variables selected as conditioning variation were considered as a way to connect the mixed-effects models. These included the realization type, the COG of frication in the affricate, and the percent frication (i.e., the dependent variables) being treated as independent variables in the other models. Each of the three was selected as a significant predictor of variation at least once elsewhere. When realization type was the dependent variable, tokens coded as [ts] favored production with a higher COG and with a longer frication period, while in the continuous models for COG and frication duration, realization type was selected as significant predictors of variation. The relationship between realization type and the two quantitative continuous factors suggests that, even if COG and percent frication do not fully map onto each other, both can be seen to play a role in perception of the affricate.

Four social variables remain to be discussed. First, the city of the politician was significantly predictive of COG, as well as in an interaction with the following context for realization type. Speakers from Madrid were most likely to produce variants with a higher COG (i.e., correlating with the alveolar production), while those from all three cities in Andalusia were less likely to have increased COGs. Furthermore, the range of production differences by city was over 120 Hz, which should fall well within the range of human perceptibility (Labov et al. 2013; Herrero de Haro 2017a; Gordon et al. 2002). Based on realization type, both speakers from Cordoba and Madrid had a nonsignificant tendency to produce [ts], but there was also a significant difference in production norms, with speakers from Cordoba, Malaga, and Seville all producing the most [ts] after [a], while those in Madrid instead produced it at higher rates after [i] and [u]. This difference in production norms could, as a reviewer points out, be related to the fact that affricate production in Madrid is incipient but the most salient, due to the highest degree of fronting, and that contexts preceding high, front vowels are a frequent context for affricate changes, from which change can spread.

Next, for speaker gender, based on the continuous analysis of COG, men favored the production of a higher normalized COG for affricates than women overall (22 Hz), although this difference falls below the expected threshold of perception. However, for realization type, this tendency is reversed, with women identified as producing around twice as many alveolar tokens as men (i.e., 37.7% vs. 16.1%). This trend merits further discussion. First, there is relatively little work that describes means of normalizing sibilant production in the articulatory space of multiple speakers. In Toda's (2007, p. 828) examination of normalized /s/ and /ʃ/ in two languages, the author finds that distinctness was easier to identify in Japanese than in French, potentially as a result of the allophonic relationship between them in French, which may lead to greater variability in production norms. Meanwhile, Gordon et al. (2002) chose to look at raw COG values across the seven languages they considered, presenting individual speakers by gender without attempting to compare norms quantitatively.

While the method of normalization used in the current article has aimed to reduce differences in individual speech, across social factors, and within audio recordings, it is unclear the extent to which the normalization and rescaling processes can erase secondary markers of variation used in perception. Admittedly, female articulatory spaces can be smaller than those of men, at times leading to sibilants with a higher COG, as seen in Figure 3,[4] but the high rate of classification of female affricates, such as [ts], suggests that there are audible differences between the alveolar and post-alveolar productions that are not captured in the continuous models, suggesting additional acoustic correlates that may be at work. This requires further analysis using perceptual instruments to determine the extent to which native speakers of the Andalusian and northern central Peninsular Spanish (NCPS) varieties perceive these differences, and to begin identifying how these various correlates and potential others work together to influence classification.

When examining speech context, there appears to exist a sociocultural environment reminiscent of Flores's (2014) study of Chilean radio speech. In her analysis of linguistic phenomena including an alveolar affricate [ts], she found that men tended to use a greater degree of vernacular forms with other men, while women used more prestige forms with other women. Meanwhile, scripted speech tended to carry the most careful and normative productions. In the current study, in the interaction for realization type between speaker gender and speech context, both male and female politicians used higher rates of [ts] with female interviewers, while only women used similar rates in scripted speeches (Figure 8). Meanwhile, in the interaction shown in Figure 7 between speaker gender and scripted speech, there is a marked dip in COG for female politicians when producing /tʃ/ with a male interlocutor for both variants, but particularly the post-alveolar affricate. This suggests a strategy of backing among female speakers to accommodate to affricate production by male colleagues.

Finally, for the model analyzing frication duration, there is another deviation from the COG and realization type models—scripted speeches have the shortest frication period, while those in unscripted interviews are longer. This may suggest an unrelated phenomenon associated with register rather than sibilant place, where frication production is slightly more evenly produced alongside the occlusion period of the affricate in more professional scripted settings. In any case, considering the social tendencies Flores identified (i.e., that same-gender dyads lead to increased rates of vernacular or prestige forms, depending on group gender), the results for COG and realization type suggest an association of more alveolar-type affricate productions with female speech, and a move away from these norms in conversations with male interviewers. Men use [ts] more with female interlocutors than in scripted speeches, while female politicians use it in both scripted speeches and with female interlocutors, but produce a considerably more backed affricate with male interlocutors, suggesting that the alveolar variant is associated with both professional prestige and feminine identity.

The final social factor under consideration was the political party of the speaker. This factor was selected as a significant predictor of variation for frication period, with the largest range in production across all the variables in the model in Table 4. Socialist politicians in the PSOE produced a frication period that occupied 53.5% of their affricates, on average, while PP politicians' frication periods were closer to 48.4%. As this tendency is not reflected in the other two models, this may provide yet more evidence of the finding of Gordon et al. (2002); namely, that duration-related measures are subject to greater variability in the production of the sibilants [s] and [ʃ]. Unlike phenomena related to liquid neutralization and opacity, referenced by Pollock (2023a), as well as intervocalic /d/ elision, described by Cruz-Ortiz (2022), Pollock and Wheeler (2022), Pollock and Wheeler (forthcoming) and others, which offer political actors of certain affiliations resources to distinguish themselves from their peers and craft an identity rooted in regional identification, the affricate is not so well-integrated into the indexical norms of Andalusian politics. While greater variation does seem to exist in cities like Madrid, these differences are not sufficiently consistent to fall outside the realm of random chance, as illustrated in Figure 6. However, there is evidence to suggest that this may be changing.

*5.3. A Labovian Sociolinguistic Marker*

The final question addressed by this study involved the role of the Spanish affricate in the Peninsular speech community, and whether it has acquired social stratification to the point that it could be considered a sociolinguistic marker, rather than an indicator. Based on the description by Labov (1972), this would mean that there is variability not only at the linguistic level, but also the extralinguistic. Given the tendencies described in the three mixed-effects regression models, it is clear that the linguistic environment, tonicity, and COG do not on their own account for all variability in use—instead, variable use also differs based on speaker gender, speech context, political affiliation, and city, as well as lexical frequency and tonicity (which may serve as initial signals of incipient change).

The contexts in which the realizations coded as alveolar, as well as productions with higher centers of gravity and longer frication periods, are used—by women, in scripted speeches, in interviews with women, as well as in Madrid and in tonic syllables—are traditionally associated with contexts of prestige and attention paid to speech (e.g., Flores 2014; Labov 1972). This comes alongside the fact that these speakers are major politicians at the regional and national level, often with considerable education and a cultivated public persona, all of which are aspects of identity associated with prestige. Finally, it is noteworthy that none of these individuals produced the deaffricated post-alveolar fricative [ʃ], which is frequently associated with rural and working-class Andalusian speech. Only the post-alveolar and alveolar affricates were identified across three dozen hours of scripted discourse and interviews, suggesting that even if politicians may use certain regional forms to index working-class identity (e.g., Hernández-Campoy and Cutillas-Espinosa 2013; Pollock 2023a), the spirantized allophone of the affricate is too stigmatized to serve as a resource for identity construction among these politicians.

Interestingly, as there was only a slight connection between political party and frication duration, and a general tendency for the alveolar affricate [ts] to be favored by Madrid politicians, this phenomenon may still be in the early stages of acquiring social meaning and becoming a sociolinguistic marker. Given its increased use preceding front vowels, as well as its appearance in tonic syllables, the phenomenon may have initially emerged from coarticulatory effects, particularly in female speech. However, based on the social and frequency-related variables that now condition duration, COG, and perception, this seems to be undergoing indexicalization as speakers become more aware of it and it takes on social meaning (Silverstein 2003).

As the Spanish sibilant inventory contains only a single affricate, variability in its production is not fully surprising. However, should alveolar production of the Spanish affricate acquire further social meaning and currency in Peninsular Spanish, there is also the opportunity for the neutralization of phonological contrast for some Andalusian speakers. As described by Torreira (2006), Ruch (2012), Del Saz (2019, 2023), and Vida-Castro and Villena-Ponsoda (2016), among others (also see Footnote 1), there is a separate process whereby the alveolar fricative /s/ undergoes elision in /st/ clusters, yielding not only productions with post-aspiration [$t^h$], but also those with affrication [ts]. Perceptual work by the aforementioned authors has also identified comparisons between [ts] resulting from /st/ and /tʃ/ segments. In varieties where both this phenomenon and the fronting of the affricate coexist, further examination is necessary to determine if productions of [ts] differ in minimal pairs like *hecho* 'fact' [etso] and *esto* 'this' [etso]. In order to avoid the loss of contrast, the Andalusian sibilant affricate inventory may develop additional complexity to allow for distinction through duration, COG, or some other measure.

## 6. Conclusions

In addition to offering insight into the phonetic, linguistic, and social outlook of the Spanish affricate, this study also opens the door to future research and methodologies to better conceptualize the phenomenon of Peninsular affricate fronting. More perceptual work is necessary not just to measure speaker attitude and awareness of the difference between the alveolar and post-alveolar variant, but also to determine the perceptual threshold of differences in segments, as a means of better reinforcing and refining the development of a normalization procedure for sibilant consonants. This, in turn, will help determine the exact status of the phenomenon in Andalusia and Madrid, in addition to offering a concrete understanding of the differentiation that native Spanish speakers carry out between adjacent fricative sounds. While it may not be the case that speakers have a completely overt awareness of the social distinctions between these two allophones, there does seem to be an awareness of differences in affricate production that maps onto some form of prestigious, feminine identity. This should be examined further, particularly outside of the realm of political speech, to understand how widespread this phenomenon may be in Peninsular varieties of Spanish.

Overall, this analysis identifies systematic differences in production norms for affricates by Peninsular politicians and determines that both linguistic and social variables condition usage. As a result, the Spanish alveolar affricate has the characteristics necessary to be described as an incipient Labovian marker that shows initial signs of social stratification. There is consistent variability in production across individuals, stratification across social groups, and signs that the alveolar production is associated with female, emphatic, and prestige speech particularly in NCPS, with growing use in Andalusian Spanish as well.

**Supplementary Materials:** The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/languages9060218/s1. Table S1: speaker and video identification.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** This study used public audio data and was deemed exempt from seeking IRB approval.

**Informed Consent Statement:** Not applicable, as the study used public data.

**Data Availability Statement:** Data available on request from the author.

**Acknowledgments:** The analysis in this paper benefited greatly as a result of the feedback and support of attendees at the 2022 Workshop on Spanish Sociolinguistics and from conversation with faculty in the Indiana University Linguistics Department.

**Conflicts of Interest:** The author declares no conflicts of interest.

## Notes

1 A correlation has also been pointed out between affricate productions of [t$^{s}$] in the wake of /s/ reduction before /t/ and a fronted variant associated with the affricate /tʃ/, by researchers including Torreira (2006), Ruch (2012), Del Saz (2019, 2023), and Vida-Castro (2022). Research in this vein has also begun to examine the production of [ts] deriving from affricates. For example, Vida-Castro and Villena-Ponsoda (2016) showed that a higher COG for center of gravity is more likely to receive perception as /tʃ/ rather than /st/ in Malaga, Spain.

2 As Cruz-Ortiz (2022) and Pollock and Wheeler (2022) have pointed out, among others, the elision of intervocalic /d/ occurs much more among politicians than even among rural and working-class speakers across Spain, suggesting that there is an association between "speaking like a politician" and eliding in these contexts. Given the demographics of the current study, it is not yet clear if the same is true for the alveolar affricate—however, Pollock (2023a), who included the alveolar affricate in a composite of regional features used by politicians, found that listeners rate it as more urban, educated, likeable, and less Andalusian than the post-alveolar variant, suggesting social associations with the form.

3 However, see the recent *Handbook of Usage-Based Linguistics* for an overview of the many ways in which these frequency-based trends have been explored in greater nuance, as well as contexts in which frequency effects have been found not to play a major role (Díaz-Campos and Balasch 2023).

4 It is also important to note that there is evidence to suggest that gendered differences can also vary cross-linguistically. Fuchs and Toda (2010) found that sibilant productions differed between gendered groups in English and German, which may have resulted from differences in the fricative inventory of the two languages. Additionally, Munson (2011) identified perceived talker gender as playing an important role in differentiating /s/ from /ʃ/, showing that there are a number of biological and sociocultural influences on sibilant fricative production that merit further study among Spanish politicians.

## References

Abramson, Arthur S., and Leigh Lisker. 1972. Voice-timing perception in Spanish word-initial stops. *Journal of Phonetics* 1: 1–8. [CrossRef]

Almeida, Manuel. 2019. Language hybridism: On the origin of interdialectal forms. In *Language Variation—European Perspectives VII: Selected Papers from the Ninth International Conference on Language Variation in Europe (ICLaVE 9), Malaga, June 2017 [Studies in Language Variation 22]*. Amsterdam: John Benjamins, pp. 9–26.

Bell, Alan. 1984. Language style as audience design. *Language in Society* 13: 145–204. [CrossRef]

Boersma, Paul, and David Weenink. 2023. Praat: Doing Phonetics by Computer [Computer Program]. Version 6.0.35. Available online: http://www.praat.org/ (accessed on 16 October 2023).

Bradley, Travis G., and Claire J. Lozano. 2022. Language Contact and Phonological Innovation in the Voiced Prepalatal Obstruents of Judeo-Spanish. *Languages* 7: 313. [CrossRef]

Bucholtz, Mary, and Kira Hall. 2005. Identity and Interaction: A Sociocultural Linguistic Approach. *Discourse Studies* 7: 585–614. [CrossRef]

Bybee, Joan. 1998. The Emergent Lexicon. *Chicago Linguistic Society* 34: 421–35.

Bybee, Joan. 2002. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change* 14: 261–90. [CrossRef]

Bybee, Joan. 2010. *Language, Usage and Cognition*. Cambridge: Cambridge University Press.

Coupland, Nikolas. 2001. Language, situation, and the relational self: Theorizing dialect-style in sociolinguistics. In *Style and Sociolinguistic Variation*. Edited by Penelope Eckert and John R. Rickford. Cambridge: Cambridge University Press, pp. 185–210.

Cruz-Ortiz, Rocío. 2022. *Sociofonética andaluza. Caracterización lingüística de los presidentes y ministros de Andalucía en el Gobierno de España (1923–2011)*. Berlin: De Gruyter.

Del Saz, María. 2019. From postaspiration to affrication: New phonetic contexts in western Andalusian Spanish. Paper presented at 19th ICPhS, Melbourne, Australia, August 5–9.

Del Saz, María. 2023. (In)complete neutralization in Western Andalusian Spanish. *Estudios de Fonética Experimental* 32: 21–42. [CrossRef]

Díaz-Campos, Manuel, and Sonia Balasch. 2023. *The Handbook of Usage-Based Linguistics*. Hoboken: Wiley-Blackwell.

Díaz-Campos, Manuel, Molly Cole, and Matthew Pollock. 2023. Re-conceptualizing affricate variation in Caracas Spanish. *Hispania* 106: 9–26.

Díaz-Campos, Manuel, Stephen Fafulas, and Michael Gradoville. 2018. Stable variation or change in progress? A sociolinguistic analysis of the alternation between para and pa. In *Current Issues in Linguistic Theory*. Edited by Jeremy King and Sandro Sessarego. Amsterdam: John Benjamins, pp. 223–45.

Eckert, Penelope. 2008. Variation and the indexical field. *Journal of Sociolinguistics* 12: 453–76. [CrossRef]

Fernández de Molina Ortés, Elena. 2020. ¿Cómo hablan los presentadores andaluces en televisión? Percepción de los espectadores y análisis lingüístico de la posible variación (socio)estilística en los medios. *LinRed* 17: 1–29. [CrossRef]

Flores, Tanya. 2014. A Socio-Phonetic Examination of Co-Articulation in Chilean Public Speech (Document No. AAI3665473). Ph.D. dissertation, Indiana University, ProQuest Dissertations and Theses Global, Bloomington, IN, USA.

Flores, Tanya. 2018. Lexical Frequency Effects in Chilean Spanish Affricate Fronting. *Southern Journal of Linguistics* 41: 17–39.

Fuchs, Susanne, and Martine Toda. 2010. Do differences in male versus female /s/ reflect biological or sociophonetic factors? In *Turbulent Sounds: An Interdisciplinary Guide*. Edited by Susanne Fuchs, Marzena Zygis and Martine Toda. Berlin: Mouton de Gruyter, pp. 281–302.

Giles, Howard, Nikolas Coupland, and Justine Coupland. 1991. Accommodation theory: Communication, context, and consequence. In *The Contexts of Accommodation*. Edited by Howard Giles, Nikolas Coupland and Justine Coupland. New York: Cambridge University Press, pp. 1–68.

Gordon, Matthew, Paul Barthmaier, and Kathy Sands. 2002. Cross-linguistic acoustic study of voiceless fricatives. *Journal of the International Phonetic Association* 32: 141–74. [CrossRef]

Hall-Lew, Lauren, Rebecca Starr, and Elizabeth Coppock. 2012. Style-shifting in the U. S. Congress: The vowels of "Iraq(i)". In *Style-Shifting in Public: New Perspectives on Stylistic Variation*. Edited by Juan Manuel Hernández-Campoy and Juan Antonio Cutillas-Espinosa. Amsterdam: John Benjamins, pp. 45–63.

Hall-Lew, Lauren, Ruth Friskney, and James M. Scobbie. 2017. Accommodation or political identity: Scottish members of the UK Parliament. *Language Variation and Change* 29: 341–63. [CrossRef]

Harrington, Jonathan. 2007. Evidence for a relationship between synchronic variability and diachronic change in the Queen's annual Christmas broadcasts. *Laboratory Phonology* 9: 125–44.
Hay, Jennifer, Stefanie Jannedy, and Norma Mendoza-Denton. 2010. Oprah and /ay/: Lexical Frequency, Referee Design, and Style. In *The Routledge Sociolinguistics Reader*. Edited by Miriam Meyerhoff and Erik Schleef. New York: Routledge, pp. 53–59.
Henríquez-Barahona, Marisol, and Darío Fuentes-Grandón. 2018. Realizaciones de los fonemas /tʃ/ y/t͡ʂ/ en el chedungun hablado por niños bilingües del alto Biobío: Un análisis espectrográfico. *Literatura y Lingüística* 37: 253–72. [CrossRef]
Hernández-Campoy, Juan Manuel, and José María Jiménez-Cano. 2003. Broadcasting standardization: An analysis of the linguistic normalization process in Murcian Spanish. *Journal of Sociolinguistics* 7: 321–47. [CrossRef]
Hernández-Campoy, Juan Manuel, and Juan Antonio Cutillas-Espinosa. 2010. Speaker design practices in political discourse: A case study. *Language and Communication* 30: 297–309. [CrossRef]
Hernández-Campoy, Juan Manuel, and Juan Antonio Cutillas-Espinosa. 2013. The effects of public and individual language attitudes on intra-speaker variation: A case study of style-shifting. *Multilingua* 31: 79–101. [CrossRef]
Herrero de Haro, Alfredo. 2017a. Four mid back vowels in Eastern Andalusian Spanish. *Zeitschrift fur romanische Philologie* 133: 82–114. [CrossRef]
Herrero de Haro, Alfredo. 2017b. The phonetics and phonology of Eastern Andalusian Spanish: A review of literature from 1881 to 2016. *Íkala, Revista de Lenguaje y Cultura* 22: 313–57. [CrossRef]
Holliday, Nicole. 2017. "My Presiden(T) And Firs(T) Lady Were Black": Style, Context, And Coronal Stop Deletion in The Speech of Barack And Michelle Obama. *American Speech* 92: 459–86. [CrossRef]
Hualde, José Ignacio, and Sonia Colina. 2014. *Los Sonidos del español*. New York: Cambridge University Press.
Hualde, José Ignacio. 2005. *The Sounds of Spanish*. New York: Cambridge University Press.
IBM. 2019. *Watson Speech to Text Demo*. [Computer Program]. Available online: https://speech-to-text-demo.ng.bluemix.net/ (accessed on 2 November 2019).
Johnson, Daniel Ezra. 2009. Getting off the GoldVarb Standard: Introducing Rbrul for Mixed-Effects Variable Rule Analysis. *Language and Linguistics Compass* 3: 359–83. [CrossRef]
Johnson, Daniel Ezra. 2014. Progress in Regression: Why Natural Language Data Calls For Mixed-Effects Models. Available online: http://www.danielezrajohnson.com/ (accessed on 15 April 2023).
Jongman, Allard. 1989. Duration of frication noise required for identification of English fricatives. *Journal of the Acoustical Society of America* 85: 1718–25. [CrossRef]
Jongman, Allard, Ratree Wayland, and Serena Wong. 2000. Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America* 108: 1252–63. [CrossRef]
Kendall, Tyler, and Erik R. Thomas. 2010. *Vowels: Vowel Manipulation, Normalization, and Plotting in R*. R Package, Version 1.1. Available online: http://ncslaap.lib.ncsu.edu/tools/norm/ (accessed on 15 October 2023).
Kirkham, Sam, and Emma Moore. 2016. Constructing social meaning in political discourse: Phonetic variation and verb processes in Ed Miliband's speeches. *Language in Society* 45: 87–111. [CrossRef]
Labov, William. 1972. *Sociolinguistic Patterns*. Oxford: Blackwell.
Labov, William, Ingrid Rosenfelder, and Josef Fruehwald. 2013. One Hundred Years of Sound Change in Philadelphia: Linear Incrementation, Reversal, and Reanalysis. *Language* 89: 30–65. [CrossRef]
Li, Aijun, and Zhiqiang Li. 2019. Spectral Analysis of Sibilant Fricatives and the Ling Sound Test for Speakers of Chinese Dialects. ICPHS Full Papers. pp. 3026–30. Available online: https://icphs2019.org/icphs2019-fullpapers/ (accessed on 1 May 2022).
Lobanov, Boris M. 1971. Classification of Russian vowels spoken by different listeners. *Journal of the Acoustical Society of America* 49: 606–8. [CrossRef]
Mazzaro, Natalia. 2022. The Effect of Language Contact on /tʃ/ Deaffrication in Spanish from the US–Mexico Borderland. *Languages* 7: 101. [CrossRef]
Melguizo-Moreno, Elisabeth. 2006. La fricatización de /č/ en una comunidad de hablantes granadina. *Interlingüística* 17: 748–57.
Munson, Benjamin. 2011. The influence of actual and imputed talker gender on fricative perception, revisited. *The Journal of the Acoustical Society of America* 130: 2631–34. [CrossRef]
Núñez-Méndez, Eva. 2022. Variation in Spanish /s/: Overview and New Perspectives. *Languages* 7: 77. [CrossRef]
Penny, Ralph. 2002. *A History of the Spanish Language*. Cambridge: Cambridge University Press.
Pollock, Matthew. 2023a. Performing Andalusian in Political Speech: Political Party and Sociophonetic Patterns Across Production and Perception. Ph.D. dissertation, Indiana University, Bloomington, IN, USA.
Pollock, Matthew. 2023b. Toeing the Party Line: Indexing Party Identity through Dialectal Phonetic Features in Spanish Political Discourse. *Languages* 8: 196. [CrossRef]
Pollock, Matthew, and Jamelyn A. Wheeler. 2022. Syllable-final /s/ and intervocalic /d/ elision in Andalusia: The Formation of Susana Díaz's Regional Identity in Political Discourse. *Language and Communication* 87: 191–204. [CrossRef]
Pollock, Matthew. forthcoming. Politics, Variation, and Politeness on Andalusian Twitter (X): The second-person plural in Peninsular Spanish identity construction. In *Issues in Hispanic and Lusophone Linguistics*. Amsterdam: John Benjamins.
Pollock, Matthew, and Jamelyn A. Wheeler. forthcoming. Language, politics, and power: A sociophonetic comparison of political and community norms in Galician Spanish. In *Issues in Hispanic and Lusophone Linguistics*. Amsterdam: John Benjamins.

Rama, José, Lisa Zanotti, Stuart J. Turnbull-Dugarte, and Andrés Santana. 2021. *Vox: The Rise of the Spanish Populist Radical Right*. London: Routledge.
Recasens, Daniel, and Aina Espinosa. 2007. An electropalatographic and acoustic study of affricates and fricatives in two Catalan dialects. *Journal of the International Phonetic Association* 37: 143–72. [CrossRef]
Recasens, Daniel. 2018. Coarticulation. In *Oxford Research Encyclopedia of Linguistics*. Edited by Mark Aronoff. Oxford: Oxford University Press.
Regan, Brendan. 2020. El [ʃ]oquero: /tʃ/ Variation in Huelva Capital and Surrounding Towns. *Estudios de Fonética Experimental* 29: 55–90.
Ruch, Hanna. 2012. Affrication of /st/-clusters in Western Andalusian Spanish: Variation and change from sociophonetic point of view. In *Proceedings of Sociophonetics, at the Crossroads of Speech Variation, Processing and Communication*. Edited by Silvia Calamai, Chiara Celata and Luca Ciucci. Pisa: Edizioni della Normale, pp. 61–64.
Samper-Padilla, José Antonio. 2011. Socio-phonological Variation and change in Spain. In *The Handbook of Hispanic Sociolinguistics*. Edited by Manuel Díaz-Campos. Malden: Wiley-Blackwell, pp. 98–120.
Schilling-Estes, Natalie. 2013. Investigating stylistic variation. In *The Handbook of Language Variation and Change*. Edited by Jack Chambers, Peter Trudgill and Natalie Schilling-Estes. Oxford: Basil Blackwell, pp. 1–23.
Sharma, Devyani. 2018. Style dominance: Attention, audience, and the 'real me'. *Language in Society* 47: 1–31. [CrossRef]
Silverstein, Michael. 2003. Indexical Order and The Dialectics of Sociolinguistic life. *Language and Communication* 23: 193–229. [CrossRef]
Tagliamonte, Sali A. 2012. *Variationist Sociolinguistics: Change, Observation, Interpretation*. Hoboken: Blackwell Publishing.
Toda, Martine. 2007. Speaker normalization of fricative noise: Considerations on language-specific contrast. Paper presented at 16th International Congress on Phonetic Sciences, Saarbrucken, Germany, August 6–10; pp. 825–28.
Torrano Moreno, Laura. 2017. Variación y cambio sociolingüístico en una comunidad de habla local: El caso de la variable (ch) y los factores ontogenético y generolectal en Ricote. *Tonos Digital* 33.
Torreira, Francisco. 2006. Coarticulation between aspirated-s and voiceless stops in Spanish: An interdialectal comparison. In *Selected Proceedings of the 9th Hispanic Linguistics Symposium*. Edited by Nuria Sagarra and A. Jacqueline Toribio. Somerville: Cascadilla Press, pp. 113–20.
Vida-Castro, Matilde. 2022. On competing indexicalities in southern Peninsular Spanish. A sociophonetic and perceptual analysis of affricate [ts] through time. *Language Variation and Change* 34: 137–63. [CrossRef]
Vida-Castro, Matilde, and Juan-Andrés Villena-Ponsoda. 2016. Percepción y análisis de pistas en la discriminación alofónica: Fusión y escisión en el estudio sociolingüístico de la ciudad de Málaga: Informe preliminar. In *53 reflexiones sobre aspectos de la fonética y otros temas de lingüística*. Edited by Anna Maria Fernández Planas. Barcelona: Universitat de Barcelona Press, pp. 131–37.
Villena-Ponsoda, Juan Andrés. 2008. Sociolinguistic patterns of Andalusian Spanish. *International Journal of the Sociology of Language* 2008: 139–60. [CrossRef]
Villena-Ponsoda, Juan Andrés. 2013. Actos de identidad: ¿Por qué persiste el uso de los rasgos lingüísticos de bajo prestigio social? Divergencia geográfica y social en el español urbano de Andalucía. In *Estudios Descriptivos y Aplicados Sobre el Andaluz*. Edited by Rosario Guillén Stuil. Seville: The University of Seville Press, pp. 173–207.

languages

MDPI

Article

# The Influence of Language Experience on Speech Perception: Heritage Spanish Speaker Perception of Contrastive and Allophonic Consonants

**Amanda Boomershine [1,*] and Keith Johnson [2]**

[1] Department of World Languages and Cultures, University of North Carolina Wilmington, 601 S College Road, Wilmington, NC 28403, USA
[2] Department of Linguistics, University of California Berkeley, 1203 Dwinelle Hall #2650, Berkeley, CA 94720, USA; keithjohnson@berkeley.edu
* Correspondence: boomershinea@uncw.edu

**Abstract:** It is well known that a listener's native phonological background has an impact on how speech sounds are perceived. Native speakers can distinguish sounds that serve a contrastive function in their language better than sounds that are not contrastive. However, the role of allophony in speech perception is understudied, especially among heritage speakers. This paper highlights a study that directly tests the influence of the allophonic/phonemic distinction on perception by Spanish heritage speakers, comparing their results to those of late bilingual and monolingual speakers of Spanish and English in the US. Building on an earlier study, the unique contribution of this paper is a study of the perceptual pattern shown by heritage speakers of Spanish and a comparison of bilingual and monolingual speakers of English and Spanish. The participants completed a similarity rating task with stimuli containing VCV sequences with the intervocalic consonants [d], [ð], and [ɾ]. The heritage speakers, who are early sequential bilinguals of Spanish and English, showed a perceptual pattern that is more like monolingual Spanish listeners than monolingual English listeners, but still intermediate between the two monolingual groups. Specifically, they perceived [d]/[ɾ] like the L1 Spanish participants, treating them as very different sounds. They perceived the pair [d]/[ð], which is contrastive in English but allophonic in Spanish, like the L1 Spanish participants, as fairly similar sounds. Finally, heritage speakers perceived [ɾ]/[ð], contrastive in both languages, as very different sounds, identical to all other participant groups. The results underscore both the importance of surface oppositions, suggesting the need to reconsider the traditional definition of contrast, as well as the importance of considering level and age of exposure to the second language when studying the perception of sounds by bilingual speakers.

**Keywords:** heritage speakers; consonant perception; speech perception; Spanish; allophones; contrastive sounds; age of exposure; bilingual; L2 perception

## 1. Introduction

Speech perception is a fundamental aspect of language processing, influencing communication and linguistic development. Understanding how heritage, bilingual, and monolingual speakers of Spanish perceive consonant sounds can provide insights into the interplay between language experience and phonological processing. This study hopes to add to the body of literature on the perception of contrastive and allophonic sounds, with a focus on the bilingual listener's linguistic experience and age of acquisition of a second language.

Listeners of one language do not always process sounds in the same way that listeners of another language might. Speech perception is dependent on the language(s) spoken by the listener and the degree and age of their exposure to the language. The phonological inventory of the language spoken affects the degree of perception, with contrast playing a significant role. Sounds that are not contrastive in the listener's native language (L1) will be more difficult, if not impossible, to perceive compared to contrastive sounds in the listener's L1. For example, Japanese listeners have difficulty distinguishing the English liquids [l]/[ɹ] because these sounds are not contrastive in Japanese (MacKain et al., 1981; Flege et al., 1996; Bradlow et al., 1999). In addition, monolingual Spanish speakers perceive English vowel pairs that are non-contrastive in Spanish, such as [i]/[ɪ], to be more similar than do monolingual English speakers (Boomershine, 2013). Additionally, studies in second language learning have found that listeners are more adept at perceiving sounds of their native language than those of a second language acquired later in life, e.g., (Polka & Werker, 1994; Strange, 1995; Dupoux et al., 1997; Best et al., 1988; Francis & Nusbaum, 2002). The study that we report here expands on this research in L2 speech perception in a study of the perceptual consequences of allophony for bilingual and heritage language speakers.

## 2. Background

### *2.1. Speech Perception*

When a person starts to learn a new language (their L2 if they currently speak just one other language—the L1) there are a few different phonetic scenarios that they may face. It has been found that the learning outcome (when it comes to developing a native-like ability to perceive the words of the L2) is determined to a substantial degree by the type of challenge that the phonetic scenario presents (Flege, 1987; Best, 1994; Kuhl, 1991; Kuhl et al., 2008; Escudero, 2005). The easiest case is when a sound in the L2 matches almost exactly a sound that has been learned in the L1. Interestingly, this sort of close match may lead to the longest lasting speech production and perception accentedness—deviation of the L2 speaker from the L1 norms for the language (Best, 1994) but generally the deviation from nativeness is relatively small.

Another scenario is when a sound in the L2 can be associated with a somewhat similar sound in the learner's L1, but only somewhat. The phonetic deviation between the languages can be quite detectable to the phonetician but escape the notice of the learner; for example when a speaker of English is learning Spanish and must produce the voicing contrasts on stops as [$\pm$voice] during the stop closure, versus the English use of [$\pm$aspiration] after the stop closure. In this case, the identity of the corresponding L1 and L2 sounds in the mind of the listener (see Kuhl et al., 2008) leads to what Flege (1987, 1995) called and "Equivalence Classification" which in turn is associated with quite substantial deviation from nativeness in the L2 speaker's pronunciation and perception.

A third type of scenario that the L2 learner may face is one in which there is no identifiable "similar" sound in the L1. This occurs, for example when an English speaker learns a click language like Zulu (Best, 1994). The clicks are not very similar to any English sound, and thus are unlikely to be confused with any existing L1 phoneme for the English speaking learner. In this non-identity scenario it is often the case that the learner can develop quite good perceptual ability in identifying the new sound (Best, 1994).

A fourth scenario that has been a focus of research in the L2 acquisition literature is the case where the languages have mismatched phonological categories (Escudero, 2005). This scenario is similar to the non-identity scenario; however the new sound is similar enough to an existing sound in the learner's L1 that they may perceptually merge two contrasting L2 sounds. For example, the English contrast between /l/ and /r/ is difficult for Japanese speakers to master. The situation is that Japanese has a sound that is similar to both /l/ and

/r/ and so initially at least the learner does not notice that in English these two are different (Lively et al., 1994). It is like an equivalence classification times two—two sounds in L2 are both being equated with a single sound in L1. Chang (2014) describes this scenario as L1 and L2 sounds not being in a 'one-to-one relationship' so that the listener must 'distinguish more categories in the L2 than exist in the L1'. Escudero (2005) also treats this category mismatch scenario as a phonological structure mismatch—where the phonology of L2 divides up a phonetic region into more categories than is done in the phonology of L1. In her theory, the solution to this problem requires a lexically-driven tuning of one's phonetic expectations—with the key element being that lexical contrast drives the development of a new phoneme. Another example of an L2 acquisition scenario that requires phonological recategorization occurs when a Spanish speaker learns English vowels. The region of vowel space that is occupied by the phoneme [i] in Spanish, must be divided between [i] and [ɪ] in English. A new phonological contrast has to be learned—phonetic space has to be more finely divided.

A fifth scenario (not so thoroughly considered in the literature) which Boomershine et al. (2008) studied in monolingual Spanish and English listeners, and to which we return now with Spanish/English bilinguals, is different from all of these. In this scenario of cross-linguistic differences in speech perception, the sounds that are presented to listeners are present in both languages, but they have different allophonic and contrastive relationships to each other in the languages' phonologies. For example, in one language, phones [a] and [b] are allophones of phoneme /A/ but in the other language, [a] is a realization of phoneme /A/ while [b] is a realization of phoneme /B/. So, in this scenario all of the listeners are familiar with the phones [a] and [b], but for one group of listeners the phones are contrastive, while for the other group of listeners they are non-contrastive allophones of the same phoneme. Boomershine et al. (2008) found a perceptual effect of language-specific allophony that differed by the language of the listener such that allophones of a single phoneme are more perceptually similar to each other than the same sounds are when they are constrastive phones. We turn to this fifth scenario on the role of allophony in the next section.

### *2.2. Perceptual Consequences of Allophony*

Research has shown that in addition to the presence or absence of contrast between two sounds in a listener's phonology (for example, that Spanish has the high front vowel /i/ but not the lax vowel /ɪ/) the allophonic relationship between sounds (for example, that Spanish /r/ is realized as either a trill or a tap depending on context) also affects how those sounds are perceived (Babel & Johnson, 2010; Boomershine et al., 2008; Chappell, 2017; Harnsberger, 2001; Huang & Johnson, 2010; Johnson, 2004; Johnson & Babel, 2010). Huang and Johnson (2010) studied the perception of Mandarin tones by native and nonnative speakers. In their study, native Mandarin listeners demonstrated sensitivity to tone contours while the English listeners attended to pitch levels during the task. The Mandarin speakers rated tones that had been contrastively neutralized due to tone sandhi rules in Mandarin as being more similar than did the English listeners—meaning the Mandarin speakers tapped into the allophonic relationships found in their tone system when completing the perceptual rating task. Harnsberger (2001) studied the perception of contrastive and allophonic nasals by Malayalam listeners, finding a near merger of allophonically-related dental and alveolar nasal consonants. These coronal nasals are in complementary distribution in the language, with the dental occurring morpheme-initially and the alveolar occurring both morpheme-finally and intervocalically. Contrastive nasals such as bilabial [m] versus velar [ŋ], on the other hand, showed greater perceptual separation in Harnsberger's study.

Boomershine et al. (2008) found that participants in their speeded discrimination and similarity rating tasks tapped into their L1 phonologies when rating the contrastive and allophonic pairs of [d]/[ɾ], [ɾ]/[ð], and [d]/[ð]. Spanish and English listeners were asked to rate these pairs using a five-point similarity scale. In Spanish, [d] and [ð] are allophones and [ɾ] is its own phoneme while in English, [d] and [ɾ] are allophones and [ð] is its own phoneme. The Spanish listeners rated [d] and [ð] as being similar while the most similar pair for the English listeners was [d] and [ɾ]. In other words, the listeners rated the pairs that are allophones in their L1 as being more similar than the other pairs. The same results were found in the speeded discrimination task with the same groups of participants, with those pairs that were allophones in Spanish ([d]/[ð]) being responded to more slowly by Spanish participants when compared to their reaction times to pairs containing an allophone and a phoneme ([d]/[ɾ], [ð]/[ɾ]). The pairs that contained allophones in English ([d]/[ɾ]) were responded to more slowly by English listeners when compared to their reaction times to pairs containing one allophone and one phoneme ([d]/[ð], [ɾ]/[ð]).

Another recent study that has investigated the role of allophony on speech perception is Babel and Johnson's (2010) study of fricative perception by Dutch and English listeners. Their study included voiceless fricatives which were either contrastive in both languages, contrastive in only one of the languages, or allophones in one of the languages. For now, we will focus on the latter category and review how their participants perceived [s] and [ʃ]. In English, these phones are contrastive, given minimal pairs such as 'sip' and 'ship', respectively. In Dutch, on the other hand, these sounds are not used contrastively. In the perceptual similarity task, Dutch listeners rated [s] and [ʃ] as being significantly more similar compared to the ratings of the English listeners.

In her 2017 study, Chappell investigated how well monolingual speakers of Costa Rican Spanish are able to perceive differences in allophonic and contrastive pairs in Spanish. The participants were asked to complete similarity rating and AX discrimination tasks, evaluating word pairs that were identical or differed only in one phoneme or allophone. She found that the listeners perceived phonemic contrasts best, followed by allophonic pairs and then identical sounds. That said, not all allophonic pairs were perceived equally, with the [s]~[z] pair being the most difficult to distinguish compared to other allophonic pairs. Chappell contends that those allophonic pairs that encode linguistic meaning are more salient than those that do not.

### *2.3. Bilingualism and Speech Perception*

Several studies have investigated the perception of sounds by early bilinguals, including the study by Bosch et al. (2000). They found that early bilinguals of Catalan and Spanish exhibited distinct vowel perception patterns for each language, suggesting that bilingual experience influences vowel categorization. They assert that balanced bilingual Catalan and Spanish speakers maintain separate vowel systems, which in turn affects their perceptual processing of vowels.

Other studies have investigated the perception of vowels by Catalan and Spanish bilinguals. In their 2005 study, Navarra et al. (2005) used an implicit method for measuring the L1 effects on the perception of L2 sounds. They asked Catalan-Spanish simultaneous bilinguals who either grew up in Spanish-speaking homes or Catalan-speaking homes to categorize the first syllable of bisyllabic stimuli, with the only difference in the stimulus items being the vowel in the second syllable—it could contain a Catalan contrastive variation (/ɛ/–/e/) or no variation. Catalan dominants responded more slowly in lists where the 2nd syllable could vary from trial to trial, suggesting an indirect effect of the /ɛ/–/e/ discrimination. Spanish dominants, however, did not suffer this interference, performing indistinguishably from Spanish monolinguals. These findings seem to suggest that even simultaneous bilinguals' perceptual systems are affected by their L1 phonology.

A similar study was conducted by Mora and Nadeu (2012) to study the effects of an L2 (Spanish) on the perception of L1 (Catalan) contrastive sounds. In their study, they asked L1-Catalan, L2-Spanish speakers who lived in a Catalan-dominant community and L1-Catalan, L2-Spanish speakers living in a Spanish-dominant community to complete an identification task and an AXB discrimination task. They found that both groups performed at near ceiling levels in the tasks, but the reaction time for those participants who have more exposure to Spanish than Catalan was slower than for the Catalan-dominant speakers. The findings suggest that extensive L2 use/exposure to Spanish and Spanish-accented Catalan in a bilingual language contact setting may modify Catalan natives' phonetic category /ɛ/.

In their study, Ning et al. (2022) studied the effect of age of L2 acquisition on the perceptual processing of lexical tones in Cantonese and Urdu speakers. Three groups of subjects, Cantonese monolinguals, Cantonese/Urdu simultaneous bilinguals, and late L1 Urdu/L2 Cantonese bilinguals, living in Hong Kong participated in a four-condition ABX task. When there were no conflicts, the simultaneous bilinguals were able to process Cantonese tones like Cantonese L1 speakers. However, when conflicts were introduced, the simultaneous bilinguals had significantly slower reaction times and produced more errors when compared to native Cantonese speakers on the same tasks. Even with years of schooling and exposure to Cantonese in and outside of the home, simultaneous bilinguals were still unable to process Cantonese tones like native Cantonese speakers.

While there has been some research on how bilinguals who live in bilingual settings where both languages are used extensively in the speech community, there has been little perception research on heritage speakers of Spanish living in the United States. Heritage speakers of Spanish differ significantly from early Catalan/Spanish bilinguals in Spain or simultaneous Cantonese/Urdu bilinguals in Hong Kong, for example, in that heritage speakers of Spanish grow up hearing and speaking Spanish in the home from birth, and then, in the case of the US, learn English upon entering school. In the US, they generally live in English-dominant communities with little to no access to schooling or literacy in Spanish. The domains available for Spanish exposure are limited to the home and culturally specific sites in the community, including church, social outings/festivities, and cultural celebrations.

### *2.4. Heritage Speakers*

The current study repeats the Boomershine et al. (2008) rating task with three new groups of listeners—bilinguals dominant in English or Spanish, and heritage speakers of Spanish in the US who grew up hearing speaking Spanish in the home, and then learned English upon entering school. The bilingual speakers in Boomershine et al. (2008) began learning English as their L2 as teenagers (Table 2). In this study we were interested in learning whether heritage speakers would pattern more like native speakers of English, given that they learned English early, had no formal education in Spanish, and that the experiment was conducted in English at an institution where English is the main mode of communication. To preview the results, bilingual speakers patterned with monolingual speakers of their L1, and heritage speakers of Spanish patterned between monolingual speakers of English and Spanish to some extent, but their pattern of responses was much more like that of Spanish monolinguals.

Given the research that has found a language effect for the perception of contrastive and allophonic sounds, with babies as young as 6 months able to perceive sounds in their L1 (Best, 1994; Kuhl et al., 1992), the early linguistic experiences of heritage speakers of Spanish are unique and given their early and extensive experience with a second language, it is not at all certain that they would pattern with other bilingual speakers who maintain active use of both languages. The limited number of studies on the speech perception of heritage speakers of Spanish (HSS) has found that HSS perceive certain sounds like Spanish speakers and other sounds like English speakers (Kim, 2011; Boomershine, 2013). Studies have shown that heritage speakers often exhibit differences in consonant perception compared to monolingual speakers. For example, Kim (2011) conducted a production and perception study of English and Spanish stop consonants with heritage Spanish speakers who were English dominant. In the production study, she found that the heritage Spanish speakers had VOTs that were nearly identical to native English speakers and significantly different from those of native Spanish speakers. In her perception study, Kim had participants listen to carrier phrases containing natural and cross-spliced stimuli and then identify the word they heard. For the natural stimuli, the HSS perceived the English stops like English speakers and the Spanish stops like Spanish speakers—in other words, like two monolinguals. However, when presented with the cross-spliced stimuli, the heritage Spanish speakers patterned like native Spanish speakers in that they judged the cross-spliced stimuli produced with pre-voicing to be voiced, whereas the native English speakers relied more on cues from the vowel portion of the stimuli when identifying the stimulus item as voiced or voiceless.

Research with HSS perception of English vowels, however, found differing results. Boomershine (2013) asked heritage, monolingual, and (late) bilingual speakers of Spanish and English to listen to pairs of English front vowels and rate them on perceptual similarity. In this study, the HSS reported being exposed to Spanish at birth and to English on average at age 2. The bilingual participants reported being exposed to their L2 on average at age 11. She found several significant differences between the heritage and monolingual groups, but no significant differences between the HSS and the L1 Spanish bilinguals. In addition, the HSS patterned similarly to the L1 English bilinguals for all vowel pairs except 'bait'/'bit', where the HSS found this pair to be more similar than did the L1 English bilinguals.

Other studies suggest that heritage speakers can maintain robust perception skills of consonants as well, finding that early exposure to the heritage language has long-lasting effects on heritage speakers' ability to distinguish heritage language phonological contrasts. When compared to Hispanic immigrants who had lived in the US for a long period of time and to Spanish native controls who had recently moved to the US, Spanish heritage speakers performed similarly to the control group when distinguishing the Spanish stop voicing contrast (Mazzaro et al., 2016). Another study comparing the perception of consonant perception by heritage Spanish speakers and Spanish native controls found that when listening to stimuli with contrasting acoustic information in the consonant portion and the vowel portion (/b/ from /be/ + /e/ from /pe/), both the heritage speakers and the native controls attended to the consonant portion more than the vowel portion (Kim, 2011).

Research on consonant perception among heritage, bilingual, and monolingual speakers of Spanish highlights the complex interplay between language experience and phonological processing. Given the limited perception research on HSS, the current paper has as its goal to better understand the perceptual processing system of heritage speakers, especially with respect to the perception of contrastive and allophonic sounds, as compared to late bilinguals and monolingual speakers of Spanish and English. As native or near-native speakers of both Spanish and English, HSS are in the unique position to shed light on how the bilingual brain processes sounds that have different phonological relationships in the two languages spoken.

Given what is known about how heritage speakers differ from late bilingual speakers in speech perception, the researchers pose the following two hypotheses:

**H1:** *Heritage speakers of Spanish will perceive contrastive sounds in their L1, Spanish, with a similar pattern of their late L1 Spanish bilinguals due to their early exposure to Spanish.*

**H2:** *Heritage speakers of Spanish will exhibit some influence from their L2, English, when perceiving pairs that are allophonic in their L2, as they have more continuous and consistent exposure to English as college students living in a majority-English community.*

To test these hypotheses, the current study will partially replicate the Boomershine et al. (2008) research on the perception of the phones [d], [ð] and [ɾ], this time testing the perception of heritage speakers of Spanish as compared to their monolingual and bilingual speakers of English and Spanish participants in a similarity rating task.

## 3. Materials and Methods

The three consonants [d], [ð], and [ɾ] exist in both the Spanish and English phonetic inventories, but have different roles in their phonological systems. In English, [d] and [ɾ] are allophones of the same phoneme, while [ð] and [d] are contrastive. On the other hand, [d] and [ð] are allophones of the same phoneme in Spanish, and [d] and [ɾ] are phonemically contrastive. On the surface, however, [d] and [ɾ] generally do not appear in the same phonetic contexts and thus are not lexically contrastive in most varieties of Spanish, including Mexican Spanish, the variety spoken by the participants in the current study (Hualde, 2013). There is, though, a surface contrast between [ð] and [ɾ] for all varieties of Spanish ([ka.ða] 'each' vs. [ka.ɾa] 'face'). These relationships are illustrated in Table 1.

**Table 1.** Relationships and examples of [d], [ð] and [ɾ] in Spanish and English.

| **Sound Pairs** | **d/ɾ** | | **d/ð** | | **ɾ/ð** | |
|---|---|---|---|---|---|---|
| **Language** | Spanish | English | Spanish | English | Spanish | English |
| **Underlying contrast** | (yes) | no | no | yes | (no) | no |
| **Surface contrast** | no | no | no | yes | yes | yes |
| **Allophony** | no | yes | yes | no | no | no |
| **Examples** | N/A | Fred<br>Freddy | en diez 'in ten'<br>a diez 'to ten' | dough<br>though | cara 'face'<br>cada 'each' | ladder<br>lather |

In order to explore the role that age and type of exposure play on the perception of contrastive and allophonic sounds in Spanish and English, the perceptual similarity rating task used by Boomershine et al. (2008) was utilized in the current study, with a new panel of listeners, including monolingual speakers of English and Spanish, bilingual speakers of English and Spanish whose L1 was either Spanish or English, and heritage speakers of Spanish. The same audio recordings that were used in Boomershine et al.'s (2008) study were used in the current study and consisted of two tokens of each of the following VCV sequences: [ada], [aða], [aɾa], [idi], [iði], [iɾi], [udu], [uðu], and [uɾu]. Boomershine et al. (2008) recorded multiple tokens of these which were produced by two native speakers of Greek, one male and one female, using a head-mounted microphone in a soundproof booth. Greek speakers were chosen because all three of the test phones, [d], [ð], and [ɾ], are contrastive in Greek and are produced naturally in intervocalic position. The speakers attempted to produce equal stress on the first and second syllables. In order to control the amplitude across tokens and speakers, the peak amplitude was equated for each of the tokens. The two best recordings for each VCV sequence were used as stimuli in their study and in this one.

Table 2 depicts the demographic information for the 50 participants in the current study. It should be noted that the heritage speakers of Spanish are considered native speakers of Spanish but were given their own category in order to distinguish their age and amount of exposure to English. The heritage speakers of Spanish were exposed to Spanish at birth and reported being exposed to English, on average, at the age of 4.3 years, usually when they started attending pre-kindergarten. This is considerably earlier than that of the late bilinguals, with an average age of exposure to English for the native Spanish, advanced English participants of 12.5 years and an average age of exposure to Spanish for the native English, advanced Spanish participants of 13.2 years. The participants labeled in this study as monolingual speakers of either Spanish or English reported having no or very little exposure to the other language, with an average self-rating of English proficiency for the monolingual Spanish speakers of 2.1 and an average self-rating of Spanish proficiency for the monolingual English speakers of 2.9 using a scale of 1–7, with 7 being 'native proficiency'. The monolingual speakers who were exposed to an L2 reported receiving this exposure as young adults in academic settings.

**Table 2.** Demographic information for participants. * Participants self-rated their L1 and L2 proficiency using a scale from 1–7.

| Native Language | L2 Language | M | F | Total | Avg Age | Average Age of L2 Exposure | * L1 Prof. Self-Rating | * L2 Prof. Self-Rating |
|---|---|---|---|---|---|---|---|---|
| **Native English** | Little to no Spanish | 1 | 11 | 12 | 20.08 | 19 | 7 | 2.9 |
| | Advanced Spanish | 6 | 2 | 8 | 26.5 | 13.2 | 7 | 6.19 |
| **Native Spanish** | Little to no English | 6 | 4 | 10 | 24.4 | 21 | 7 | 2.1 |
| | Advanced English | 2 | 5 | 7 | 32 | 12.5 | 7 | 5.28 |
| **Heritage Spanish** | Native/Fluent | 4 | 9 | 13 | 19.3 | 4.3 | 6 | 6.97 |

The participants in this study were affiliated with the University of North Carolina Wilmington as students, staff, or acquaintances of students or staff. Thirteen heritage speakers of Spanish participated in the current study. They were all students at the University of North Carolina Wilmington and were enrolled in the advanced Spanish for Bilingual Speakers course. They received course credit for participating in this study. The participants completed a post-experiment demographic questionnaire, and only those participants that were exposed to Spanish in the home since birth, had parents from Mexico (and thus spoke Mexican Spanish), and were exposed to English by age 5 were included in the study. The L1 Spanish bilingual speakers were all native speakers of Mexican Spanish that reported that their primary exposure to Spanish was with other speakers of Mexican Spanish. Speakers of other varieties of Spanish were excluded from this study in order to control for variation in the realization of intervocalic stops in some varieties of Spanish. All L1 English participants in this study were speakers of American English and all L2 speakers of English reported having lived in the US as their only English-speaking country of residence. None of the participants reported any hearing, speech, or language disorders.

Participants were seated in front of a PC with an attached five-button response box and were presented with the stimuli through headphones. Participants heard pairs of physically different stimuli separated by one second of silence, such as [ada] < 1 s silence> [aða], and responded with a rating score from 1 (very similar) to 5 (very different). The pairs were presented in a different random order for each participant, using E-Prime software (v. 1.1; Psychological Software Tools, Pittsburgh, PA, USA). The talker and vowel context were the same for every pair so that the only difference in each pair was the consonant. The stimuli presented in each pair were always physically different tokens, even when they were both examples of a single sound (e.g., [ada]. . .[ada]). The participants were given four practice trials, and then the opportunity to ask questions before proceeding to the four test blocks (360 test trials total). Participants were able to take a short break between test blocks if needed, and the average participation duration was less than one hour. They received no feedback in this experiment.

Boomershine et al. (2008) also presented pairs for speeded discrimination ('same' vs. 'different' responses) and measured perceived similarity as a function of the time it takes to press the 'different' button when the pair is physically different. We chose in this study to only use the rating task because it is statistically less noisy and thus more reliable, and because it taps a level of perception that is more sensitive to linguistic knowledge, as opposed to being based on purely auditory discrimination.

In order to account for some participants not using the response option endpoints, the rating scores for each speaker were normalized to compensate for differences in use of the 5-point scale. The scores were normalized using a standard z-score transformation, such that each participant's scores were centered around 0, with scores above zero indicating "more different" and scores below zero indicating "more similar". This normalization allows the participants' similarity ratings to be compared with each other and with those of the previous study (Boomershine et al., 2008).

## 4. Results

This study compared the normalized similarity ratings of three consonant pairs across five groups of participants with varying levels of experience and age of exposure to Spanish and English. The results of this similarity task were analyzed across sound pairs. We will first look at the results by the first language of each participant, followed by the results based on second language exposure of each participant group.

### *4.1. L1 Effects on Perception*

To better understand the role of L1 on the perceptual rating of contrastive and allophonic sounds, the participants were first divided into two groups based on their first language (the language they first acquired). In this study, 30 of the 50 total participants acquired Spanish as their first language (the monolingual Spanish, L1 Spanish bilingual, and heritage Spanish speakers), with the remaining 20 participants being L1 speakers of English (monolingual and L1 English bilingual speakers). The normalized results by the first language of the participants (leaving out the heritage speakers) are shown in Figure 1.

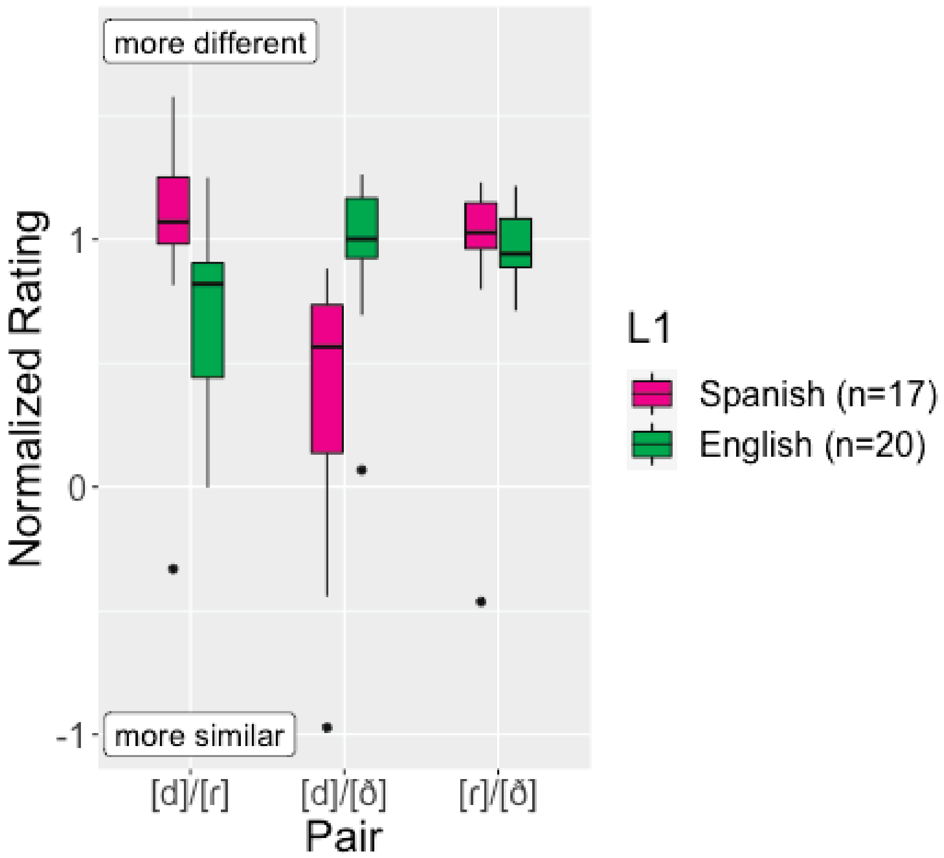



**Figure 1.** Rating task results. Normalized similarity rating for the pairs [d]/[ɾ], [d]/[ð], and [ɾ]/[ð] by first language (L1) of the participants.

A repeated measures ANOVA revealed a significant main effect for consonant pair, F(1,65) = 4.835, $p < 0.05$, $\eta^2 = 0.092$, as well as a significant interaction of first language on the similarity rating of consonant pairs, F(1,65) = 17.854, $p < 0.001$, $\eta^2 = 0.271$.

Overall, the perception results patterned around the participants' first language, with contrastive sounds in each L1 being perceived as more different for that language group. To determine if these similarity ratings were significantly different based on the first language of the participants, independent samples t-tests were run. In both Spanish and English, [ɾ]/[ð] has a surface contrast (as in 'cara'/'cada' and 'ladder'/'lather', respectively), leading both participant groups to rate those sounds as being different ($p = 0.761$). L1 (English vs. Spanish) did not produce a significant difference in similarity rating as both L1 English and L1 Spanish participants rated [ɾ]/[ð] as being different. The pair [d]/[ð], with a surface contrast in English (as in 'dough'/'though') and an allophonic relationship in Spanish (as in 'en dos'/'a dos'), was found to be significantly more different by the L1 English participants than by the L1 Spanish participants ($p < 0.001$). While the pair [d]/[ɾ] does not have a surface contrast in Spanish or English, it does have an underlying contrast in Spanish and is allophonic in English. Accordingly, the L1 Spanish participants found that pair to be significantly more different compared to the L1 English participants ($p < 0.001$).

### 4.2. L2 Exposure Effects on Perception

Knowing that a person's first language affects their rating of these consonant pairs, the participants were subdivided into five groups based on the level of experience in their L2 and their age of exposure to that L2. The five participant groups are monolingual English, monolingual Spanish, L1 Spanish bilingual (late sequential bilingual), L1 English bilingual (late sequential bilingual), and heritage Spanish (early sequential bilingual) speakers. The normalized similarity ratings of the three consonant pairs across five groups of participants with varying levels of experience and age of exposure to Spanish and English are shown in Figure 2. To determine the effects of L2 exposure on the perception of these consonant pairs, a repeated measures ANOVA was conducted to compare the means of the five groups, revealing a significant effect of participant group on the similarity rating of consonant pairs, $F(1,62) = 6.005$, $p < 0.05$, $\eta^2 = 0.109$.

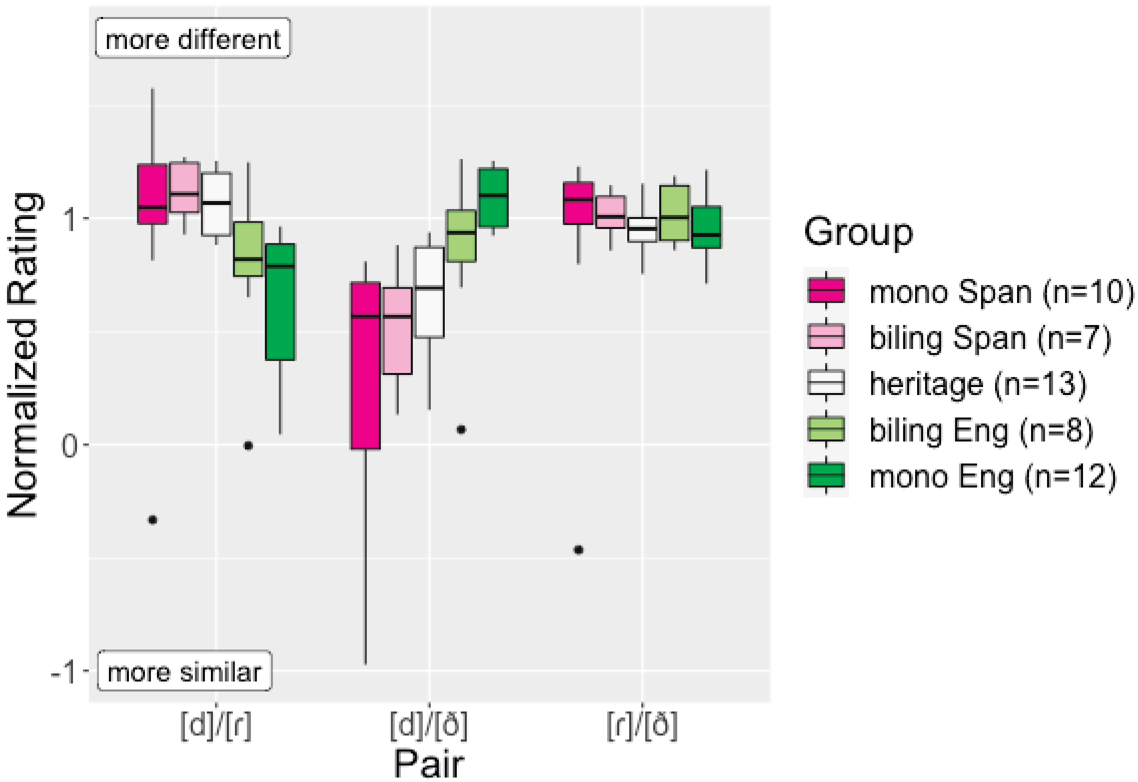



**Figure 2.** Rating task results. Normalized similarity rating for the pairs [d]/[ɾ], [d]/[ð], and [ɾ]/[ð] for all participant groups.

In order to determine if the participants' language experience had a significant effect on their perception of these pairs of sounds, a Tukey's HSD post hoc test was conducted. For the pair [ɾ]/[ð], the results showed that there was no significant difference by participant language group, with all participant groups rating that pair as being different ($p = 0.901$). This is expected since this pair is contrastive in both English and Spanish.

For the pair [d]/[ð], the Tukey's HSD post hoc test showed a significant difference between monolingual English participants and heritage Spanish participants ($p < 0.05$), monolingual English participants and L1 Spanish bilingual participants ($p < 0.05$), and monolingual English participants and monolingual Spanish speakers ($p < 0.01$). There was also a significant difference between the L1 English bilinguals and the monolingual Spanish participants ($p < 0.05$). This pair is contrastive in English but allophonic in Spanish.

Tukey's HSD post hoc test results for the pair [d]/[ɾ] revealed a significant difference between the monolingual English speakers and L1 Spanish bilinguals ($p < 0.05$) and the monolingual English speakers and the heritage Spanish speakers ($p < 0.05$). There was not a significant difference among the participants for whom Spanish is their first language, regardless of their English level, nor among the participants for whom English is their first language, also regardless of their Spanish level. This pair is allophonic in English and has an underlying contrast in Spanish, but no surface contrast in either language.

## 5. Discussion

Previous research has found that across languages, speakers of a language in which a particular pair of sounds is contrastive at a phonemic level perceive that pair as being more perceptually distinct when compared to speakers of a language in which the pair is not phonemically contrastive (Boomershine et al., 2008). The present study aimed to examine how monolingual and early and late bilingual speakers of Spanish and English perceive sounds whose phonological structures differ across the two languages, and specifically whether heritage speakers of Spanish in the US perceived sounds using their first language (Spanish) or their dominant (English) sound system.

### 5.1. L1 Effects on Perception

The first language of participants in this similarity rating task was found to significantly impact their perception of contrastive and allophonic sounds, with those pairs that are allophonic in the L1 being rated as more similar than sounds that are contrastive in that language. The results of this study align with those of the Boomershine et al. (2008) study in which the L1 speakers of English found the contrastive pair [d]/[ð] to be different while the L1 speakers of Spanish, where this pair is allophonic, found it to be similar, even for L1 speakers of Spanish who had very early exposure to English and considered themselves to be English-dominant. The opposite is true for the pair [d]/[ɾ], which is allophonic in English, resulting in L1 speakers of English finding it to be similar when compared to the L1 speakers of Spanish.

The findings of these studies add to evidence that the presence or absence of a sound in a language's inventory is not the only source of information when a speaker perceives that and other sounds. The phonological relationship of sounds within a language is an integral piece to the perception puzzle. Sounds that are contrastive, either at the surface or underlying level, are perceived as more different by native speakers of that language than are sounds that are allophonic in that language (Kuhl, 2000; Pisoni, 1997; Johnson, 1997).

### 5.2. L2 Exposure Effects on Perception

When the participants are grouped not only by their first language but are further grouped by their age and type of exposure to their L2, we find that both early and late bilinguals tend to perceive sounds in a way that is similar to their L1 monolingual counterparts. The age and type of exposure to the L2 was clearly a factor in the similarity rating of the pair [d]/[ð], which is allophonic in Spanish but contrastive in English. The monolingual English speakers rated that pair as the most different, followed by the L1 English (late) bilinguals, with both groups rating it significantly more different than monolingual Spanish speakers. The monolingual Spanish speakers found the pair to be the most similar, followed closely by the L1 Spanish (late) bilinguals. The heritage speakers of Spanish, who were exposed to Spanish since birth but acquired English at an average age of 4.3 and were then surrounded by English throughout their lives including in school, in the community, etc, found this pair to be significantly more similar when compared to monolingual English speakers, as did the L1 Spanish (late) bilinguals and Spanish monolinguals. The fact that the heritage Spanish speakers patterned like monolingual and late bilingual speakers of Spanish, rather than monolingual English speakers provides evidence that even early sequential bilinguals tend to perceive sounds with their L1 phonology. However, it is notable that the heritage speaker ratings were also not reliably different from those of the L1 English bilinguals.

Turning to the pair [d]/[ɾ], which is allophonic in English but has an underlying contrast in Spanish, the monolingual English speakers found that pair to be more similar than did the Spanish L1 groups. Interestingly, the heritage speakers of Spanish again had a significantly different similarity rating for this pair as they did for [d]/[ð] when compared to monolingual English speakers, rating it as a more different sound than the English speakers did. This finding follows previous studies on L1 and L2 speech perception that found that even early bilinguals perceived L2 sounds by reference to perceptual expectations that are based on their L1 phonological categories (Archila-Suerte et al., 2012; Baigorri et al., 2019; Flege, 1991; Flege et al., 1999; Flege & MacKay, 2004; Levy, 2009; Levy & Strange, 2008; Morrison, 2002; Pallier et al., 1997). So, for example, Flege et al. (1997) found that L1 Korean listeners' perception of the English contrast between the vowels of "beat" and "bit" (which is signaled by both vowel duration and vowel formant frequency differences) used vowel duration to distinguish the vowels and did not use the vowel formant differences. It should be noted that the Korean listeners in the study were not heritage speakers of Korean, but instead were L1 speakers who were exposed to English upon moving to the US as adults. Flege et al. attributed this pattern of results to the fact that Korean has long and short i: versus i, but no distinction between tense [i] and lax [I]. Returning to heritage speakers of Spanish for a moment, it is interesting to find that these participants patterned nearly identically to L1 Spanish (late) bilinguals, even though the heritage speakers in this study were exposed to English at a much younger age (4.3 years of age) than the L1 Spanish (late) bilinguals (12.5 years of age). In addition, the heritage speakers of Spanish's exposure to both Spanish and English was different from that of their L1 Spanish late bilingual counterparts, as the heritage speakers were born and raised in a country where English is the majority language (the United States), whereas the L1 Spanish (late) bilinguals were born and raised in a Spanish-speaking country. The amount of exposure to schooling in English and Spanish differs across the two groups, with heritage speakers receiving schooling exclusively in English up until high school, where they may have taken a Spanish elective course, and the L1 Spanish (late) bilinguals attending school in Spanish, with some exposure to elective English classes at times. As seen in Table 2, the heritage speakers of Spanish self-rated as being dominant English speakers (6.97 out of 7) as compared to a score of 6 out of 7 for Spanish, whereas the L1 Spanish (late) bilinguals rated themselves as dominant Spanish speakers (7 out of 7) and less proficient in English (5.28 out of 7). Montrul's (2008) study on the linguistic abilities of heritage speakers of Spanish suggests that heritage speakers may face difficulties in distinguishing phonemic contrasts in their L1 that native speakers can easily perceive. Similarly, Rao and Myers' (2014) found that heritage Spanish speakers perceive vowel duration and prosodic contours differently from native speakers, often leading to non-native-like speech patterns. The results of the current study differ from those of Montrul and Rao and Myers, finding that heritage Spanish speakers' perception was in fact very similar to that of their L1 Spanish (late) bilingual counterparts. More research needs to be done to determine how heritage speakers of Spanish perceive other allophonic vs. contrastive sounds in the languages they speak to determine if the pattern found in this study is due to L1 sound categories or if there are other factors at play.

The results reported here suggest a nuanced view of the critical period hypothesis for second language bilingualism. Two groups of L1 Spanish, L2 English bilinguals were observed to have responded largely in line with Spanish monolinguals: early sequential bilinguals (heritage speakers), and later L2 English learners. Though the heritage speakers showed greater similarity to the English monolinguals' pattern of responses in the allophone similarity task, as might be expected by their early acquisition of English, it is noteworthy that their pattern of responses was strikingly similar to that of Spanish monolinguals. Heritage speakers learned English well within the critical time period for language acquisition as most authors would delimit it, yet their perception responses were very similar to native Spanish listeners: both monolinguals and bilingual listeners who acquired English later in life. These findings are similar to those reported by Kim

(2016) in her study comparing the perception and production of Spanish lexical stress by heritage and L2 Spanish speakers, finding that heritage speakers did not differ statistically from native speakers in the perception of Spanish lexical stress, unlike their L2 Spanish counterparts who performed at chance. This, coupled with our current findings, suggests that very early linguistic experience has a lasting effect on speech perception.

*5.3. Theoretical Implications*

Of the theoretical approaches to L2 speech perception that we presented in Section 2.1, Kuhl et al.'s (2008) perceptual magnet theory is the most evocative of the allophonic similarity phenomenon, in that a linguistic category serves as a metaphorical "magnet" in that members of the category are more similar to each other than would be expected on purely psychophysical grounds. However, like the recategorization phenomenon that inspired Escudero's (2005) phonological theory of L2 perception and production, allophonic similarity is a phonological effect.

What has been demonstrated in this paper (replicating and extending the findings of Boomershine et al., 2008) is that allophonic effects in perception are language-specific and may be tied to the amount and type of the listener's linguistic experience. So, what kind of theory could account for experience-based language-specific allophonic perceptual similarity? Although there may be several alternatives, we will just mention one here: the exemplar resonance theory of Johnson (2006). This model envisions an activation loop between abstract and phonetic levels of representation, where the phonetic level is populated by phonetically detailed exemplar traces in memory. The resonance loop feeds activation from abstract categories back into the exemplar cloud which can result in co-activation of exemplars of different allophones of the same phoneme. This co-activation of the exemplars of allophones results then in increased similarity between the perceptual activation patterns generated by the allophones of a phoneme -even when they are not terribly phonetically similar—like [d] and [ð]. And this increased similarity in activation patterns will result, so the model predicts, in decreased perceptual separation between them. This model predicts the pattern of results found in this study, that allophones will be more perceptually similar to each other than would be predicted by their psychophysical representation alone, and that this pattern of perceptual similarity will be language-specific (driven by the language's phonology) and will be sensitive to the listener's specific linguistic experience (their clouds of exemplars).

However, the data here are challenging for the exemplar resonance theory because the 'abstraction' layer in that model is lexical/semantic (Johnson, 2004), so in an experiment such as the one reported here we have to imagine that the non-word [aCa] stimuli activate specific lexical items. Future research developing exemplar resonance theory will need to develop a specific account for how non-words can activate lexical items and how activation in a bilingual linguistic system may be regulated. Particularly, it seems important to note that the L1 pattern of allophony-driven speech perception is evidently learned at an early stage of L1 acquisition, and persists despite the early and extensive exposure to L2 that is experienced by heritage speakers. This is consistent with the suggestion that "the speech perceptual system does not seem to be prone to modify initial phonemic categories." (Sebastián-Gallés & Soto-Faraco, 1999, p. 112; see also Mack, 1989; Pallier et al., 1997).

*5.4. Limitations and Future Directions*

While this study begins to address the question of age of L2 exposure as a factor in the perception of contrastive and allophonic sounds, there are some limitations. Ideally, more participants from each of the L1/L2 groups would be included in a future study. Also, future studies could include simultaneous bilinguals, or speakers of Spanish and

English that acquired both languages from birth, rather than only early and late sequential bilinguals as were included in this study. Additionally, it would be useful to see how heritage speakers of Spanish who had received academic exposure to Spanish and English from an early age, such as via a Dual Language/Immersion program, would perform in this task, especially based on the research already available concerning not just the age of acquisition, but also the availability of input (Montrul, 2016).

Additionally, the realization of the consonants in question varies by dialect, in both English and Spanish. In this study, only speakers of Mexican Spanish and American English were included to minimize dialectal differences in production and perception. However, it would be of interest to have speakers of British English, where intervocalic alveolar stops are not produced as taps, participate in this study, as it would for speakers of some Central American and highland varieties of Spanish where voiced stops often do not undergo lenition in expected environments, instead being realized as voiced stops (Carrasco et al., 2012).

Turning to the discussion on the role of the phonological relationship of L1 and L2 sounds in speech perception, more studies are needed in this area. For instance, a study that compares the perception and production of the sounds in the current study would be helpful in determining the heritage and late bilinguals' proficiencies with the two languages generally, and the sounds at hand, in particular. It would also be helpful to have studies similar to this one but that look at the perception of other allophonic and contrastive pairs to determine if they pattern the same way as these sounds do. For example, how do heritage speakers of Spanish perceive nasals as compared to bilingual and monolingual speakers? This and other studies looking at the role of contrast and allophony in the perception of L1 and L2 sounds by heritage and late bilingual speakers will continue to shed light on how the bilingual brain processes sounds.

**Author Contributions:** Conceptualization, A.B. and K.J.; methodology, A.B. and K.J.; formal analysis, A.B. and K.J.; investigation, A.B.; writing—original draft preparation, A.B.; writing—review and editing, A.B. and K.J. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of the University of North Carolina Wilmington (H1213-023, 9/10/2012).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The data presented in this study are available on request from the author.

**Acknowledgments:** The authors would like to thank Kathleen Currie Hall and Beth Hume for their work on this topic, including their collaboration on the design of the study and stimuli, as well as their encouragement to extend the study to heritage Spanish speakers.

**Conflicts of Interest:** The author declares no conflicts of interest.

## References

Archila-Suerte, P., Kevin, J., Bunta, F., & Hernandez, A. (2012). Age of acquisition and proficiency in a second language independently influence the perception of non-native speech. *Bilingualism: Language and Cognition*, *15*(1), 190–201. [CrossRef]

Babel, M., & Johnson, K. (2010). Accessing psycho-acoustic perception and language-specific perception with speech sounds. *Laboratory Phonology*, *1*(1), 179–205. [CrossRef]

Baigorri, M., Campanelli, L., & Levy, E. S. (2019). Perception of American–English vowels by early and late Spanish–English bilinguals. *Language and Speech*, *62*(4), 681–700. [CrossRef] [PubMed]

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. C. Goodman, & H. C. Nusbaum (Eds.), *The development of speech perception* (pp. 167–224). MIT Press.

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance 4*, 45–60. [CrossRef] [PubMed]

Boomershine, A. (2013). The perception of English vowels by monolingual, bilingual, and heritage speakers of Spanish and English. In *Selected proceedings of the 15th hispanic linguistics symposium* (pp. 103–118). Cascadilla Proceedings Project.

Boomershine, A., Hall, K. C., Hume, E., & Johnson, K. (2008). The Impact of Allophony vs. Contrast on Speech Perception. In K. Rice (Ed.), *Contrast in phonology: Perception and acquisition*. Mouton de Gruyter. [CrossRef]

Bosch, L., Costa, A., & Sebastián-Gallés, N. (2000). First and second language vowel perception in early bilinguals. *European Journal of Cognitive Psychology*, *12*(2), 189–221. [CrossRef]

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics*, *61*(5), 977–985. [CrossRef]

Carrasco, P., Hualde, J. I., & Simonet, M. (2012). Dialectal differences in Spanish voiced obstruent allophony: Costa Rican versus Iberian Spanish. *Phonetica*, *69*, 149–179. [CrossRef]

Chang, C. B. (2014). Transfer effects in perception of a familiar and unfamiliar language. *The Journal of the Acoustical Society of America*, *135*(4_Supplement), 2355–2355. [CrossRef]

Chappell, W. (2017). Costa Rican Spanish speakers' phonetic discrimination. *Journal of Experimental Phonetics*, *26*, 13–61.

Dupoux, E., Pallier, C., Sebastian, N., & Mehler, J. (1997). A destressing "deafness" in French? *Journal of memory and language*, *36*(3), 406–421. [CrossRef]

Escudero, P. (2005). *Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization*. LOT Dissertation Series 113. Utrecht University.

Flege, J. E. (1987). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*, *15*, 47–65. [CrossRef]

Flege, J. E. (1991). The interlingual identification of Spanish and English vowels: Orthographic evidence. *The Quarterly Journal of Experimental Psychology Section A*, *43*(3), 701–731. [CrossRef]

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). York Press.

Flege, J. E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, *25*, 437–470. [CrossRef]

Flege, J. E., & MacKay, I. R. A. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition*, *26*(1), 1–34. [CrossRef]

Flege, J. E., MacKay, I. R. A., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *The Journal of the Acoustical Society of America*, *106*(5), 2973–2987. [CrossRef] [PubMed]

Flege, J. E., Takagi, N., & Mann, V. (1996). Lexical familiarity and English-language experience affect Japanese adults' perception of /ɹ/ and /l/. *The Journal of the Acoustical Society of America*, *99*(2), 1161–1173. [CrossRef]

Francis, A. L., & Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(2), 349–366. [CrossRef] [PubMed]

Harnsberger, J. D. (2001). On the relationship between identification and discrimination of non-native nasal consonants. *Journal of the Acoustical Society of America*, *110*(1), 489–503. [CrossRef]

Hualde, J. I. (2013). *Los sonidos del español: Spanish language edition*. Cambridge University Press. [CrossRef]

Huang, T., & Johnson, K. (2010). Language specificity in speech perception: Perception of Mandarin tones by native and nonnative listeners. *Phonetica*, *67*, 243–267. [CrossRef]

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson, & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 143–165). Academic Press.

Johnson, K. (2004). Cross-linguistic perceptual differences emerge from the lexicon. In W. W. Agwuele, & P. Sang-Hoon (Eds.), *Proceedings of the 2003 Texas linguistics society conference: Coarticulation in speech production and perception* (pp. 26–41). Cascadilla Press.

Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, *34*, 485–499. [CrossRef]

Johnson, K., & Babel, M. (2010). On the perceptual basis of distinctive features: Evidence from the perception of fricatives by Dutch and English speakers. *Journal of Phonetics*, *38*(1), 127–136. [CrossRef]

Kim, J. Y. (2011). *Discrepancy between the perception and production of stop consonants by Spanish heritage speakers in the United States* [Unpublished MA Thesis, Korea University].

Kim, J. Y. (2016). *The perception and production of prominence in Spanish by heritage speakers and L2 learners* [Doctoral dissertation, University of Illinois at Urbana-Champaign].

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, *50*, 93–107.

Kuhl, P. K. (2000). Language, Speech, and Hearing. *Annual Review of Psychology*, *51*(1), 211–236.

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*, 979–1000. [CrossRef] [PubMed]

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetics perception in infants by 6 months of age. *Science*, *255*, 606–608. [CrossRef] [PubMed]

Levy, E. S. (2009). Language experience and consonantal context effects on perceptual assimilation of French vowels by American–English learners of French. *Journal of the Acoustical Society of America*, *125*(2), 1138–1152. [CrossRef]

Levy, E. S., & Strange, W. (2008). Perception of French vowels by American English adults with and without French language experience. *Journal of Phonetics*, *36*(1), 141–157. [CrossRef]

Lively, S., Pisoni, D. B., Yamada, R. A., Tohkura, Y. I., & Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/: III. Long-term retention of new phonetic categories. *Journal of the Acoustical Society of America*, *96*(4), 2076–2087. [CrossRef]

Mack, M. (1989). Consonant and vowel perception and production: Early English-French bilinguals and English monolinguals. *Perception and Psychophysics*, *46*, 189–200. [CrossRef]

MacKain, K. S., Best, C. T., & Strange, W. (1981). Categorical perception of English/r/and/l/by Japanese bilinguals. *Applied Psycholinguistics*, *2*(4), 369–390. [CrossRef]

Mazzaro, N., Colantoni, L., & Cuza, A. (2016). Age effects and the discrimination of consonantal and vocalic contrasts in heritage and native Spanish. In *Romance linguistics 2013: Selected proceedings of the 43th linguistic symposium on romance languages*. John Benjamins.

Montrul, S. (2008). Incomplete acquisition and attrition of Spanish syntax in adult heritage speakers. *Bilingualism: Language and Cognition*, *11*(3), 201–218.

Montrul, S. (2016). Age of onset of bilingualism effects and availability of input in first language attrition. In E. Nicoladis, & S. Montanari (Eds.), *Bilingualism across the lifespan: Factors moderating language proficiency* (pp. 141–161). American Psychological Association. [CrossRef]

Mora, J. C., & Nadeu, M. (2012). L2 effects on the perception and production of a native vowel contrast in early bilinguals. *International Journal of Bilingualism*, *16*(4), 484–500. [CrossRef]

Morrison, G. S. (2002). Perception of English /i/ and /ɪ/ by Japanese and Spanish listeners: Longitudinal results. In G. S. Morrison, & L. Zsoldos (Eds.), *Proceedings of north west linguistics conference 2002* (pp. 29–48). Simon Fraser University Linguistics Graduate Student Association.

Navarra, J., Sebastián-Gallés, N., & Soto-Faraco, S. (2005). The perception of second language sounds in early bilinguals: New evidence from an implicit measure. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(5), 912. [CrossRef] [PubMed]

Ning, J., Peng, G., Liu, Y., & Li, Y. (2022). The effect of simultaneous exposure on the attention selection and integration of segments and lexical tones by Urdu-Cantonese bilingual speakers. *Frontiers in Psychology*, *13*, 918737. [CrossRef] [PubMed]

Pallier, C., Bosch, L., & Sebastian-Galles, N. (1997). A limit on behavioral plasticity in speech perception. *Cognition*, *64*(3), B9–B17. [CrossRef]

Pisoni, D. B. (1997). Factors affecting the perception of speech. In *Speech perception: Theoretical and experimental issues* (pp. 115–144). Springer.

Polka, L., & Werker, J. F. (1994). Developmental changes in perception of non-native vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 421–435.

Rao, Z., & Myers, J. (2014). Heritage speakers' perception of prosody and vowel duration in Spanish. *International Journal of Bilingualism*, *18*(3), 311–325.

Sebastián-Gallés, N., & Soto-Faraco, S. (1999). Online processing of native and non-native phonemic contrasts in early bilinguals. *Cognition*, *72*, 111–123. [CrossRef]

Strange, W. (Ed.). (1995). *Speech perception and linguistic experience*. York Press.

*languages*

MDPI

*Article*

# Loanword Phonology of Spanish Anglicisms: New Insights from Corpus Data

**Linda Bäumler**

Department of Romance Studies, University of Vienna, 1010 Wien, Austria; linda.baeumler@univie.ac.at

**Abstract:** Previous research shows that several factors influence the adaptation of English phonemes in Spanish Anglicisms: speaker age, English proficiency, and geographic distance from the U.S.A, among others.Due to globalization, increased mobility, and the ubiquitous availability of English media, the question arises whether these factors are still relevant in today's world. For the present study, 70 speakers from Mexico and Spain read a word list containing Anglicisms aloud. A generalized linear mixed effects model was applied to analyze which factors directly influence pronunciation. Results show that the realization of Spanish grapheme-phoneme correspondences plays a major role in the adaptation process. Moreover, the analysis shows that it is exposure to the English language that mainly influences the pronunciation: the more exposure speakers from both countries have to the English language, the more likely they are to imitate the English pronunciation instead of the realization of Spanish grapheme-phoneme correspondences. Finally, the analysis revealed differences not only between the phonemes and the speakers but also between the words included in the study and once more highlighted that every word has a history of its own.

**Keywords:** loanword phonology; adaptation; Anglicisms; Spanish; globalization; language contact

## 1. Introduction

Loanwords that enter a target language often possess phonological, orthographic, or morphological features originating from their source language that may not align with those of the target language. For instance, the open-mid back unrounded vowel /ʌ/ does not exist in the Spanish phoneme system and thus might be difficult to pronounce for native Spanish speakers when encountering English loanwords like *club* [klʌb]. Therefore, speakers often adapt the loanword to fit the system of their native language (*loanword adaptation*, (Chang 2008, 2009; Paradis and Tremblay 2009; Peperkamp and Dupoux 2003)). In loanword adaptation, there are two main options: speakers either substitute the foreign sound with a similar native one or they attempt to imitate it (Calabrese 2009; Gómez Capuz 2001; Kang 2011; Pratt 1980; Pustka 2021; Van Coetsem 1988). For example, if a Spanish speaker encounters the Anglicism *brunch* via the oral channel, he or she might either modify the unfamiliar vowel /ʌ/ to the Spanish vowel [a] or import the foreign sound.

Peperkamp et al. (2008, p. 159) suggest that phoneme importations occur when a "non-native structure is not perceptually assimilated", which means that a Spanish speaker only imports the foreign vowel [ʌ] when she or he perceives the difference between [ʌ] and the native vowel [a]. If, due to the influence of their first language (L1 filter) they map the foreign sound to a similar sound in their native phoneme system, they produce the corresponding native sound and do not import the foreign phoneme. In this context, many researchers argue that loanword adaptation occurs during the perception of foreign input and that speakers may adapt the foreign phoneme to the acoustically closest sound of their own L1 (Calabrese 2009; Peperkamp et al. 2008; Vendelin and Peperkamp 2004), for another view, see (LaCharité and Paradis 2005). Drawing from Best (1994), Peperkamp et al. (2008) underline that some non-native sounds are easier to perceive than others, making non-adaptation, i.e., importation more likely. In this context, some authors also emphasize the role of salience and similarity in loanword phonology (Kang 2003; Kenstowicz 2007). A well-known example is the differentiation between English /æ/ and /ɛ/ (e.g., in English *bat* and *bet*), which is difficult for German learners (Llompart and Reinisch 2019): they might, therefore, not import the foreign vowel /æ/ when using a loanword, but adapt it to the similar sound of their L1. For Spanish in the United States, Flege and Wayland (2019) conducted a perceptual mapping experiment among Spanish natives coming from different Spanish-speaking countries living in the U.S., where they had arrived between the ages of 17 and 46. Participants had to map the different English vowels to the Spanish vowels /i e a o u/ or could indicate the option "not Spanish". Results show that for instance for the English vowel /æ/ less than 20% of the participants claim that it is not a Spanish vowel, for the vowel /ʌ/, less than only 30% and for the vowel /ɛ/, less than 40% do so. Only the vowel /ɝ/ was identified as not Spanish in 70 to 80% of cases.

However, Peperkamp et al. (2008, p. 160) note that loanword adaptations are not entirely equivalent to perceptual assimilation, as perceptual assimilation displays intra- and inter-speaker variation, whereas loanword adaptations tend to be homogeneous within speech communities. This suggests that loanword adaptations undergo a standardization process, although the outcome may change over time, where certain non-native structures are not adapted in younger loanwords compared to older ones. Using the example of *brunch*, some speakers may initially import the foreign sound [ʌ] if they perceive the difference, but over time, as more speakers adopt it, they might also adapt it to the native sound [a].

In previous research, different factors have been presented that might influence the adaptation process. Independently of the individual speakers, orthography and the borrowing channel might influence the adaptation process. The role of orthography has been debated: while some researchers consider it unimportant in loanword adaptation (LaCharité and Paradis 2005), others emphasize its significance (Flege and Wayland 2019; Meisenburg 1992; Peperkamp and Dupoux 2003; Peperkamp et al. 2008; Pratt 1980; Pustka 2021; Vendelin and Peperkamp 2006). Orthography may influence the pronunciation of foreign phonemes insofar as speakers align their pronunciation of the foreign sound with the grapheme-phoneme correspondences of their native language. Even when loanwords enter the recipient language via the auditory channel (*ear loans*, as opposed to *eye loans* which are loans that are borrowed via the written channel, (Pratt 1980)), speakers may nevertheless be influenced by orthography if they are proficient in the source language's writing system (Vendelin and Peperkamp 2006, p. 1004). Other than English, which is known for its opaque orthography, "Spanish is said to have a transparent orthography" (Kwok et al. 2017, p. 2105), since "most letters equal one sound and most single sounds equal one letter" (Lallier et al. 2014, p. 1178). Therefore, through grapheme-phoneme conversion, most Spanish words can be read in a non-lexical way (Ferreres and López 2014, p. 522)—although its orthography is not completely transparent (Bravo-Valdivieso and Escobar 2014, p. 443). However, given the shallow orthography of Spanish (Seymour et al. 2003, p. 145), this may lead to the realization of Spanish grapheme-phoneme correspondences in Anglicisms, in particular.

It has to be noted that it is often impossible to determine the original path of loanwords, especially nowadays, given the extensive array of auditory and visual platforms through which Anglicisms can find their way into the Spanish language. An Anglicism may enter the Spanish language through different kinds of auditory channels such as movies or podcasts, and at the same time through visual channels such as commercials, blogs, or Instagram posts. If we think of examples like Anglicisms written in TV adverts, they show that Anglicisms can also enter through both channels—auditory and visual—simultaneously (Winter-Froemel 2011, p. 276). Moreover, an Anglicism can also initially enter the target language as an eye loan and later also as an ear loan, or vice versa. This scenario can lead to a shift wherein a pronunciation, originally grounded in Spanish grapheme-phoneme correspondences, is replaced by the perception-based realization. The path of entrance is also related to the diffusion process of the Anglicisms. Starting from different innovators and early adopters (Milroy and Milroy 1985; Rogers and Shoemaker 1971) who might have come into contact with the Anglicisms via different channels, the Anglicisms take different pathways and thus their pronunciation might vary.

Moreover, the perceived prestige of the source language might also play a decisive role if speakers imitate the sounds of this language (Heffernan 2007; Meisenburg 1992; Pustka 2021). Concerning lexeme-specific factors, researchers have attributed phonetic variability to the time the loanword was integrated into the recipient language, assuming that early, i.e., older loanwords are integrated to a higher degree to the recipient language than recent, i.e., younger loanwords (Haspelmath 2009; Poplack et al. 1988). Rodríguez González (2017) for instance names *flirt* realized as [fliɾt] and *girl* realized as [guel] as examples: *flirt* being a much older and *girl* a more recent loanword. In the case of Spanish, older Anglicisms are thought to be pronounced in accordance with Spanish grapheme-phoneme correspondences, whereas more recent loanwords are more likely to be pronounced following the English model (Gómez Capuz 2001; Rodríguez González 2017). In addition, the frequency of use of a loanword is also believed to influence the adaptation process. Loanwords that are frequently employed by the speech community tend to undergo greater adaptation to the recipient language compared to loanwords which are only rarely used (Poplack et al. 1988; Rodríguez González 2017).

Furthermore, geographic proximity to the U.S. is also said to influence the adaptation of English loanwords. Previous research assumes a higher influence of English in Hispanoamerica than in Spain (Oncins-Martínez 2009; Pratt 1980; Pustka 2021). Gimeno Menéndez and Gimeno Menéndez (2003, p. 301), who analyzed the use of Anglicisms in Spanish newspaper articles from Spain and the U.S., discovered that the U.S. press uses more Anglicisms than the Spanish press. Additionally, the *Proyecto de estudio coordinado de la norma lingüística culta de las principales ciudades de Iberoamérica* (Lope Blanch 1986), which involved speakers from various Spanish-speaking countries completing a lexical questionnaire (Comisión de Lingüística Iberoamericana 1971), indicated that participants from Madrid reported fewer Anglicisms compared to those from Latin American capitals such as Mexico City, Santiago de Chile, and San Juan (Dworkin 2012; Moreno de Alba 1992). The economic colonization of vast regions of Hispanoamerica by the U.S. likely plays a significant role in this phenomenon (Pratt 1980). Baumgardner (2006) and Despagne (2010) emphasize the close economic and political ties between the U.S. and Mexico, which have heightened the importance of English in Mexico in education, entertainment, tourism and economy and have led to the incorporation of English elements in Mexican commerce. Nevertheless, Baumgardner (2006, p. 263) also underlines the importance of English as a world language nowadays and outlines that Mexico—as well as many other countries—is part of the Expanding Circle (Kachru 1985), whose cultures and languages have been influenced by English. He claims that "proximity is no longer of prime importance" (Baumgardner 2006, p. 263).

Besides these factors, a variety of speaker-specific factors have also been suggested to influence the adaptation process in previous literature: The individual's attitude towards the language can be linked with the generally perceived prestige of the language in a speech community. This personal language attitude can also influence the adaptation process (Haspelmath 2009, p. 42). If the speaker holds a favorable attitude towards the source language, they may exhibit a tendency to adapt the Anglicism less to the recipient language and adhere more closely to the origin. However, it has to be noted that language attitudes are not easy to measure objectively because, first, speakers might not be aware of certain attitudes and second, social desirability might play a role when speakers are asked directly about their attitudes. Different methods, such as the matched guise method (Lambert et al. 1960) have been developed to measure these language attitudes.

The level of competence in the foreign language is also considered a relevant factor in previous literature: more competent speakers might adapt less than speakers who are not so competent in the source language (Gómez Capuz 2001; Haspelmath 2009; Haugen 1950; Poplack et al. 1988; Rodríguez González 2017). Furthermore, previous research also assumes that the age of the speaker influences the way Anglicisms are realized, with younger speakers being more inclined to imitate the English model compared to older speakers (Gómez Capuz 2001; Rodríguez González 2018). This assumption is likely tied to other external variables such as English proficiency, exposure to the English language, and attitudes towards it, since younger speakers may have higher levels of English proficiency, as English language instruction in schools is becoming more prevalent. Additionally, they may have more frequent exposure to English through modern media, their work, and travel experiences, which could further enhance their language skills. Moreover, younger speakers might also exhibit more positive attitudes towards the English language.

These possible interactions should be taken into account in an analysis that includes several different words and speakers, to allow a systematic comparison of the adaptation of English phonemes between different speakers and Anglicisms. Statistical models should be used to determine whether a factor has a significant effect on the pronunciation of foreign phonemes and to reveal whether this effect is indeed a direct one or an indirect one through another factor, thus acting as a confounding variable. To my knowledge, there is currently no large-scale corpus-phonological dataset available that would allow such an analysis. Since Anglicisms are rarely found in spontaneous speech, this study aims to fill this research gap by conducting an analysis of the pronunciation of Anglicisms on corpus-phonological data involving elicited data, namely the reading of a word list. With this method, I aim to analyze first, whether foreign phonemes get imported and second, which variables in a globalized world influence the realization of English phonemes in Spanish Anglicisms. In the following sections, I will begin by describing the methodology used before presenting the results and discussion.

## 2. Materials and Methods

In the following section, I will describe the methodology applied in this study, namely the corpus which served as the database, as well as the conducted analysis (audiovisual coding, calculation of inter-rater agreement and statistical analysis).

### *2.1. Corpus*

#### 2.1.1. Participants

The English Influence on Spanish (EIS) corpus was used as the database for the present study. The corpus was designed for the dissertation project of the author (Bäumler 2023) and includes 70 native Spanish speakers from Mexico and Spain. Speakers were recruited in the respective capitals of the two countries—Mexico City and Madrid—and in rural areas surrounding these cities: speakers from Pedrezuela (5892 residents (INE 2019)) in Spain as well as Chiconcuac (19,656 residents (INEGI 2009)) in Mexico were also included in the corpus. I chose these two areas since they seem relatively comparable regarding their infrastructure and social composition. Moreover, they are both equally distanced from the respective capital. Since this distance is rather short (45 min by car) one can assume that speakers of the rural areas use the same variety as the speakers of the respective capitals. I recorded 50 speakers in the capitals and 20 speakers in the rural areas. For the recruitment, I used the 'friend of a friend' approach (Milroy and Gordon 2003, p. 32), which assured speakers' confidence in me as an investigator, since I got to know the participants as a friend of a friend and not as a complete stranger or outsider (Milroy and Gordon 2003, p. 32). Moreover, this approach not only created a trustworthy and comfortable environment during the recordings but also an economic way of efficiently finding new participants (Pustka et al. 2018).

#### 2.1.2. Tasks

The EIS-corpus includes naturalistic as well as experimental data, as is often the case in corpus phonology (Chaudron 2007; Eychenne 2021) and was created following the research protocol established in previous projects like *Phonologie du Français Contemporain* (Durand et al. 2002; Eychenne and Laks 2012), *Phonologie de l'Anglais Contemporain* (Durand and

Przewozny 2012), and *Fonología del Español Contemporáneo* (Pustka et al. 2016, 2018). This is particularly relevant in the case of loanword phonology since Anglicisms are rarely present in spontaneous speech and are, therefore, only marginally comparable between speakers (cf. Section 1). Besides various other tasks, speakers read a word list consisting of 102 items that contained 77 Anglicisms and 25 filler items. These Anglicisms also included 29 Anglicisms which are morphologically and/or orthographically integrated into the Spanish language or semantic loans (e.g., *administración*) and can, therefore, also be regarded as fillers, since they distract the participants from the Anglicisms in the list. By having participants read this list, it is possible to analyze the realization of various phonemes. Unlike spontaneous speech, this approach facilitated a systematic comparison of pronunciation across different speakers and across different words (Pustka et al. 2018, p. 2). However, it is important to note that, with this chosen method, speakers are also exposed to the orthography, potentially influencing the pronunciation of loanwords. Therefore, when interpreting the data, one must consider the potential impact of orthography and list effects. Nevertheless, these limitations were deemed acceptable in light of the advantages gained through systematic comparison. Table 1 gives an overview of the words included in the list and the respective English phonemes they contain, which served as the basis for the present study.

**Table 1.** Foreign English phonemes and the respective Anglicisms integrated in the word list.

| Category | English Phoneme | Anglicisms Integrated in Word List |
|---|---|---|
| Consonants | <r>[1] word-finally | *baby sitter, chárter, contáiner, gangster, cheer leader, leader, manager, trailer* |
| | <r> before a consonant | *interview, marketing, flirt, airbag, derby, overbooking, chárter, jersey, cheer leader, iceberg, kindergarden, parking, standard* |
| | <r> before a vowel | *trailer, brunch, copyright, ginger-ale, Rock'n'Roll* |
| | /h/ | *hobby, high-society, hot-dog* |
| | /v/ | *festival, interview, detecive, overbooking* |
| | /ŋ / | *marketing, overbooking, gangster, meeting, parking* |
| | /d͡ʒ / | *gentleman, ginger-ale, jersey, get-lag, manager, jazz* |
| Vowels | /ə/ | *contáiner, alien, gentleman, festival, slogan, high-society* |
| | /ʌ/ | *club, brunch, bluff, puzzle* |
| | /æ/ | *match, airbag, gangster, sándwich, fan, WhatsApp, flashback, jet-lag, manager, standard, jazz* |
| | /ɑ/ (GA) /ɒ/ (RP) | *hobby, copyright, chárter, hot-dog, kindergarden, parking, Rock'n'Roll* |
| | /ɜ/ | *flirt, derby, jersey, iceberg* |
| Diphthong | /oʊ/ (GA), /əʊ/ (RP) | *overbooking, slogan, Rock'n'Roll* |

Note that General American English as well as Received Pronunciation have been taken into account. These two varieties have been chosen since they are spoken by a large number of speakers, and therefore, it is likely that Spanish natives encounter them via media, while traveling or in other international settings. Nevertheless, one can assume that General American is more present in the media: since the Second World War, when the United States of America claimed a leading global position which led to English becoming the lingua franca, the influence of American English on other languages has been increasing (Kowner and Rosenhouse 2008; Schweickard 1991). Moreover, American cultural imperialism (Gray 2007; Hamm and Smandych 2005) led to an increasing influence of American English through the worldwide diffusion of culture, trademarks, values, media and language of the United States.

The word list recordings were made using the digital audio recorder ZOOM H4n Handy at a sampling rate of 44.1 kHz and a 16-bit depth. The condenser microphone AKG C 520 was used to further improve audio quality.

#### 2.1.3. Variables

As the aim of this study was, besides the description of the realization of the different phonemes, to reveal the factors that influence a more 'Spanish' or a more 'English' way of realization (see Section 2.1.4), multiple variables were collected. In the following section, I first describe the external variables that relate to the speakers before describing the internal variables that relate to the words included in the study.

Table 2 gives an overview of the external variables that relate to the speakers.

**Table 2.** Variables collected in the sociodemographic interview.

| Variable | Value | Summary |
|---|---|---|
| *Country* | categorical: Mexico/Spain | Spain 34 (48.6%), Mexico 36 (51.4%) |
| *Area* | categorical: city/rural | city 50 (71.4%), rural 20 (28.6%) |
| *Age* | numerical (in years) | mean 48.06 (SD 19.17) |
| *Sex* | categorical: female/male | female 35 (50%), male 35 (50%) |
| *English level* | ordinal: zero/beginner/intermediate/advanced | zero 31 (44.3%), beginner 11 (16.0%), intermediate 13 (18.6%), advanced 14 (20%) |
| *Education level* | ordinal: elementary/sec. I/sec. II/university/doctorate | elementary 1 (0.1%), sec. I 15 (21.4%), sec. II 16 (22.6%), university 36 (51.4%), doctorate 2 (2.9%) |
| *Time spent abroad* | numerical (in months) | mean 10.66 (SD 23.88) |
| *GS* | numerical | mean 19.7 |
| *LEAS* | numerical | mean 11.99 (SD 6.91) |

I ensured an equal distribution between females and males as well as between three different age groups (18 to 35 years, 36 to 65 years, and older than 65 years) in each survey point. The mean age of the speakers was 48 years, while the youngest speaker was 21 years and the oldest 92 years old. Besides the classic sociolinguistic variables, I included a variety of variables that potentially influence the realization of Anglicisms in the analysis. I also aimed for a balanced sample in terms of knowledge of the English language: 44.3% of the speakers claimed having no knowledge of the English language and 55.7% reported proficiency in English. Among the latter, 16.0% indicated having basic, 18.6% intermediate and 20.0% advanced knowledge. Moreover, two variables, built-in scores, were designed especially for the EIS-Corpus, which require further explanation:

First, a score called *Language Exposure and Affinity Score* (LEAS) was created in order to describe the participants' affiliation with the English language and American culture. It consolidates various variables into a single score: Data for this score came from a sociodemographic interview, in which participants responded to different questions regarding their English media use or travel habits. Participants reported their frequency of reading in English, watching English movies, using the internet in English, and having English language interactions in their daily communication. Moreover, they rated their attitude towards the English language and American culture and indicated how many times they had visited an English-speaking country. The first six questions answered on a 5-point Likert scale, contributed four points if the most positive value was chosen and zero points for the most negative value. For the non-Likert scale variable (number of visits to English-speaking countries), zero points were added for those who had not visited, two points for one visit, and four points for two or more visits, allowing for a maximum of 28 points. The analysis of LEAS revealed a mean score of 11.99 points, with a median of 13 points.

Moreover, a *Globalization Score* (GS) was established in order to assess the perspectives of speakers regarding globalization processes and their open-mindedness towards foreign cultures. It was derived from the sociodemographic interview by amalgamating the answers to ten different questions. These questions were adopted from or inspired by various surveys conducted by relevant opinion research institutes, namely the World Values Survey (Inglehart et al. 2014), European Values Study (EVS 2010), and a study conducted by the University Diego Portales of Chile addressing cosmopolitanism (Gayo and Teitelboim 2008). The questions explored aspects such as speakers' concerns about their home country's culture being influenced by the U.S., their participation in religious or familial traditions, the importance they place on their children engaging with their home country's culture and traditions, and their views on dependency on scientific and technological advancements. Respondents referred to a 4-point Likert scale for their answers. The forced-choice format was chosen to obtain distinct tendencies in participants' attitudes. Points were assigned accordingly: three for the most negative response, two for 'not that much', one for 'a little bit', and zero for 'very much'. Speakers were also asked to rate, on a 4-point Likert scale, the significance they attributed to their children getting to know other countries, understanding other cultures and learning foreign languages. They were also asked if they thought that the opportunity to easily connect with people from other countries leads to a better world and whether they thought that future generations would have more opportunities because of science and technology. In these cases, zero points were awarded for the most negative response, one for 'not that much', two for 'a little bit', and three for 'very much'. Additionally, speakers indicated whether they considered themselves citizens of the world, of Europe/America, of Spain/Mexico, or of their local community. Points were assigned as follows: four for identifying as citizens of the world, two for identifying with their continent, and zero for identifying with their country or local community. This led to a maximum score of 31 points. Notably, only two speakers identified with their local community while also feeling a connection to their entire country. Analysis of GS showed that the mean score of all participants was 19.8 points, with a median of 20, a maximum of 29 and a minimum of 13 points. Importantly, no speaker scored very low (below 13 points), which could be attributed to social desirability: speakers may have hesitated to express negative views, especially when addressing their children's exposure to other cultures.

In addition to the sociodemographic metadata on the speakers, I also gathered metadata for the Anglicisms examined in this study. To assess whether word frequency impacts the adaptation of foreign phonemes, I extracted frequency values from the online corpora *American Spanish Web 2011* (esamTenTen11; (Kilgarriff and Renau 2013)) and *European Spanish Web 2011* (eseuTenTen11; (Kilgarriff and Renau 2013)), both accessible via the Sketch Engine software (Kilgarriff et al. 2014). By including data from both corpora, I aimed to address potential frequency differences between the two continents. Consequently, I assigned the frequency values of the American corpus to speakers from Mexico, while values from the European corpus were aligned to speakers from Spain.

Moreover, to examine the impact of the age of the loanwords, I determined the age of the Anglicisms based on an analysis of the *Nuevo Tesoro lexicográfico de la lengua Española* (NTLLE) (Real Academia Española 2019), which consolidates a comprehensive selection of different Spanish dictionaries. For this study, I extracted the date of the first appearance of each word and the respective dictionary in which it first appeared. Note that using the year of incorporation into a Spanish dictionary as an indicator of the time when the loanword is known in the speech community comes with inevitable methodological challenges. Loans or neologisms are typically known in the speech community for a certain, often indeterminable amount of time before their inclusion in dictionaries. Therefore, the year of first appearance should be viewed as an approximation.

#### 2.1.4. Analysis

In order to analyze the realization of the Anglicisms, I processed the data in PRAAT (Boersma and Weenink 2021) before segmenting and orthographically transcribing. Following this, three linguists who have undergone phonetic training—namely, the author of the article, a master's degree student, and another person holding a master's degree in linguistics—phonetically transcribed the Anglicisms. Note that 82% of the data were transcribed by all three linguists, while 100% of the data were transcribed by the first two linguists. The first two of the transcribers have German as their first language and are proficient in English and Spanish as foreign languages, while the third transcriber is a native English speaker. All three have studied linguistics in Spanish and/or English, demonstrating extensive knowledge in either Spanish or English linguistics, or both. They independently transcribed the data to minimize mutual influence during the process, and subsequent comparisons were made to ensure objectivity. During this process, each coder performed audiovisual transcription, aligning the phonetic transcription with the acoustic signal.

To measure the degree to which the transcribers agreed in their transcriptions, I calculated interrater agreement, applying Fleiss' Kappa (Fleiss et al. 2003) for the cases with three transcriptions available and Cohens' Kappa (Cohen 1960) for the cases with two available transcriptions. The R package DescTools (Signorell 2024) was employed for the calculation. Landis and Koch (1977) proposed a classification for interpreting Kappa values, as outlined in Table 3.

**Table 3.** Interpretation of kappa value (Landis and Koch 1977, p. 165).

| Kappa Statistics | Interpretation |
|---|---|
| <0.00 | Poor agreement |
| 0.00–0.20 | Slight agreement |
| 0.21–0.40 | Fair agreement |
| 0.41–0.60 | Moderate agreement |
| 0.61–0.80 | Substantial agreement |
| 0.81–1.00 | Almost perfect agreement |

It is worth noting that Landis and Koch (1977) underlined the arbitrary nature of these divisions, emphasizing their role as useful benchmarks. Critics have pointed out that the effects of prevalence and bias on kappa must be considered when interpreting its magnitude (Sim and Wright 2005, p. 264). It is essential to acknowledge that kappa can be influenced by the number of categories in the measurement scale: the more categories, the potential for disagreement among raters is increased, leading to a lower kappa with many categories compared to fewer categories. Recognizing these limitations, the interpretation provided by Landis and Koch (1977) can only serve as an orientation in this study. Given the diverse potential categories in phonetic transcription, it can be assumed that calculated kappa values in phonetic studies are artificially low.

For the following statistical analysis, I chose the transcriptions in which all two or three transcribers matched. For cases where all transcribers provided differing results, the author re-listened to the phoneme and made a final decision on one solution.

Subsequently, I processed the data in R (R Core Team 2021) and analyzed them in two steps: (1), a descriptive analysis of the transcribed data was conducted to gain insights into the phoneme realizations and their corresponding frequencies. (2), statistical modeling was applied to understand the impact of variables on the realization of the specific phoneme. For this purpose, the R package *lme4* (Bates et al. 2015) was employed for fitting linear mixed-effects models. The transcribed data were dichotomously coded for statistical modeling: when speakers realized the foreign phoneme in accordance with Spanish grapheme-phoneme correspondences, I coded such realizations as 'Spanish'. Conversely, when speakers imitated the English model, these realizations were coded as 'English'. For instance, for the vowel /ʌ/, data were coded as 'English' when the English phoneme was imported (e.g., [bɹʌnt͡ʃ] for *brunch*) or when speakers adapted the foreign phoneme without adhering to Spanish grapheme-phoneme correspondences (e.g., [bɹant͡ʃ]), assuming that they intended to imitate the English language. Realizations aligning with Spanish grapheme-phoneme correspondences were coded as 'Spanish' (e.g., [bɾunt͡ʃ]). Figure 1 provides a visual representation of the dichotomous coding process.

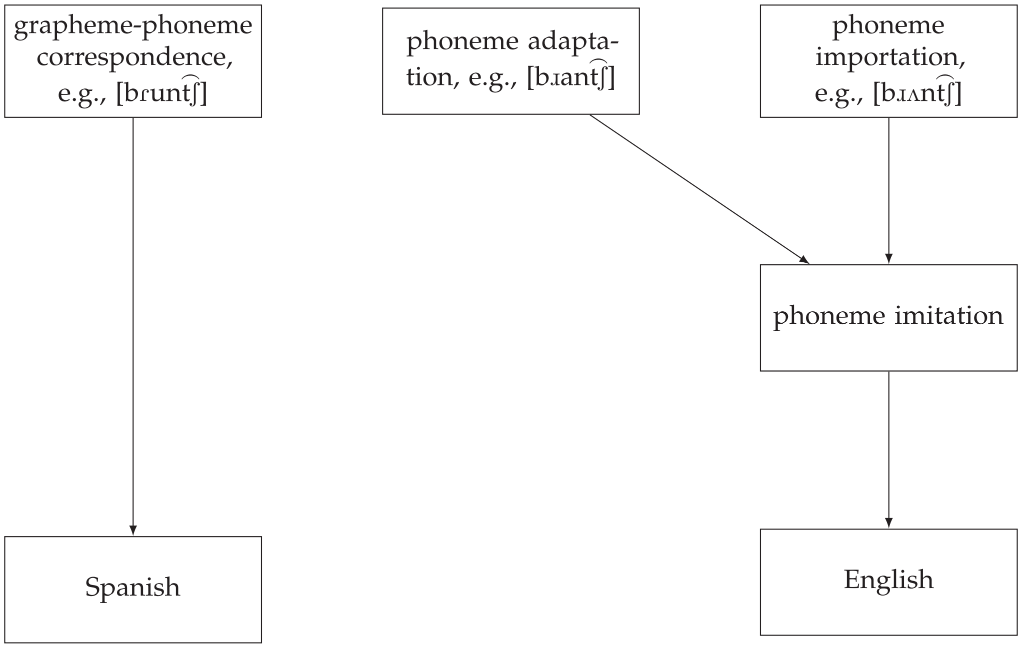



**Figure 1.** Dichotomous coding of the variable.

To identify those variables that are actually (i.e., in terms of statistical significance) and directly (i.e., not confounding by another independent variable) influencing the pronunciation of the foreign phoneme, a generalized linear mixed-effects model was constructed using both bottom-up and top-down approaches. For enhanced precision in the results, I log-transformed the variable *word frequency* and scaled all continuous variables. Reference levels were set as follows: city (*area*), Spain (*country of origin*), female (*sex*). Moreover, the random effects *speaker*, *word* and *phoneme* were included to account for group-level variations.

## 3. Results

### *3.1. Interrater Agreement*

The computation of the Kappa values yielded a mean score of 0.6, indicating "moderate agreement" according to (Landis and Koch 1977). Note that it was highest for the consonants and lowest for the vowels (0.68 and 0.5), where obviously inconsistencies between the phonetic transcriptions appear more often. Given the inherently diverse nature of phonetic transcriptions with numerous categories, this mean Kappa value signifies a satisfactory level of interrater agreement.

### *3.2. Distribution of the Realizations of the Individual Phonemes*

The descriptive analysis of the data showed that some English phonemes are more likely to be imported by the Spanish speakers than others. Table 4 depicts these results.

The results not only show that consonants are more frequently imported than vowels but also that the realization of Spanish grapheme-phoneme correspondences plays a major role in the adaptation process. Overall, the English phonemes are realized according to Spanish grapheme-phoneme correspondences in 73.1% of the cases.

**Table 4.** Rates of phoneme importation and realization of grapheme-phoneme correspondences (GPC).

| Phoneme | Importation (in %) | Realization of GPC (in %) | Error (in %) |
|---|---|---|---|
| <r> word-finally | 15.9 | 78.2 | 6.2 |
| <r> before a consonant | 19.2 | 70.9 | 6.3 |
| <r> before a vowel | 25.2 | 66.6 | 6.6 |
| /h/ | 47.4 | 4.2 | 7.0 |
| /v/ | 10.2 | 84.9 | 4.6 |
| /ŋ / | 48.1 | 41.7 | 7.0 |
| /d͡ʒ / | 40.4 | 18.1 | 2.6 |
| /ə/ | 5.6 | 82.7 | 7.1 |
| /ʌ/ | 13.0 | 65.5 | 5.6 |
| /æ/ | 7.0 | 81.6 | 6.5 |
| /ɑ/ (GA) /ɒ/ (RP) | 2.1 | 90.4 | 6.7 |
| /ɜ/ | 18.3 | 75.7 | 5.3 |
| /oʊ/ (GA), /əʊ/ (RP) | 18.7 | 71.8 | 8.0 |

### *3.3. Variables Influencing the Realization of the Phonemes*

Remember that for the calculation of the influence of different variables, data were coded dichotomously. The analysis revealed that a significant number of speakers (31 speakers) exhibited an imitation rate below 20%. Notably, the speakers who showed the highest rates of phoneme imitation (in over 50% of cases), were all from Mexico (9 speakers)—except for one speaker. In contrast, the three speakers who showed an imitation rate of less than 10% were all from Spain.

In order to analyze which variables actually and directly influence the realization of the phonemes, a generalized mixed model (cf. Section 2.1.4) was applied, with the type of realization serving as the dichotomous dependent variable. The model incorporated 5546 observations, with *speaker*, *word*, and *phoneme* set as random effects. For each participant, there were 52 to 87 (mean 79, median 83) observations included. The results are presented in the forest plot in Figure 2.

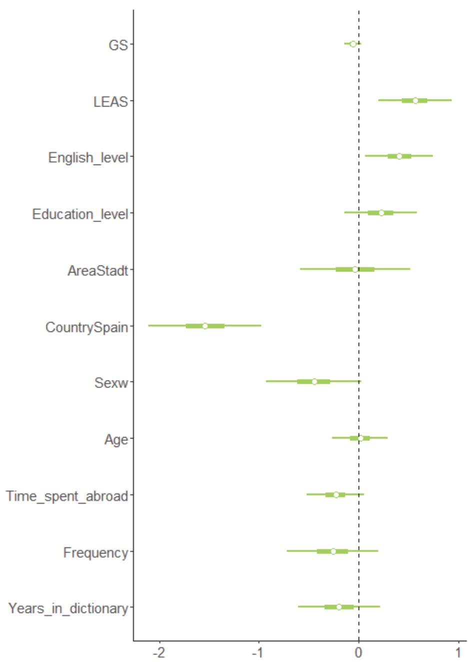



**Figure 2.** Results of the generalized linear mixed effects model for all phonemes.

In this plot, each independent variable's impact on the realization is shown by a horizontal bar, illustrating the 95 percent (thin bar) and 50 percent (thicker bar) credibility intervals of the predicted effect. The x-axis represents the probability of the type of loanword realization: positive values indicate a higher likelihood of imitating the English phoneme, while negative values indicate a higher likelihood of following grapheme-phoneme correspondence. The mean effect of each predictor is marked with a dot. When interpreting these plots, it is crucial to note that if the 95 percent credibility interval overlaps the zero-line (dashed vertical line), the model does not provide a clear direction of the effect for that predictor, indicating it is statistically insignificant.

The model highlighted that the variable *country* significantly influences phoneme realization: Spaniards are more likely to realize Spanish grapheme-phoneme correspondences than Mexicans, indicating that Mexicans are more likely to imitate the English model than Spaniards. Additionally, the model revealed a significant influence of LEAS on phoneme realization; the higher a speaker's LEAS, the more likely they are to imitate the English model. While the forest plot also indicates a significant influence of the variable *English level*, this effect is likely an artifact of the regression model, as it diminishes when it is included with LEAS in the model. It can be assumed that the English level only indirectly impacts phoneme realization, with speakers possessing a higher LEAS potentially also having a higher English level. This indirect influence can be assumed for *age* and *education level* as well, both of which showed significant effects in the univariate analysis. Conversely, the variables GS, *frequency*, *years in the dictionary*, *time spent abroad*, *area*, and *sex* did not exhibit a significant influence on phoneme realization. The analysis of the random effect *phoneme* corroborated descriptive findings, indicating that the phonemes /h/, /d͡ʒ /, /ŋ/, and /ʌ/ were more likely to be imitated than the phonemes /v/, /ɹ/ in word-final position, /æ/, and /ɑ/. These results are visualized in Figure 3.

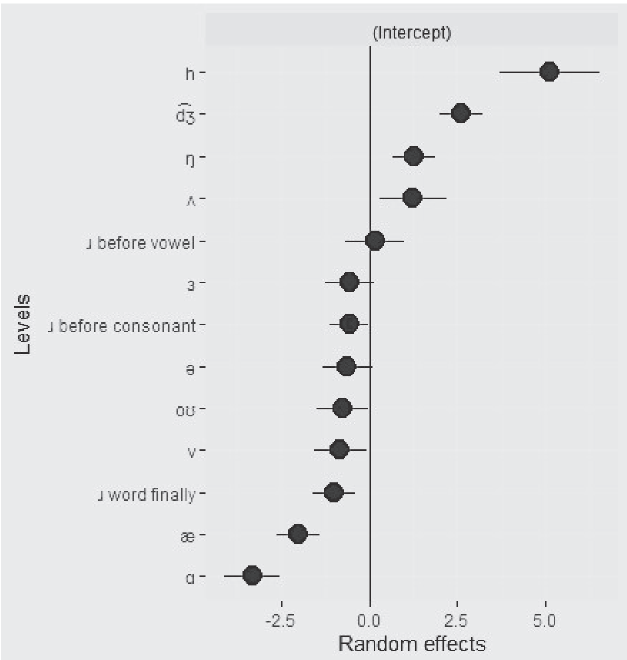



**Figure 3.** Forest plot of group-level intercepts per phoneme.

The difference between Mexicans and Spaniards is depicted by the bar plot in Figure 4, indicating that, overall, Mexicans imitated the English model more frequently (33.5%) than Spaniards (20.1%). Regarding LEAS, a descriptive analysis indicated that speakers who imitated the English model had a higher LEAS (mean 14.18, SD 6.49) compared to speakers who followed Spanish grapheme-phoneme correspondences (mean 11.77, SD 6.85), as shown in Figure 5.

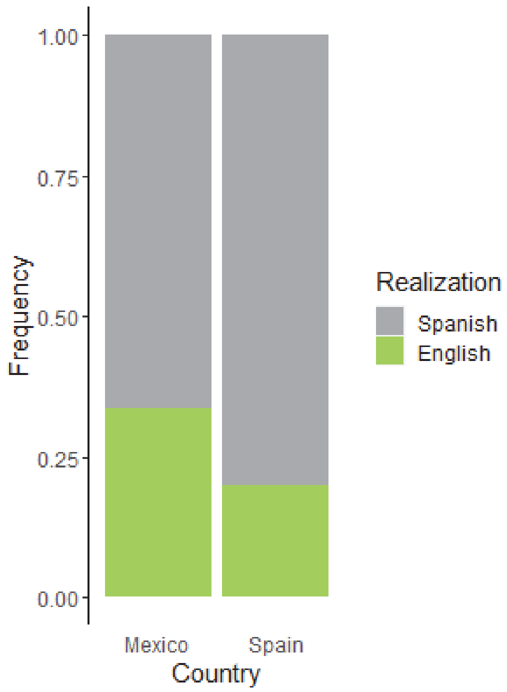



**Figure 4.** Distribution of *country* by realization across all phonemes.

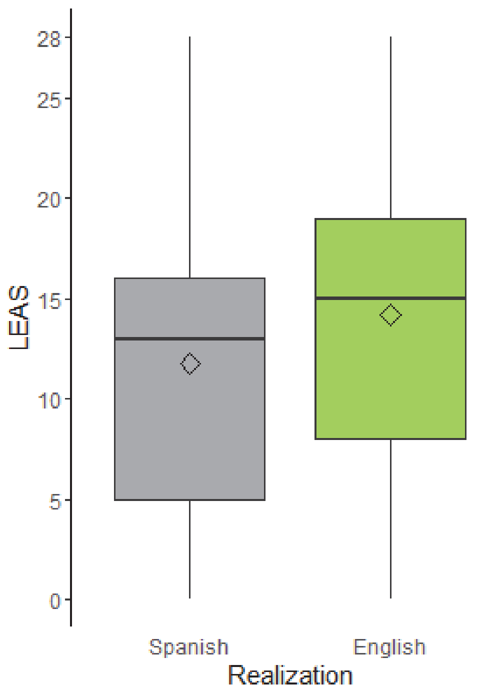



**Figure 5.** Distribution of LEAS by type of realization across all phonemes.

## 4. Discussion

### *4.1. Realization of Grapheme-Phoneme Correspondence, Importation and Adaptation*

The analysis of the data revealed many cases of grapheme-phoneme correspondence realizations (in 73.1% of the cases across all phonemes). A plausible explanation for this might be due to the characteristics of Spanish orthography. The importance of orthography in loanword phonology has been underlined by many studies (see Section 1) and is supported by the present data: given the transparent nature of Spanish orthography, this transparency may influence the realization of Spanish grapheme-phoneme correspondences in loanword adaptation. Spanish speakers may naturally tend towards realizing

Spanish grapheme-phoneme correspondences in Anglicisms since they are accustomed to a transparent orthographic system in their L1. However, it is important to note that the chosen method (reading of a word list) may also have impacted the results since speakers were exposed to orthography which potentially promotes the realization of Spanish grapheme-phoneme correspondences.

Regarding the rates of importation, it has been shown that the English phonemes are imported to varying degrees: overall, vowels showed lower rates of phoneme importation (10.8% across all vowels) than consonants (29.4% across all consonants). This suggests that vowels may be more frequently adapted to align with existing vowels in the Spanish vowel system. Since previous research emphasizes the role of salience and similarity in loanword phonology (Kang 2003; Kenstowicz 2007), it can be assumed that consonants such as /d͡ʒ/, /h/, or /ŋ/ may possess salient features that Spanish speakers aim to import. The differences between the English and Spanish vowels, in contrast, might not be as noticeable for the speakers. Previous research shows that the difference between [æ] and [ɛ] is difficult for English learners, which appears to be true as well for the Spanish natives when realizing English loanwords. In the realization of Anglicisms, speakers may, therefore, adapt the English vowel to the perceptually closest sound in their native system (Brannen 2002). This means that if the non-native sound is easily discernible, non-adaptation (i.e., phoneme importation) is more likely to occur (see Section 1). This aligns with the perspective of many researchers, who underline that loanword adaptation takes place during the perception of foreign phonemes and that speakers consequently might adapt the foreign segment to the acoustically closest sound of their L1 (see Section 1). However, one must consider that the transcription process might also have contributed to the observed differences between consonants and vowels since it is not possible to rule out that the transcribers are influenced by their L1. This limitation, however, only concerns the differentiation between phoneme importation and imitation (e.g., *fan* realized with [æ] or [ɛ]) and not the differentiation between phoneme imitation and realization of Spanish grapheme-phoneme correspondence (e.g., *fan* realized with [æ] or [a]).[2] The latter differentiation was the basis for the statistical model, the results of which are discussed below.

### *4.2. Variables Influencing the Realization of the Anglicisms*

The results of the generalized linear mixed model showed that LEAS influences the realization of the phonemes significantly: speakers with a higher LEAS, meaning that they are more frequently in contact with the English language and showed themselves more open towards it, were more likely to imitate the foreign phoneme. One can assume that speakers who show a higher LEAS might be better trained in the pronunciation of the English language since they are in frequent contact with it. Moreover, since it is these speakers who show a more positive attitude towards the language, they might also be more willing to imitate it than speakers with a less positive attitude. Speakers' level of English, which was often thought to influence the realization of Anglicisms (Gómez Capuz 2001; Rodríguez González 2017), interacted with LEAS in the model. A bidirectional causal relationship between the two variables can, therefore, be assumed: speakers who are frequently in contact with the English language are also more likely to have a higher language competence due to more frequent use of the language and vice versa. The present data, therefore, confirm previous research (Gómez Capuz 2001; Rodríguez González 2017), which suggests that speakers with a higher level of competence in English are more likely to imitate the English model. However, the implementation of LEAS suggests that speakers' degree of exposure to a foreign language is more important than speakers' English level.

Moreover, speakers' education level did not impact the realization significantly, as supposed by previous research (Gómez Capuz 2001; Rodríguez González 2017). This suggests that speakers do not necessarily require a high level of education to have contact with the English language. Conversely, individuals with higher levels of education may more frequently encounter puristic tendencies and might have been exposed to stricter ideologies regarding the avoidance of foreign elements in Spanish. Furthermore, age did not significantly impact the realization of Anglicisms. Previous studies often presumed that younger speakers would be more inclined to imitate the English language compared to older speakers (Rodríguez González 2017). However, the statistical analysis unveiled that age merely acts as a confounding factor for LEAS. In essence, it is not the age of the speakers but rather their exposure to the English language that influences the pronunciation of loanwords.

Moreover, the area of residence (urban vs. rural) did not influence the realization significantly: In both areas observed, individuals have access to diverse English media, enabling them to engage with the English language. Nevertheless, it has to be noted, that the rural areas included in this study are within reach of the metropolis. Future studies should, therefore, also include far more remote rural areas even if the varieties spoken in these areas might differ from those spoken in the capitals. Furthermore, GS was not a significant factor. It appears that the realization of phonemes is not affected by speakers' attitudes toward globalization and the connection between different cultures and languages. However, in this case, some limitations should be considered since GS showed relatively similar results for the different speakers. This might be due to a social desirability bias (Diekmann 2009; Fisher 1993), which means that participants tend to avoid giving answers that they think are not accepted by society. It is crucial, however, to assess whether the selected questions, despite being adapted from reputable opinion research institutes' questionnaires, effectively differentiated between speakers' attitudes towards globalization (see for instance the question about the willingness to show their children foreign cultures).

Moreover, the analysis showed that the country of origin of the speakers significantly influences the realization, as speakers from Mexico are more likely to imitate the English phoneme than speakers from Spain. This finding confirms previous research, which underlined a higher influence of English in Hispanoamerica than in Spain (Oncins-Martínez 2009; Pustka 2021; Rodríguez González 2017), and indicates that the imitation of the foreign phoneme is still more common in Mexico than in Spain.

Various factors, beyond just geographic proximity, could contribute to the greater influence of English in Mexico compared to Spain. Spain may be less economically influenced by the U.S. compared to Mexico, but also other factors such as linguistic purism might play a role. Moreover, it is plausible that certain Anglicisms have different historical trajectories in Mexico and Spain, which means that, in Spain, they may have been adopted primarily through the visual channel (*eye loans*), whereas in Mexico, they may have entered the language through auditory channels (*ear loans*, see Section 1).

The model showed that individual words have an influence (random effects): frequency of use and age of the loanword, in general, did not impact the realization significantly, but there may be other factors that explain the influence of the random effects. As already stated, whether the words entered the lexicon via the oral or the written channel can influence the realization of the loanword. As Stone (1957, p. 149) already assumed in his analysis of Anglicisms in Spain, the Anglicism *jersey* entered the Spanish lexicon in Spain as an eye loan. The present data showed that the lexeme is consistently pronounced with grapheme-phoneme correspondence by the Spaniards, while the English phoneme /d͡ʒ/ is imitated by the Mexicans. Besides the path of entrance, the different diffusion processes might also influence the adaptation of the loanwords (see Section 1). Therefore, the grapheme <u> might be pronounced with grapheme-phoneme correspondence in the Anglicism *club* by Mexicans, while they pronounce the same grapheme by imitating the English language in the Anglicism *puzzle*. The involvement of different social groups might also play a decisive role in the diffusion process: as shown by the data, contact with the English language plays an important role in the adaptation process. If the loanword enters the language via an innovator who is frequently in contact with the English language, this might promote the diffusion of the imitated form of the pronunciation.

To summarize, as the disparities among the Anglicisms cannot be elucidated solely by considering the age and frequencies of the loanwords, and given the observed intra-speaker variation, it is reasonable to posit that the unique histories associated with each word also influence the pronunciation of the various phonemes. Therefore, the present data show that "each word has its own history".[3]

## 5. Conclusions

Concerning the factors impacting the phonic adaptation of Anglicisms, it became evident that exposure to and affinity towards the English language significantly influenced the adaptation process. Individuals with more frequent contact with English and a more positive attitude towards it were inclined to imitate English pronunciation more than those with less exposure and a less favorable attitude. Furthermore, Mexican speakers exhibited a higher tendency to imitate the English model compared to speakers from Spain, which is in line with previous research that assumes a greater influence of English in Hispanoamerica than in Spain (Oncins-Martínez 2009; Pustka 2021; Rodríguez González 2017).

However, several limitations must be kept in mind when interpreting the data: the complex LEAS variable was calculated from self-reported data. It is possible that participants understand the Likert scale differently, and therefore, rank the same amount of media use differently. In addition, self-reported data have the limitation that participants provide socially desirable answers and thus might, for example, not reveal their actual media consumption. Furthermore, it has to be remembered that the data are based on a reading task, namely the reading of a word list. This may impose several limitations, since, on the one hand, speakers might tend to realize Spanish grapheme-phoneme correspondences, as already discussed, when exposed to the graphic code. On the other hand, some speakers may also tend to imitate the English model in cases where they would not have done so if the word were embedded in a Spanish text or in spontaneous speech. Future research should, therefore, also analyze the pronunciation of Anglicisms using elicited data, such as a discourse completion task, in which speakers are not exposed to the graphic code.

The implementation of the statistical model allowed us to analyze multiple predictors and to show which factors directly influence loanword adaptation. The analysis of the random factors also showed that the Anglicisms themselves influence the adaptation of the English phonemes. Future research should, therefore, be based on comparative corpora that include, along with reading tasks, elicited or spontaneous data and comparison of different Spanish varieties.

**Funding:** This research was funded by the University of Vienna, who financed the contracts for transcriptions and field research in Mexico and Spain as well as by the the Rosita Schjerve-Rindler Fonds which contributed financially for the field research stay in Spain.

**Institutional Review Board Statement:** The procedure was approved by the doctoral program of the University of Vienna.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The raw data supporting the conclusions of this article will be made available by the authors on request.

**Acknowledgments:** Thank you to Elissa Pustka and two anonymous reviewers for their helpful comments on this paper. ChatGPT was used for grammar improvement. Thanks to Hannah Meurice for proofreading. Thank you to Luke Green and Martin Baumgartner for transcribing the data and to all the speakers who were willing to participate in the study. Open Access Funding by the University of Vienna.

**Conflicts of Interest:** The author declares no conflicts of interest.

## Notes

1. Note that for the realization of <r>, the grapheme is indicated since the pronunciation varies between General American and Received Pronunciation.
2. See Bäumler and Hartmann (2023) for an acoustic analysis of the realization of the English vowel /ʌ/ e.g., in the loanword *brunch*.
3. Pustka (2007, p. 40) states that the quote "chaque mot a son histoire" can be assigned to Gilliéron: "Cette phrase ne se trouve pas explicitement dans l'oeuvre de Gilliéron, mais lui est attribuée par de nombreux auteurs".

## References


Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67: 1–48. [CrossRef]

Baumgardner, Robert J. 2006. The appeal of English in Mexican commerce. *World Englishes* 25: 251–266. [CrossRef]

Bäumler, Linda. 2023. Anglicisms in Mexico and Spain in Times of Globalization: Use, Pronunciation and Perception. Ph.D. Thesis, Universität Wien, Wien, Austria.

Bäumler, Linda, and Frederik Hartmann. 2023. Sociogeographical differences in the pronunciation of Spanish Anglicisms: Acoustic vowel analysis using multilevel regression models. In *Corpus Dialectology: From Methods to Theory (French, Italian, Spanish)*. Amsterdam: John Benjamins.

Best, Catherine. 1994. The Emergence of Native-Language Phonological Influences in Infants: A Perceptual Assimilation Model. In *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*. Edited by Judith Claire Goodman and Howard C. Nusbaum. Cambridge, MA: MIT Press, pp. 167–224.

Boersma, Paul, and David Weenink. 2021. Praat: Doing Phonetics by Computer [Computer Program].

Brannen, Kathleen. 2002. The Role of Perception in Differential Substitution. *Canadian Journal of Linguistics/Revue Canadienne de Linguistique* 47: 1–46. [CrossRef]

Bravo-Valdivieso, Luis, and José-Pablo Escobar. 2014. How transparent is spanish orthography? *Estudios de Psicologia* 35: 442–49. [CrossRef]

Calabrese, Andrea. 2009. Perception, production and acoustic inputs in loanword phonology. In *Loanword Phonology*. Edited by Andrea Calabrese and W. Leo Wetzels. Amsterdam and Philadelphia: John Benjamins, pp. 59–114. [CrossRef]

Chang, Charles B. 2008. Phonetics vs. Phonology in Loanword Adaptation: Revisiting the Role of the Bilingual. *Berkeley Linguistics Society* 34: 61–72. [CrossRef]

Chang, Charles B. 2009. English loanword adaptation in Burmese. *Journal of the Southeast Asian Linguistics Society* 1: 77–49.

Chaudron, Craig. 2007. Data collection in SLA research. In *The Handbook of Second Language Acquisition*. Edited by Catherine J. Doughty and Michael H. Long. Malden: Blackwell, pp. 762–828.

Cohen, Jacob. 1960. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement* 20: 37–46. [CrossRef]

Comisión de Lingüística Iberoamericana. 1971. *Cuestionario para el estudio coordinado de la norma lingüística culta de las principales ciudades de Iberoamérica y de la Península Ibérica. III – Léxico*. Madrid: Consejo Superior de Investigaciones Científicas.

Despagne, Colette. 2010. The Difficulties of Learning English: Perceptions and Attitudes in Mexico. *Canadian and International Education* 39: 55–74. [CrossRef]

Diekmann, Andreas. 2009. *Empirische Sozialforschung*, 20th ed. Reinbek: Rowohlt.

Durand, Jacques and Anne Przewozny. 2012. La phonologie de l'anglais contemporain : Usages, variétés et structure. *Revue Française de Linguistique Appliquée* 17: 25–37. [CrossRef]

Durand, Jacques, Bernard Laks, and Chantal Lyche. 2002. La *Phonologie du Français Contemporain*: Usages, variétés et structure. In *Romanistische Korpuslinguistik—Korpora und gesprochene Sprache/Romance Corpus Linguistics—Corpora and Spoken Language*. Edited by Claus Pusch and Wolfgang Raible. Tübingen: Narr, pp. 93–106.

Dworkin, Steven N. 2012. *A History of the Spanish Lexicon: A Linguistic Perspective*. Oxford: Oxford University Press. [CrossRef]

EVS. 2010. European Values Study 2008: Spain (evs 2008). Available online: https://search.gesis.org/research_data/ZA4756?doi=10.4232/1.10153 (accessed on 31 October 2022).

Eychenne, Julien. 2021. Corpus phonology. In *Manual of Romance Phonetics and Phonology*. Edited by Christoph Gabriel, Randall Gess and Trudel Meisenburg. Berlin and Boston: De Gruyter, pp. 319–41.

Eychenne, Julien and Bernard Laks. 2012. Le programme *Phonologie du Français Contemporain* : Bilan et perspectives. *Revue Française de Linguistique Appliquée* 7: 7–24. [CrossRef]

Ferreres, Aldo-Rodolfo, and Cynthia-Valeria López. 2014. Orthographic transparency and acquired dyslexias (alexias) in spanish speakers: A review. *Studies in Psychology* 35: 519–44. [CrossRef]

Fisher, Robert J. 1993. Social desirability bias and the validity of indirect questioning. *Journal of Consumer Research* 20: 303–15. [CrossRef]

Flege, James E., and Ratree Wayland. 2019. The role of input in native Spanish Late learners' production and perception of English phonetic segments. *Journal of Second Language Studies* 2: 1–44. [CrossRef]

Fleiss, Joseph L., Bruce Levin, and Myunghee Cho Paik. 2003. *Statistical Methods for Rates and Proportions*, 3rd ed. Hoboken: John Wiley & Sons.

Gayo, Modesto, and Berta Teitelboim. 2008. Localismo cosmopolitismo y gustos musicales. *Reporte Encuesta Nacional de Opinión Pública de la Universidad Diego Portales (4ª versión): Radiografía social, política y económica de Chile* 2: 111–20.

Gimeno Menéndez, Francisco, and María Victoria Gimeno Menéndez. 2003. *El desplazamiento lingüístico del español por el inglés*. Madrid: Cátedra.

Gómez Capuz, Juan. 2001. Estrategias de integración fónica de los anglicismos en un corpus de español hablado: Asimilación, compromiso y efectos estructurales. *Estudios de Lingüística* 15: 5–86. [CrossRef]

Gray, Jonathan. 2007. Imagining America: The Simpsons Go Global. *Popular Communication* 5: 129–48. [CrossRef]

Hamm, Bernd, and Russell Smandych, eds. 2005. *Cultural Imperialism. Essays on the Political Economy of Cultural Domination*. Toronto: University of Toronto Press.

Haspelmath, Martin. 2009. Lexical borrowing: Concepts and issues. In *Loanwords in the World's Language: A Comparative Handbook*. Edited by Martin Haspelmath and Uri Tadmor. Berlin and New York: De Gruyter Mouton, pp. 35–54.

Haugen, Einar. 1950. The analysis of linguistic borrowing. *Language* 26: 210–31. [CrossRef]

Heffernan, Kevin. 2007. The role of phonemic contrast in the formation of Sino-Japanese. *Journal of East Asian linguistics* 16: 61–86. [CrossRef]

INE, Instituto Nacional de Estadística. 2019. Cifras oficiales de población resultantes de la revisión del Padrón municipal a 1 de enero. Available online: https://www.ine.es/jaxiT3/Datos.htm?t=2881#!tabs-tabla (accessed on 17 November 2021).

INEGI, Instituto Nacional de Estadística Geografía e Informática. 2009. Prontuario de información geográfica municipal de los Estados Unidos Mexicanos. Available online: https://www.inegi.org.mx/contenidos/app/mexicocifras/datos_geograficos/15/15030.pdf (accessed on 17 November 2021).

Inglehart, Ronald, Christian Haerpfer, Alejandro Moreno, Christian Welzel, Kseniya Kizilova, Jaime Diez-Medrano, Marta Lagos, Pippa Norris, Eduard Ponarin, and Bi Puranen. 2014. World Values Survey: Round Six. Available online: https://www.worldvaluessurvey.org/WVSDocumentationWV6.jsp (accessed on 31 October 2022).

Kachru, Braj B. 1985. Standards, codification and sociolinguistic realism: The English language in the outer circle. In *English in the World. Teaching and Learning the Language and Literatures*. Edited by Randolph Quirk and Henry G. Widdowson. Cambridge: Cambridge University Press, pp. 11–30.

Kang, Yoonjung. 2003. Perceptual similarity in loanword adaptation: English postvocalic word-final stops in Korean. *Phonology* 20: 219–73. [CrossRef]

Kang, Yoonjung. 2011. Loanword Phonology. In *The Blackwell Companion to Phonology*. Edited by Marc van Oostendorp, Colin Ewen, Elizabeth Hume and Keren Rice. Malden: John Wiley & Sons, Ltd., pp. 1–25.

Kenstowicz, Michael. 2007. Salience and similarity in loanword adaptation: A case study from Fijian. *Language Sciences* 29: 316–40. [CrossRef]

Kilgarriff, Adam and Irene Renau. 2013. esTenTen, a vast web corpus of Peninsular and American Spanish. *Procedia-Social and Behavioral Sciences* 95: 1–8. [CrossRef]

Kilgarriff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý, and Vít Suchomel. 2014. The sketch engine: Ten years on. *Lexicography* 1: 7–36. [CrossRef]

Kowner, Rotem and Judith Rosenhouse. 2008. The hegemony of English and determinants of borrowing from its vocabulary. In *Globally Speaking. Motives for Adopting English Vocabulary in Other Languages*. Edited by Judith Rosenhouse and Rotem Kowner. Clevedon: Multilingual Matters, pp. 4–18. [CrossRef]

Kwok, Rosa Kit Wan, Fernando Cuetos, Rrezarta Avdyli, and Andrew W. Ellis. 2017. Reading and lexicalization in opaque and transparent orthographies: Word naming and word learning in english and spanish. *Quarterly Journal of Experimental Psychology* 70: 2105–29. [CrossRef] [PubMed]

LaCharité, Darlene, and Carole Paradis. 2005. Category preservation and proximity versus phonetic approximation in loanword adaptation. *Linguistic Inquiry* 36: 223–58. [CrossRef]

Lallier, Marie, Sylviane Valdois, Delphine Lassus-Sangosse, Chloé Prado, and Sonia Kandel. 2014. Impact of orthographic transparency on typical and atypical reading development: Evidence in french-spanish bilingual children. *Research in Developmental Disabilities* 35: 1177–90. [CrossRef]

Lambert, Wallace E., Richard C. Hodgson, Robert C. Gardner, and Samuel Fillenbaum. 1960. Evaluational reactions to spoken languages. *Journal of Abnormal and Social Psychology* 60: 44–51. [CrossRef] [PubMed]

Landis, J. Richard and Gary G. Koch. 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics* 33: 159–74. [CrossRef] [PubMed]

Llompart, Miquel, and Eva Reinisch. 2019. Imitation in a Second Language Relies on Phonological Categories but Does Not Reflect the Productive Usage of Difficult Sound Contrasts. *Language and Speech* 62: 594–622. [CrossRef]

Lope Blanch, Juan M. 1986. *El estudio del español hablado culto. Historia de un proyecto*. Ciudad de México: Universidad Nacional Autónoma de México.

Meisenburg, Trudel. 1992. Graphische und phonische Integration von Fremdwörtern am Beispiel des Spanischen. *Zeitschrift fur Sprachwissenschaft* 11: 47–67. [CrossRef]

Milroy, James, and Lesley Milroy. 1985. Linguistic Change, Social Network and Speaker Innovation. *Journal of Linguistics* 21: 339–84. [CrossRef]

Milroy, Lesley, and Matthew Gordon. 2003. *Sociolinguistics: Method and Interpretation*. Chichester: Wiley-Blackwell.

Moreno de Alba, José G. 1992. *Diferencias léxicas entre España y América*. Madrid: Mapfre.

Oncins-Martínez, José. 2009. Towards a corpus-based analysis of Anglicism in Spanish: A case study. *International Journal of English Studies* 9: 115–32.

Paradis, Carole, and Antoine Tremblay. 2009. Nondistinctive features in loanword adaptation: The unimportance of English aspiration in Mandarin Chinese Phoneme categorization. In *Loan Phonology*. Edited by Andrea Calabrese and Wetzels L. Wetzels. Amsterdam and Philadelphia: John Benjamins, pp. 211–24.

Peperkamp, Sharon, and Emmanuel Dupoux. 2003. Reinterpreting loanword adaptations: The role of perception. Paper presented at 15th International Congress of Phonetic Sciences, Barcelona, Spain, August 3–9; vol. 367, pp. 367–70.

Peperkamp, Sharon, Inga Vendelin, and Kimihiro Nakamura. 2008. On the perceptual origin of loanword adaptations: Experimental evidence from Japanese. *Phonology* 25: 129–64. [CrossRef]
Poplack, Shana, David Sankoff, and Christopher Miller. 1988. The social correlates and linguistic processes of lexical borrowing and assimilation. *Linguistics* 26: 47–104. [CrossRef]
Pratt, Chris. 1980. *El anglicismo en el español peninsular contemporáneo*. Madrid: Gredos.
Pustka, Elissa. 2007. *Phonologie et variétés en contact. Aveyronnais et Guadeloupéens à Paris*. Tübingen: Narr.
Pustka, Elissa. 2021. Loanword phonology in Romance. In *Manual of Romance Phonetics and Phonology*. Edited by Christoph Gabriel, Randall Gess and Trudel Meisenburg. Berlin and Boston: De Gruyter, pp. 503–28.
Pustka, Elissa, Christoph Gabriel, and Trudel Meisenburg. 2016. Romance Corpus Phonology: From *(Inter-)Phonologie du Francais Contemporain* (I)PFC to *(Inter-)Fonología del Español Contemporáneo* (I)FEC. In *Tagungsband der 12. Tagung Phonetik und Phonologie im deutschsprachigen Raum*. Edited by Christoph Draxler and Felicitas Kleber. München: Ludwig-Maximilians-Universität, pp. 151–54.
Pustka, Elissa, Christoph Gabriel, Trudel Meisenburg, Monja Burkard, and Kristina Dziallas. 2018. (Inter-)Fonología del Español Contemporáneo (I)FEC: Metodología de un programa de investigación para la fonología de corpus. *Loquens* 5: 1–16. [CrossRef]
R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
Real Academia Española. 2019. Nuevo tesoro lexicográfico (NTLLE). Available online: https://apps.rae.es/ntlle/SrvltGUISalirNtlle (accessed on 10 February 2019).
Rodríguez González, Félix. 2017. Variaciones fonológicas en el uso del anglicismo: Panorama y revisión crítica. *Revista Española de Lingüística* 47: 99–134.
Rodríguez González, Félix. 2018. Aspectos ortográficos del anglicismo. *Lebende Sprachen* 63: 350–73. [CrossRef]
Rogers, Everett M., and F. Floyd Shoemaker. 1971. *Communication of Innovations. A Cross-Cultural Approach*. New York: The Free Press.
Schweickard, Wolfgang. 1991. Anglizismen im Spanischen. *Terminologie et Traduction* 1: 75–86.
Seymour, Philip H.K., Mikko Aro, and Jane M. Erskine. 2003. Foundation literacy acquisition in european orthographies. *British Journal of Psychology* 94: 143–74. [CrossRef]
Signorell, Andri. 2024. DescTools: Tools for Descriptive Statistics. R Package Version 0.99.44. Available online: https://github.com/AndriSignorell/DescTools/ (accessed on 15 January 2024).
Sim, Julius, and Chris C. Wright. 2005. The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Physical Therapy* 85: 257–69. [CrossRef]
Stone, Howard. 1957. Los anglicismos en España y su papel en la lengua oral. *Revista de Filología Española* 41: 141–60. [CrossRef]
Van Coetsem, Frans. 1988. *Loan Phonology and the Two Transfer Types in Language Contact*. Dordrecht: Foris. [CrossRef]
Vendelin, Inga, and Sharon Peperkamp. 2004. Evidence for phonetic adaptation of loanwords: An experimental study. *Actes des Journées d'Etudes Linguistiques* 129–31.
Vendelin, Inga and Sharon Peperkamp. 2006. The influence of orthography on loanword adaptations. *Lingua* 116: 996–1007. [CrossRef]
Winter-Froemel, Esme. 2011. *Entlehnung in der Kommunikation und im Sprachwandel: Theorie und Analysen zum Französischen*. Berlin and Boston: De Gruyter.

MDPI AG
Grosspeteranlage 5
4052 Basel
Switzerland
Tel.: +41 61 683 77 34

*Languages* Editorial Office
E-mail: languages@mdpi.com
www.mdpi.com/journal/languages

MDPI

Academic Open
Access Publishing

mdpi.com

ISBN 978-3-7258-6313-6