
This is the **accepted version** of the journal article:

Repiso-Puigdelliura, Gemma. «Development of Vowel Intrusion in Spanish Heritage Speakers». *Language and Speech*, Vol. 68, Num. 3 (2025), p. 633-654
DOI 10.1177/00238309251315068, PMID 40012513

This version is available at <https://ddd.uab.cat/record/336178>

under the terms of the  license.

INTRUSIVE VOWELS IN SPANISH HERITAGE SPEAKERS

Development of Vowel Intrusion in Spanish Heritage Speakers

Gemma Repiso-Puigdelliura

Universitat Autònoma de Barcelona

Abstract

This study investigates the sound system of heritage speakers (henceforth, HSs) as they shift dominance from their heritage language to the majority language. Specifically, it analyzes the production of intrusive vowels in Spanish consonant clusters across the lifespan of HSs, focusing on tautosyllabic clusters (i.e., /Cr/) and heterosyllabic clusters (i.e., /r.C/). Semi-spontaneous speech was elicited from three age groups of Spanish HSs—younger children, older children, and adults—for whom American English is the majority language, as well as from three age-matched groups of non-heritage Spanish speakers raised in Mexico. The presence and duration of intrusive vowels were examined, with the latter calculated as a ratio of intrusive vowel duration to tap duration. Overall, the results indicate that all speaker groups more frequently produce consonant clusters with intrusive vowels than without. However, Spanish HSs produced fewer and shorter intrusive vowels compared to their non-heritage counterparts. In addition, heritage and non-heritage speakers exhibited a lower rate of intrusive vowel production in the 5-to-8-year-old group compared to adults. Despite this, the absence of an interaction between age group and speaker type suggests a parallel pattern of change in intrusive vowel production across all ages. Voicing emerged as the most consistent phonetic factor, predicting a higher rate of intrusive vowels with a longer duration relative to the tap.

Development of Vowel Intrusion in Spanish Heritage Speakers

Introduction

Heritage speakers (i.e., HSs henceforth) are a type of early bilingual who speak a socio-politically minority language as their first language in a bilingual or multilingual context. The most extensively studied social context that gives rise to heritage bilingualism, and the focus of this study, is the migratory context, where a heritage language is spoken by second-generation individuals whose parents grew up in the homeland and migrated during adulthood (Montrul, 2023). In the domain of speech perception and production, extensive research over the last decade has shown that, despite initial findings showing that HSs overperform L2 speakers (Au et al., 2008, 2002; Oh et al., 2003), HSs demonstrate divergences in their speech system when compared to non-heritage native speakers (Rao, 2024; Kim, 2024, see for recent reviews). These differences can be attributed to the particular contextual effects of this type of unbalanced bilingualism, which sometimes involve a shift in language dominance possibly triggered by reduced exposure to the heritage language and increased exposure to the majority language—from early dominance in the heritage language before the primary school years to late dominance in the majority language (Benmamoun et al., 2013). Examining HSs' speech production during this shift in dominance is crucial for understanding the restructuring of linguistic representations within their systems, leading to these observed differences when HSs are compared to non-heritage native speakers.

When examining heritage language speech systems across the HSs' lifespan, it is important to consider the dynamic nature of language development and drift. Polinsky (2018) proposes four possible scenarios for how specific linguistic properties might evolve: 1) a given property is equally produced by child HSs, adult HSs, and the baseline group, 2) a parallel change is observed in both child and adult HSs when compared to the baseline group, 3) a given property is maintained in child HSs but not in adult HSs, or 4) a property is found in adult HSs but not in child HSs. Developmental studies on the sound

systems of HSs primarily support the third and fourth scenarios. The third scenario is often understood as the attrition, loss, or lack of activation of a specific feature in the heritage language, which is likely replaced by a corresponding feature from the majority language. Kupisch et al. (2021), for example, provide evidence for attrition-like scenarios, as the authors found that Russian HSs with German as the majority language sound more foreign-like in Russian, their heritage language, than in German, their majority language, during the primary school years (i.e., from 7 to 9 years old), while they sound more foreign-like in German than in Russian during the preschool years (i.e., from 4 to 6 years old). The authors attribute this shift to both quantitative and qualitative changes in exposure to German, such as the onset of schooling and increased interaction in formal settings, relative to exposure to Russian. The fourth scenario, by contrast, entails either a reanalysis or protracted development of a linguistic feature during adulthood, or stronger activation of that feature—often at the expense of a corresponding property in the majority language. Evidence also exists of adult HSs outperforming child HSs. In a study on the production of connected speech, Spanish child HSs with English as the majority language (i.e., 5 to 11 years) demonstrated higher rates of English-like glottalization compared to their monolingual peers, while adult HSs did not show this difference compared to their monolingual peers (Repiso-Puigdelliura, 2023). These results suggest that either accumulated experience in the heritage language or an improved ability to inhibit the activation of non-target language (i.e., the language not intended for use in the given context) reduces the influence of the majority language. Moreover, a specific instance of protracted development can be observed with late-developing sounds, in which older HSs tend to incorporate these sounds more extensively than younger HSs due to a possible slower development when both input and output in the heritage language are reduced. This is the case of the Spanish trill, which has been found to continue to develop beyond the age of monolingual acquisition (Repiso-Puigdelliura and Kim, 2021; Menke, 2018).

The varying developmental trends in these studies suggest that the effects of a

dominance shift do not occur uniformly across all domains of the HSs' sound systems. This variation supports a modular view of the HSs' sound systems, in which the domains of sound representation and production are more or less affected by the particular contextual effects of unbalanced bilingualism. Natvig (2021) recently put forth a formal approach to a modular integrated sound system for multilingual speakers consisting of the phonological module (i.e., contrastive representations), the phonology-phonetics interface (i.e., implementation of speech gestures), and the phonetics module (i.e., online production of speech sounds). In this framework, the phonological module pertains to grammatical knowledge and is less affected by contextual changes, whereas the phonology-phonetics interface and the phonetics module are linked to procedural knowledge and are more vulnerable to change in differing contextual variations.

At the procedural level, most research on the heritage language sound system has focused on how contact with the majority language impacts the production of individual sounds (e.g., Voice Onset Time values, degrees of intervocalic lenition) (Rao, 2019, 2024; Chang, 2021; Kim, 2024, see for recent overviews). However, less attention has been given to how these segments are coordinated with each other (Beristain, 2023, for example see). Nevertheless, from a heritage language developmental perspective, examining the malleability of speech gesture coordination is crucial because studies have found that segment-to-segment coordination extends into late childhood, as evidenced by findings in English (Nitttrouer, 1993) and German (Rubertus et al., 2015; Rubertus and Noiray, 2020; Noiray et al., 2018). Early stages of phonological development appear to be holistic, with a lesser differentiation of speech gestures. A refinement of the speech motor control during phonological development leads to a greater separation of the speech gestures and, thus, lower rates of gestural overlap (Nitttrouer, 1993; Noiray et al., 2018; Rubertus et al., 2015; Rubertus and Noiray, 2020). There is some support, however, for an early stage of reduced interaction between speech gestures, with gestural overlap increasing as children grow older (Barbier et al., 2020; Nitttrouer et al., 1989). Despite these varying results, the overall trend

that segment-to-segment coordination develops during the preschool years and continues into primary school indicates that language-specific gestural timing may be particularly susceptible in a heritage language, especially during the period in which HSs receive systematic exposure to the majority language. This highlights the importance of examining segment-to-segment coordination in the development of heritage language speech systems.

This study specifically investigates the development of Spanish intrusive vowels, which are optional phonetic vocoids arising in the open transitions of consonant clusters (e.g., [ˈti.ɣere] for [ˈti.ɣre] ‘tigre’ *tiger*). These vowels are not unique to Spanish and have been observed cross-linguistically in languages such as Catalan, Dutch, Norwegian, Polish, and Turkish, among others (Bellik, 2018; Recasens, 2018; Bradley, 2002; Booij, 1995). Importantly, in American English, intrusive vowels are mostly unattested, with gestures in consonantal clusters described as overlapping (Byrd, 1996; Zsiga, 2000; Recasens et al., 1993; Colantoni and Steele, 2006). Consequently, the difference in intrusive vowel production between Spanish and American English provides an interesting test case for studying potential transfer effects when developing segment-to-segment coarticulation in Spanish HSs for whom English is the majority language.

Background

Intrusive vowels in consonant clusters

Unlike other cases of inserted vowels, such as phonological epenthetic vowels, intrusive vowels do not repair phonologically ill-formed structures and appear to be transparent to the phonological system; for example, they do not form syllable nuclei (Hall, 2006). The production of intrusive vowels is explained as a phonetic transition resulting from specific timing between consonant gestures, specifically as a voiced lag between non-overlapping consonantal gestures (Bradley, 2007; Gibson et al., 2019). However, it remains unclear whether intrusive vowels have an acoustic target and occur at some level of speech planning, or if they are an aeroacoustic by-product of speech gesture execution. In support of the explanation of active laryngeal timing, Gibson et al. (2021) examined

aeroacoustic properties of open transitions (e.g., intensity, duration, amplitude) in Georgian, Spanish, and Moroccan Arabic. Gibson et al. (2021) found that machine-learning models (random forest classifiers) successfully learned the specific parameters of each language, suggesting that the same speech gestures result in differentiated and learnable aeroacoustic parameters. Moreover, anecdotal evidence suggests that singers can manipulate the length and/or production of the intrusive vocoid (Kochetkova, 2019). An explanation based on speech gesture execution posits that a low gestural overlap between the two consonantal gestures allows for a lag between the constrictions, which can be voiced and acoustically manifested as a vocalic element (Bradley, 2007; Gibson et al., 2019). This explanation finds support in research showing that the production of intrusive vocoids is affected by the phonetic context, such as the place and manner of articulation of the consonants in the consonant cluster, their syllable organization, or the type of voicing in the cluster (Schmeiser, 2020; Colantoni and Steele, 2007; Savu, 2013). For instance, a consistent environment for intrusive vowels is the consonant-rhotic cluster, as evidenced by cross-linguistic studies in Turkish (Bellik, 2018), Spanish, Italian, and Catalan (Recasens, 2018). Furthermore, within these consonant-rhotic clusters, the apical gesture elicits a higher rate of intrusive vowels compared to the uvular gesture (Lancien and Côté, 2019).

Like intrusive vowels, intergestural timing has also been found to differ across consonant cluster types, varying by voicing, place of articulation, manner of articulation, and syllabic organization (Hardcastle and Roach, 1979; Byrd, 1996; Recasens et al., 1993; Pouplier et al., 2022). As for place of articulation, for example, Byrd (1996) provides evidence for a back-to-front effect in English, where tongue tip consonants are more overlapped when followed by a tongue body consonant [e.g., d # g] than when a tongue tip consonant follows a tongue body consonant [e.g., g # d]. This effect could be explained by considering the recoverability of the first consonant (i.e., C1). With more posterior places of articulation for the second consonant (i.e., C2), information can be recovered from the release of C1 (Chitoran et al., 2002, p. 424). In the specific case of [e.g., d # g], Byrd

(1996, p. 235) also argues that speakers may make less effort to preserve weaker perceptual cues. In addition, Byrd (1996) found that onset clusters (e.g., *Type a scab again*) were less overlapped than heterosyllabic clusters (e.g., *Type bass cap again*) or clusters in coda position (e.g., *Type mask amp again*). More recent cross-linguistic evidence (Poupplier et al., 2022) proposes that languages use a core set of common overlap patterns, with the range of these patterns varying across languages and clusters. For example, Poupplier et al. (2022) found that the obstruent and its associated phonation influenced the intergestural plateau interval. For example, although /kn/ clusters had a comparably large intergestural plateau interval across different languages, this interval was substantially devoiced in all cases.

In particular, our study focuses on Spanish clusters that contain an apicoalveolar tap. The apicoalveolar tap is produced by a primary gesture of the tongue tip involving a single rapid ballistic movement (*hit-and-run*) with the alveolar ridge (Catford, 1977, p. 130) and a secondary dorsal component (Proctor, 2011). The consonants in these clusters can be arranged in the same syllable (i.e., tautosyllabic position - /Cr/) or in different syllables (i.e., heterosyllabic position - /r.C/). In /Cr/ clusters, the tap may precede a labiodental fricative (e.g., [ˈfre.sa] *fresa* ‘strawberry’), a stop (e.g., [tren] *tren* ‘train’), or a spirant (e.g., [ˈo.ɣro] *ogro* ‘ogre’). In /r.C/ clusters, the tap appears before a nasal (e.g., [ar.ˈma.rjo] *armario* ‘closet’), a voiceless stop (e.g., [ˈar.te] *arte* ‘art’), a fricative (e.g., [per.ˈfek.to] *perfecto* ‘perfect’) or a spirant (e.g., [ˈar.βol] *árbol* ‘tree’).

Intrusive vowels have been reported in /Cr/ and /r.C/ clusters across various dialects, including Castilian Spanish (Gili Gaya, 1921; Blecua, 2001; Schmeiser, 2006), Argentine Spanish (Schmeiser, 2006; Colantoni and Steele, 2007), highland Ecuadorian Spanish (Bradley, 2004), Mexican and Guatemalan Spanish (Schmeiser, 2009; Ramírez, 2006), as well as Colombian, Salvadorian and Peruvian Spanish (Ramírez, 2006). The presence of intrusive vowels in these clusters is often linked to the articulation of the tap. In /Cr/ clusters, the tongue must retract to increase the velocity to allow the tip of the tongue to come into contact with the alveolar ridge, which can contribute to the extended

lag between the two consonants (Gibson et al., 2019, 2018). In /r.C/ clusters, intrusive vowels may result from the rapid release of the tap and the subsequent consonantal gesture (Recasens, 2018). The reported rate of intrusive vowels varies across studies, ranging from 63% to 97.7% (Ramírez, 2006; Blecua, 2001; Schmeiser, 2020; Colantoni and Steele, 2005). The broad range in percentages may stem from phonetic effects or dialectal variation. In a cross-dialectal study, Schmeiser (2006, p. 50) shows that Ecuadorian Spanish demonstrates the highest rate of vowel intrusion (84.21%), while speakers from Spain, Argentina, Mexico, Guatemala, or Colombia exhibit rates between 50% and 66%. Intrusive vowels in /r.C/ clusters have received less attention in research but also show variability, with occurrences ranging from 61.2% in Castilian Spanish (Gili Gaya, 1921) to 84.21% in Argentine Spanish (Schmeiser, 2006), or 61.2% in Guatemalan Spanish (Schmeiser, 2009). Direct comparisons with the same speakers reveal, in addition, higher rates of intrusive vowels in /Cr/ clusters than in /r.C/ clusters (Schmeiser, 2006, p. 50). It's important to note that these aggregated percentages may reflect a distribution pattern where some speakers consistently produce intrusive vowels, while others do so only rarely.

In terms of duration, intrusive vowels have a length similar to unstressed vowels (Malmberg, 1965) and are generally longer than the tap gesture (Blecua, 2001; Gili Gaya, 1921). Apart from being influenced by the syllabic affiliation of the consonant cluster, the rate and duration of intrusive vocoids also depend on other factors associated with their phonetic context. A common predictor of the rate and duration of intrusive vocoids is the voicing of the flanking consonant; intrusive vowels are longer and appear more often when the tap is flanked by voiced obstruents compared to voiceless ones (Colantoni and Steele, 2005; Schmeiser, 2020; Colantoni and Steele, 2007). Regarding the place of articulation of the non-tap consonant, the rate and duration of intrusive vowels are greater in heterorganic than in homorganic clusters, with longer durations found when the tap neighbors a dorsal consonant (Colantoni and Steele, 2005; Schmeiser, 2020), likely due to the slow excursion of the tongue body in velar consonants after and before an apical gesture (Recasens, 2018,

p. 128).

A crucial question for the learnability of intrusive vowels is whether they provide an advantage to the listener in discriminating between segments. In particular, the intrusive vocoid can enhance the perceptibility of the tap gesture, a brief segment whose perception is likely improved when produced between vowels (Bradley, 2002). Experimental evidence indicates that while intrusive vocoids are not essential for tap perception, they facilitate the identification of /Cr/ clusters (Romero, 2008). Romero (2008) presented Castilian Spanish listeners (Spanish-Catalan bilinguals) with /s/ and /r/ minimal pairs (e.g., *casta* ‘cast’ vs. *carta* ‘letter’) for which three reduction continua were created: (1) the /s/ portion was incrementally removed, (2) an intrusive vowel was added to the reduction continua in the first continuum, and (3) a smaller portion of the vowel was added to the first continuum. While listeners identified /r/ in reduced /s/ tokens, the presence of the vocalic element increased the perception of /r/ for the higher levels of consonant reduction.

Unlike in Spanish, where intrusive vowels are common, they are generally not attested in English. English shows considerable overlap between consonants (Byrd, 1996; Zsiga, 2000; Recasens et al., 1993; Pouplier et al., 2022), thus reducing the possibility of intrusive vocoids. Pouplier et al. (2022), for example, in a cluster analysis showed that 68% of the American English tautosyllabic consonant clusters exhibited high overlap. Colantoni and Steele (2006) examined the production of /Cr/ clusters (e.g., [ˈpɹaj.si] ‘pricey’ [ˈbɹi.zi] ‘breezy’) in ten English-speaking Spanish learners and found negligible rates of epenthesis and considerable rates of affrication. However, while intrusive vowels are rare, they are not impossible. Buchwald et al. (2007), for example, examined speech errors in an English-speaking aphasic patient who produced intrusive vowels (e.g., [ˈbəlɪd] ‘bleed’). Acoustic and ultrasound analyses comparing lexical vowels to these intrusive vowels showed similar influences from the stressed vowel and comparable rates of articulation and tongue contours, providing evidence for an acoustic target.

Regarding the identity of the rhotic sound in English, the closest approximation to

the Spanish apicoalveolar tap is the American English tap/flap. In English, this sound occurs between vowels, specifically when the second vowel is unstressed, as an allophone of /t/ and /d/ (e.g., ['bʌrəf] 'butter') (Kahn, 1976). Otherwise, American English has a voiced approximant that appears in consonant clusters (e.g., ['bɹeɪm] 'brain', ['ɑɹm] 'arm') (Ladefoged and Maddieson, 1998). This rhotic is characterized by a primary constriction in the oral cavity, achieved by raising the tip of the tongue (retroflex) or by raising the middle of the body of the tongue with the tip down (bunched), along with two secondary constrictions in the pharyngeal cavity and at the lips (Delattre and Freeman, 1968; Alwan et al., 1997; Tiede et al., 2004).

Acquisition and crosslinguistic influence of intrusive vowels in consonant clusters

Learnability of intrusive vowel production involves executing non-overlapped consonantal gestures and precise laryngeal coordination to achieve voiced intergestural transitions. Voiceless-voiced clusters require the ability to initiate voicing during the open transition before reaching the target of the following voiced consonant, while voiced-voiced and voiced-voiceless clusters demand the ability to maintain voicing during these transitions. In bilingual speakers, who are simultaneously exposed to a language with overlapped transitions, interference may appear at the level of syllable-timing or at that of the laryngeal configuration during the intergestural transitions. Gibson et al. (2018) examined syllable-level timing patterns in Spanish-English bilingual preschoolers aged 4;4 to 6;1 years, both with and without cochlear implants. Among other acoustic measures, the authors included the duration of the intergestural plateau interval, defined as the time between the acoustic offset of C1 and the acoustic onset of C2. Their results show that bilinguals, regardless of cochlear implant status, had acquired Spanish-like timing patterns for complex onsets and demonstrated the expected consonantal differences between stop+lateral and stop+rhotic clusters. Nevertheless, the taxing nature of the acquisition of intergestural timing in consonant clusters is supported by their findings showing that

INTRUSIVE VOWELS IN SPANISH HERITAGE SPEAKERS

vowels following complex onsets were lengthened, which the authors interpret as a slowdown in speech rate resulting from increased mechanical complexity. Similarly, the same speakers in English also show longer consonants in English stop-sonorant clusters (Gibson et al., 2022). In adult speakers, Tobin et al. (2019) examined /CrV/ and /CIV/ productions of two L1 American English - L2 Spanish speakers, measuring the duration between the release of the C1 gestural plateau and the onset of the C2 gestural plateau. They found that the L2 Spanish learners produced similar degrees of overlap compared to Spanish native speakers. Despite these similarities in speech timing, the L2 Spanish speakers did not exhibit examples of intrusive vowels, which could indicate a failure to initiate or maintain voicing in the open transitions. Along with these findings, an acoustic analysis of stop-rhotic consonant clusters by Colantoni and Steele (2006) indicated that intrusive vowels in Spanish L2 learners develop with increasing proficiency. The difference is more pronounced in voiceless clusters (e.g., [praðo] ‘prado’ *meadow*) than in voiced clusters (e.g., [siðra] ‘sidra’ *cider*). In voiceless clusters, intermediate learners produce intrusive vowels at a rate of 24%, a lower rate than the 94% observed in Spanish monolingual speakers. However, advanced learners show a rate of 80% vowel intrusion, bringing them closer to the monolingual rate. In voiced clusters, intermediate learners produce intrusive vowels at 79%, while advanced learners, at 97%, perform similarly to monolingual speakers, who also produce intrusive vowels at 97%.

Overall, these findings suggest that, even if Spanish-English bilinguals acquire language-specific intergestural timing, initiating or maintaining voicing in the open transitions can still be affected by interference from English. In addition, the findings suggesting a slowdown in speech rate in the child’s production of consonant clusters reveal the taxing nature of consonant cluster production, which suggests that, even in speakers raised in monolingual environments, consonantal transitions are mechanically complex.

Aims of the study

This study investigates the production of intrusive vowels in /Cr/ and /r.C/ clusters in Spanish. The high rate of intrusive vowels in Spanish compared to their negligible presence in English creates an interesting scenario for examining potential crosslinguistic transfer in the speech production of child and adult HSs. Specifically, Spanish HSs influenced by English may exhibit lower rates of intrusive vowels in the production of /Cr/ and /r.C/ consonant clusters. This is likely due to the adoption of English-like consonantal transitions, characterized by either overlapped consonantal gestures or voiceless C1-C2 intergestural plateaux. If intrusive vowels are produced, their implementation will likely differ from non-heritage native Spanish patterns, potentially resulting from semi-overlapped transitions between consonants or shorter periods of voicing between the two consonants.

The first research question examines whether Spanish HSs show influence from English in the production of intrusive vowels. This question will be addressed by analyzing the rate of intrusive vowel production in /Cr/ and /r.C/ consonant clusters, as well as the relative duration of these vowels compared to the duration of the preceding or following tap.

- We predict that, due to their exposure to English—a language with negligible rates of intrusive vowels in consonant clusters—Spanish HSs will produce a lower rate of intrusive vowels than Spanish speakers raised in monolingual environments. If intrusive vowels are present, their duration (normalized as tap-to-intrusive-vowel duration) is expected to be shorter in Spanish HSs compared to non-heritage Spanish native speakers

The second research question investigates whether the realization of intrusive vowels in /Cr/ and /r.C/ clusters, which depends on the production of precise consonantal gestures, varies as Spanish speakers develop adult-like gestural coordination skills.

- Given the mechanical complexity of producing consonant clusters and the observed slowdown in speech rate during their production, children likely experience an early

stage in which intrusive vowels are longer and more frequent until they develop a mature vocal tract with refined motor skills.

The third research question examines whether differences in the realization of intrusive vowels in /Cr/ and /r.C/ clusters between Spanish HSs and non-heritage Spanish native speakers vary with age.

- Attrition-like scenario: HSs are likely to show increased influence from English as they gain systematic exposure to it, potentially coupled with reduced exposure to Spanish starting around the onset of schooling. Therefore, results from adult HSs would reflect the adoption of English-like patterns of consonant cluster production. Under this hypothesis, we predict that younger children will produce intrusive vowels at a higher rate than older children, and older children will exhibit more intrusive vowels than adults. In addition, intrusive vowels are expected to be longer in younger children compared to older children, and longer in older children compared to adults.
- Protracted development-like scenario: Due to the influence of English, the development of Spanish-like coordination of consonantal gestures may be protracted, continuing across childhood and into adulthood. According to this hypothesis, due to intensive exposure to English, children might experience a period during primary school when consonant clusters are produced with English-like consonantal-gesture-coordination and laryngeal timing. In this scenario, younger child HSs are expected to produce a lower rate of intrusive vowels compared to older child HSs, with older child HSs showing a lower rate of intrusive vowels than adult HSs.

The fourth research question examines whether intrusive vowel production is influenced by phonetic context. The following hypotheses are based on previous research showing asymmetrical rates and duration of intrusive vocoids:

- Syllabic affiliation of the consonant cluster: We predict that, based on research showing a higher prevalence of intrusive vowels in /Cr/ clusters than in /r.C/ clusters

(Schmeiser, 2006), the rate and duration of intrusive vowels will be greater in /Cr/ clusters compared to /r.C/ clusters.

- Voicing: In line with previous studies (Colantoni and Steele, 2005; Schmeiser, 2006; Gibson et al., 2019), the rate and duration of intrusive vowels will be greater when the tap is flanked by a voiced obstruent compared to when it is flanked by a voiceless obstruent.
- Place of articulation: Consistent with previous research (Colantoni and Steele, 2005; Schmeiser, 2020), we predict that the rate and duration of intrusive vowels will be greater in heterorganic clusters compared to homorganic clusters. The longest duration is expected in clusters containing a dorsal obstruent.

Methodology

Participants

In total, 50 Spanish heritage speakers and 50 Spanish speakers raised in Mexico participated in the study. The Spanish heritage speaker (HS) group included 34 child Spanish HSs (16 F, 18 M; M = 9;1 years [SD = 24.15 months], age range = 5;3–11;11 years) and 16 adult Spanish HSs (13 F, 3 M; M = 20 years [SD = 2.06 years]) born in the US or who arrived in the US before the age of 3. Participants had at least one caregiver who emigrated from Mexico (two Mexican caregivers: N = 43; one caregiver from Mexico and one from an English-speaking country: N = 3; one caregiver from Mexico and one HS: N = 1). The HSs were exposed to Spanish since birth and to English before the age of 5 (children AoA: M = 2 years [SD = 1.57 years]; adults AoA: M = 3 years [SD = 1.44 years]).

The group of Spanish speakers raised in monolingual environments (i.e., SpanMonoSs hereafter) consisted of 34 child Spanish speakers (12 F, 22 M; M = 8;9 years [SD = 21.79 months], age range = 5 years–11;8 years) and 16 adult Spanish speakers (12 F, 4 M; M = 21.5 years [SD = 1.91 years]) recruited from Northern and Central Mexico.

INTRUSIVE VOWELS IN SPANISH HERITAGE SPEAKERS

While the SpanMonoSs also had some exposure to English at school, only participants with at least 70% of exposure to Spanish were included in the study.

To create comparable groups of children (i.e., HSs, SpanMonoSs), we divided the child participants at a cut-off point of eight years. This resulted in four groups of children: a group of younger child HSs ($N = 14$; $M = 6;11$ years [$SD = 8.89$ months], age range = 5;3–8;2 years), a group of younger child SpanMonoSs ($N = 14$; $M = 6;7$ years [$SD = 12.06$ months], age range = 5 years–7;9 years), a group of older child HSs ($N = 20$; $M = 10;6$ years [$SD = 14.45$ months], age range = 8 years–11;11 years) and a group of older child SpanMonoSs ($N = 20$; $M = 9;9$ years [$SD = 13.13$ months], age range = 8 years–11;8 years).

The adult participants and the caregivers of the child participants completed a language background questionnaire. This questionnaire gathered relative percentages of input and output in Spanish and English, along with the time spent with each family member. For the children, the amount of input was calculated as the percentage of time that each caregiver and sibling spoke in Spanish relative to English, weighted by the time each family member spent with the child. Instructional time was weighted by calculating the percentage of hours the child was instructed in Spanish. The amount of output was calculated as the relative percentage of Spanish spoken by the child to each family member, weighted by the time spent with each family member. For the adults, the questionnaire included data from three periods: primary school to middle school, high school, and university. The same weighting system used for the children was applied to average the weights, with the three periods weighted equally. Output was an averaged relative percentage of Spanish input and output for each speaker. For instance, a speaker with 20% Spanish input would have 80% English input. The results of the language background questionnaire show that child HSs had a mean input of 59% ($SD = 18\%$) and a mean output of 59% ($SD = 27\%$), while adult HSs had a mean input of 47% ($SD = 13\%$) and a mean output of 51% ($SD = 21\%$). The child SpanMonoSs had a mean input of 85% ($SD = 15\%$) and a mean output of 97% ($SD = 8\%$), while adult SpanMonoSs had a mean input of

84% (SD = 10%) and a mean output of 89% (SD = 13%).

Task

Semi-spontaneous speech was elicited using the wordless picture book *Frog, Where Are You?* (Mayer, 1969) (i.e., *frog story* henceforth). The story elicited /C_r/ and /ɾ.C/ clusters in which the obstruent varied in voicing (voiced: e.g., *grande* ['grɑ̃.de] 'big'; voiceless: e.g., ['kre.o] *creo* 'I think'), manner of articulation (fricative: e.g., ['fras.ko] *frasco* 'jar'; stop: e.g., ['troŋ.ko] *tronco* 'log'; or spirant: e.g., ['pje.ðra] *piedra* 'stone'), and place of articulation (labial: e.g., [pro.'βle.ma] *problema* 'problem'; coronal: ['o.tros] *otros* 'others'; dorsal: ['lo.ɣran] *logran* 'they achieve'). The *frog story* provides comparable data on the acquisition of rhotics, as it has been previously used to examine the rhotic system of Spanish HSs in semi-spontaneous speech (Henriksen, 2015; Repiso-Puigdelliura, 2022). The *frog stories* were collected as part of a larger study consisting of two production sessions, one in Spanish and one in English. Participants were recruited online during the COVID-19 pandemic, and oral consent was obtained during the first experimental session. Participants were compensated with a \$15 gift card. The experiments were proctored over Zoom by the author of this study. Before the start of the experimental session, participants were asked to record their speech using the smartphone app ShurePlus Motiv®, which records in an uncompressed audio format (48 kHz, 16 bits), with the device situated approximately 4 inches from the participant's mouth. In the case of child participants, a caregiver was asked to hold the phone approximately 4 inches from the child's mouth during the recording.

Data analysis

Segmentation and coding

Semi-spontaneous speech was annotated and segmented using Montreal Forced Aligner (McAuliffe et al., 2017). Instances of obstruent + tap (i.e., C1 + C2) were labeled in the data. In addition, C1 was annotated for the manner of articulation of C1 (i.e., stop, approximant, fricative), place of articulation of C1 (i.e., labial, coronal, dorsal), and voicing of C1 (i.e., voiced, voiceless). Intrusive vowels were manually segmented whenever formant

structure was observed in the spectrogram between the obstruent and the following tap (see Figure 1 for examples by SpanMonoSs and Figure 2 for examples by HSs) and consonantal sequences were coded with absence (0) or presence (1) of an IV.

The type of tap was categorized using criteria from previous studies (Amengual and Chamorro, 2016; Henriksen and Willis, 2010; Kim et al., 2020). **True taps** were identified by the presence of a visible constriction in the spectrogram, discontinuation of the formant structure, and a clear dip in intensity between the tap and the following vowel ($N = 933$, $\text{Freq} = 49.4\%$). **Approximant taps** were classified when there was a continuation of the formant structure between the tap and the following vowel and/or the preceding IV, without a clear dip in intensity ($N = 909$, $\text{Freq} = 48.12\%$). Segments in which an apicoalveolar tap could be perceived but not observed in the spectrogram were labeled as **perceptual taps** ($N = 5$, $\text{Freq} = 0.26\%$). Non-tap articulated segments were categorized as **other**, which included fricatives ($N = 24$, $\text{Freq} = 1.27\%$), laterals ($N = 10$, $\text{Freq} = 0.53\%$), retroflex articulations ($N = 1$, $\text{Freq} = 0.05\%$), affricates ($N = 2$, $\text{Freq} = 0.11\%$), glides ($N = 3$, $\text{Freq} = 0.16\%$), and elided taps ($N = 2$, $\text{Freq} = 0.11\%$). Two trained research assistants classified the tokens, and inter-rater agreement was evaluated using Cohen’s Kappa on a sample of 536 tokens. The resulting Kappa score of 0.79 ($z = 27.24$) indicates substantial agreement between the raters. To focus on intrusive vowel production, only tap segments articulated as **true taps** or **approximant taps** were included in the analysis, as previous research suggests that non-tap realizations may influence the occurrence of intrusive vowels (Colantoni and Steele, 2006).

To account for speech rate differences in the duration analysis, we normalized the duration of the intrusive vowel by computing a duration ratio relative to the tap segment (i.e., $\text{IV}/\text{tap duration ratio}$) (see Table 1). A duration ratio of 1 indicates that the intrusive vowel and the tap have the same duration, values greater than 1 indicate that the intrusive vowel is longer than the tap, and values less than 1 indicate that the intrusive vowel is shorter than the tap.

Statistical analyses

To analyze the data, we fitted mixed-effects models using *lme4* (Bates et al., 2014) and p-values were extracted using *lmerTest* (Kuznetsova et al., 2017). Following the recommendations of Matuschek et al. (2017, p. 312), we used standard model selection to find a middle ground between the Type I error rate and power. To do so, we fitted the maximal random effects structure and stepwise simplified the random slopes while comparing the Akaike Information Criterion across models. Given the research design, random intercepts for the participants and words were retained in the model. The estimated marginal means for the models were extracted using *emmeans* (Lenth et al., 2020). The results are reported with the package *easystats* (Lüdtke et al., 2022).

Results

Rate of realization of intrusive vowels

We fitted a logistic mixed model to predict the rate of intrusive vowel (i.e., presence coded as 1, absence coded as 0) with TYPE OF SPEAKER (i.e., SpanMonoSs, HSs), AGE GROUP (i.e., younger children, older children, adults) and their interaction, SYLLABIC AFFILIATION (i.e., /Cr/ and /r.C/), PLACE OF ARTICULATION (i.e., labial, coronal, dorsal), and VOICING of the flanking consonant (i.e., voiced, voiceless) (formula: `IV.bin type.speaker * age.group + syllabic.affiliation + voicing + poa`). The model included PARTICIPANT and LEMMA as random effects (formula: `list( 1 | Participant, 1 | Lemma)`). The model's total explanatory power is substantial (conditional $R^2 = 0.26$) and the part related to the fixed effects alone (marginal R^2) is of 0.08. Manner of articulation was excluded from the model to prevent an unbalanced factorial design, as two levels of this factor (spirants and fricatives) did not have data for the voiceless level of VOICING.

Within the model (see Table 2 and Figure 3), TYPE OF SPEAKER is statistically significant, indicating that SpanMonoSs ($M = 81.16\%$, $SE = 2.61\%$) have a higher probability of producing consonant clusters with intrusive vowels than HSs ($M = 67.91\%$, $SE = 4.01\%$). As for the effect of AGE GROUP, younger children ($M = 68.02\%$, $SE =$

5.61%) show a lower rate of intrusive vowels compared to adults ($M = 80.68\%$, $SE = 2.96\%$). With the younger children age group as the baseline, the older children age group does not show a simple effect. No significant interaction is found between AGE GROUP and TYPE OF SPEAKER.

No significant effects are found for AGE GROUP. As for the effects of the phonetic context, a significant effect of SYLLABIC AFFILIATION indicates that the probability of producing an intrusive vowel is higher in /Cr/ clusters ($M = 78.8\%$, $SE = 2.42\%$) than in /r.C/ clusters ($M = 71.04\%$, $SE = 2.98\%$). In addition, an effect of VOICING shows that intrusive vowels are more likely to be produced when the non-tap consonant is voiced ($M = 84.45\%$, $SE = 2.19\%$) than when it is voiceless ($M = 62.68\%$, $SE = 3.24\%$).

IV/tap duration ratio

We fitted a linear mixed model to predict log-transformed duration with TYPE OF SPEAKER (i.e., SpanMonoSs, HSs), AGE GROUP (i.e., younger children, older children, adults) and their interaction, SYLLABIC AFFILIATION (i.e., /Cr/ and /r.C/), PLACE OF ARTICULATION (i.e., labial, coronal, dorsal), and VOICING of the flanking consonant (i.e., voiced, voiceless) (formula: $\log(\text{duration ratio}) \sim \text{type.speaker} * \text{age.group} + \text{syllabic.affiliation} + \text{voicing} + \text{poa}$). The model included PARTICIPANT and LEMMA as random effects (formula: $\text{list}(1 \mid \text{Participant}, 1 \mid \text{Lemma})$). The model's total explanatory power is moderate (conditional $R^2 = 0.21$) and the part related to the fixed effects alone (marginal R^2) is 0.06. Recall that higher values indicate a longer duration of the intrusive vowel in relationship with the tap gesture.

Our findings (see Table 3 and Figure 4) show an effect of TYPE OF SPEAKER indicating that the IV/tap duration ratio is lower in the HSs ($M = 0.87$, $SE = 0.04$) compared to the SpanMonoSs ($M = 1.06$, $SE = 0.04$). Neither the effect of AGE GROUP nor the interaction between TYPE OF SPEAKER and AGE GROUP is significant. Concerning the effects of the phonetic context, a negative effect of VOICING indicates that the IV/tap duration ratio is shorter with a voiceless flanking obstruent ($M = 1.03$, $SE = 0.04$) than

with a voiced one ($M = 0.9$, $SE = 0.03$).

Discussion

The present study examined the production of intrusive vowels in /Cɾ/ and /ɾ.C/ clusters produced by child and adult Spanish HSs and non-heritage Spanish native speakers. We elicited semi-spontaneous speech to calculate the rate of intrusive vowel production and the duration ratio relative to the tap segment. In this section, we will discuss the results in relation to the research questions of the study.

The heritage language effect on intrusive vowel production

To return to the aims of this study, the first research question asked whether Spanish HSs produce a different rate and duration of intrusive vowels compared to non-heritage Spanish speakers. Using semi-spontaneous speech, we assessed differences in the presence and length of intrusive vowels among three groups of Spanish HSs and three groups of approximately age-matched Spanish speakers raised in monolingual environments.

Overall, the Spanish HSs produced intrusive vowels in more than half of the consonant clusters they produced ($M = 67.91\%$, $SE = 4.01\%$). Nonetheless, this rate is lower compared to that of the Spanish speakers raised in monolingual environments ($M = 81.16\%$, $SE = 2.61\%$). Moreover, when normalized by tap duration, the HSs produced shorter intrusive vowels than the non-heritage Spanish speakers.

To discuss the findings in terms of influence from the majority language, it is important to recall that it remains an open question whether intrusive vowels have an acoustic target and occur at the level of speech gesture planning, or whether they lack an acoustic target and arise at the execution level. According to the modular integrated sound system proposed by Natvig (2021), the former possibility is likely to correspond to the phonetics-phonology interface and the latter, to the phonetics module. In either case, however, the production of intrusive vowels is related to procedural knowledge, rather than non-contrastive knowledge, and is more likely to exhibit contextual and social variation (Natvig, 2021, p. 10). Therefore, the probabilistic nature of intrusive vowels, combined

with their lack of functional load in the phonology and their role as part of procedural knowledge—whether as a by-product of C1-C2 timing or as an acoustic target during speech planning—may contribute to their vulnerability to transfer within a multilingual system.

The next question arising from these results is whether HSs transfer intergestural timing or voicing patterns from their English into their Spanish speech systems. Our categorical data on the presence of intrusive vowels cannot determine whether the lower rate of intrusive vowels observed in HSs can be attributed solely to the transfer of English laryngeal demands into Spanish, resulting in a voiceless intergestural plateau interval—consistent with Tobin et al. (2019)’s findings, which demonstrate similar articulatory patterns between L2 Spanish speakers and native Spanish speakers, but an absence of intrusive vowels—or whether it also involves the transfer of English-to-Spanish intergestural timing, leading to the production of closed transitions in Spanish. However, our duration results showing a significant difference in the IV/tap duration ratio suggest that, when the laryngeal requirements to initiate or sustain voicing are met, and intrusive vowels are produced, the difference between the two groups of speakers (i.e., HSs and SpanMonoSs) may lie in the length of the intergestural plateau interval, and, therefore, in the degree of overlap between the two consonants. This, in turn, indicates that C1-C2 timing patterns from English can influence those of Spanish. This would be in line with previous studies showing crosslinguistic influence between timing patterns. Zsiga (2003), for example, investigated the transfer of timing patterns in L2 English and L2 Russian /VC#CV/, /VC#V/, and /V#CV/ sequences. While English shows a great deal of overlap in word-to-word articulation, Russian demonstrates little or no overlap between gestures. L2 English learners who were Russian native speakers showed transfer from the L1 timing patterns, producing little overlap in L2 English, as measured by the percentage of released final consonants and sequence-to-singleton duration. In contrast, L1 English-Russian L2 learners did not transfer the extensively overlapped consonants from

their L1 English into their L2 Russian. The Spanish HSs in this study may provide evidence for transfer in the high-overlap to low-overlap direction.

At the level of formalization, for the cases in which intrusive vowels are present but appear to have a shorter IV/tap duration ratio, one possibility is that, during the evaluation of Spanish clusters, English-like C1-C2 timing patterns may be active to some degree alongside Spanish C1-C2 timing patterns. At this point, the timing patterns of the two languages may interact, leading to either a full transfer of overlap (i.e., low overlap) or hybrid consonantal timing (i.e., mid-overlap). This suggests that Spanish HSs may temporarily broaden the range of intergestural timing patterns available in Spanish. Recall that, in cases where intrusive vowels are not produced, the crosslinguistic interaction during the intergestural plateau interval may also occur in the initiation of voicing in voiceless-voiced clusters or the maintenance of voicing in voiced-voiced clusters. To determine whether the observed acoustic differences between groups are due to variations in intergestural timing or differences in voicing during the plateau, future research should analyze articulatory data to investigate how intrusive vowels are realized in Spanish as a heritage language.

Furthermore, it is important to consider that the patterns observed in the HSs might be influenced by the nature of the linguistic input they receive. Specifically, we cannot rule out the possibility that the observed patterns arise from exposure to an environment with a low rate of intrusive vowels due to ongoing language change in the community. This possibility could be investigated by directly comparing the speech productions of HSs with those of their caregivers, or by including input-related variables in the models. Although research on this topic remains limited, recent studies have shown that the type of input children are exposed to significantly influences their speech production (Ramon-Casas et al., 2023; Stoehr et al., 2019). Therefore, future research should consider eliciting semi-spontaneous speech from both child HSs and their primary caregivers to better understand whether the observed differences result from individual

crosslinguistic transfer or from contact-induced changes at the community level.

Development of intrusive vowels in Spanish

The second research question asked whether child HSs produce intrusive vowels at a higher rate and with longer duration than adult HSs, possibly due to the development of speech motor skills and the fine-tuning of gestural coordination in the production of consonant clusters. Contrary to our prediction, our findings reveal that 5 5-to-8-year-olds produce a lower rate of intrusive vowels than adult speakers. As in the case of transfer, these results may be associated with the production of overlapped consonants or the devoicing of the intergestural plateau. If the former is the case, our results would align with a holistic account of phonological development in which children go through a stage of overlapped consonantal gestures that increasingly become independent as the motor control skills are refined (Nitttrouer, 1993; Noiray et al., 2018; Rubertus et al., 2015; Rubertus and Noiray, 2020). If, instead, the lack of intrusive vowels arises from devoiced C1-C2 transitions, we could attribute our findings to the mechanical complexity of achieving and maintaining voicing through the transition. However, the relatively small increase in the predicted likelihood between the two groups, along with the absence of significant differences between the 5-to-8-year-olds and the 8-to-11-year-olds, highlights the need for further research to explore this question.

The third research question pertained to the development scenarios posited by Polinsky (2018) and asked whether any differences in the realization of intrusive vowels between the two groups of speakers are mediated by age. Our results show that the slight differences between younger child HSs ($M = 57.55\%$, $SE = 9.63\%$) and older children ($M = 69.03\%$, $SE = 5.66\%$), as well as between older children and adults ($M = 74.68\%$, $SE = 4.98\%$), are not statistically significant at the population level. In other words, our findings do not support the hypothesis that the patterns generally observed in adult HSs result from attrition as HSs gain systematic exposure to the language. Similarly, our data do not support a hypothesis of protracted but continued development of intrusive vowels during

childhood. These results suggest that the null hypothesis cannot be rejected, aligning with the second developmental scenario proposed by Polinsky (2018). This scenario proposes a parallel change in the heritage language of both child and adult HSs relative to the baseline group. Given a true null effect at the population level, this would suggest that, unlike other aspects of phonetic and phonological development, the influence of the majority language may remain consistent throughout the HSs' lifespan.

Although there was no significant interaction between speaker type and age group, a closer examination of the data may reveal whether the means for each speaker group reflect similar distributions among children and adults. Figure 5 shows the distribution of participants based on the aggregated rates of intrusive vowel production. Overall, it is more common for child HSs to exhibit intrusive vowel realization rates below 50% ($N = 11$) compared to adult HSs ($N = 2$). The spread in the distributions narrows with age, with younger speakers showing the greatest variance ($\sigma^2 = 0.15$), followed by older children ($\sigma^2 = 0.04$) and adults ($\sigma^2 = 0.03$). This suggests that individual variation decreases with age. Notably, in the group of younger children, three participants display intrusive vowel rates below 25% (i.e., CHBILES53 [7;0 years old], CHBILES58 [7;6 years old], CHBILES45 [5;6 years old]). These children produce true taps at a relatively low rate (27.78%) compared to approximant taps (38.89%) and also show other variants, including perceptual taps (5.56%), laterals (16.67%), and fricatives (11.11%). This pattern indicates that mastery of tap gestures is necessary for the production of intrusive vowels, as proposed by Colantoni and Steele (2006), who explain that lower rates of intrusive vowels in L2 Spanish speakers result from their failure to achieve a target-like realization of tap gestures.

When comparing the proportion of intrusive vowels based on tap type (i.e., approximant tap, true tap), younger child HSs show an 18.99% difference, with true taps occurring at a higher rate ($M = 65.15\%$, $SE = 5.91\%$) compared to approximant taps ($M = 46.15\%$, $SE = 9.97\%$). In contrast, older child HSs, and adult HSs exhibit nearly identical proportions of intrusive vowels in true taps (older children: $M = 61.22\%$, $SE =$

4.95%; adults: $M = 67.14\%$, $SE = 3.25\%$) and approximant taps (older children: $M = 66.15\%$, $SE = 4.17\%$; adults: $M = 67.19\%$, $SE = 3.42\%$), with approximant taps slightly more likely to show intrusive vowels than true taps. It's important to note that the acoustic difference between tap types is based on the degree of constriction observed in the spectrogram, but acoustic analysis alone cannot determine the cause of a weakened constriction. Moreover, approximant taps are not unique to HSs. In an acoustic study on Madrid Spanish, Perry et al. (2024) found evidence of two types of tap variants using intensity differences, with more reduced variants appearing in fast speech and after stressed vowels (which are associated with a lower tone and jaw position), supporting an explanation of articulatory undershoot in these weakened taps. To further explore the relationship between the mastery of tap gestures and the production of intrusive vowels during acquisition, articulatory studies should examine the characteristics of approximant taps and their specific correlation with intrusive vowel production.

Effects of phonetic context on intrusive vowels

The fourth aim of this study was to examine the phonetic effects on the presence and duration of intrusive vowels. The effect of voicing was significant in both mixed-effects models, indicating that intrusive vowels are more likely to occur and are longer with a voiced flanking consonant than with a voiceless one. These findings are robust in the literature (Colantoni and Steele, 2005; Schmeiser, 2006; Gibson et al., 2019) and suggest that the specific laryngeal requirements for the production of intrusive vowels are more likely met in voiced clusters. For transitions involving an intrusive vocoid between voiced consonants in both $/Cr/$ and $/r.C/$ clusters, vocal fold vibration is maintained throughout the transition from a voiced C1 to a voiced C2. In clusters with a voiceless flanking consonant, producing an intrusive vocoid requires anticipatory voicing in $/Cr/$ clusters and precise coordination to open the glottis in time for the obstruent in $/r.C/$ clusters. These results also align with crosslinguistic evidence showing that the phonation of the obstruent consonant has a significant effect on the voicing of the intergestural plateau interval

(Poupplier et al., 2022).

Next, the effect of syllabic affiliation shows that a greater rate of intrusive vowels is produced in onset clusters (i.e., /Cr/) when compared to heterosyllabic sequences (i.e., /r.C/). Assuming a supralaryngeal-based account, we could refer our findings to the specific articulation of the tap segment. First, consider the findings in Gibson et al. (2019) showing less overlap for /Cr/ clusters than for /Cl/ clusters. The authors explain the asymmetry between consonant clusters by positing that the initial tongue body gesture of the tap may inhibit the overlap with C1 (Gibson et al., 2019, p. 470). In the present study, the anticipatory tongue body gesture would be produced between C1 and C2 in /Cr/ clusters, but before C1 in /r.C/ clusters. Because of these differences in timing, the articulatory demands of the tap are more likely to result in less overlap between the consonants in /Cr/ clusters compared to /r.C/ clusters.

Finally, we did not find significant effects for a labial-coronal-dorsal place of articulation. Recall that we included both /Cr/ and /r.C/ clusters in our models, which may have masked potential differences in articulation between back-to-front and front-to-back sequences. However, the average percentages of intrusive vowel production split by syllabic affiliation in the clusters suggest that intrusive vowels are produced at a higher rate in back-to-front sequences (/Cr/: M = 76.77%, SE = 3.40%, /r.C/: M = 84.17%, SE = 3.66%) than in front-to-back sequences (/Cr/: M = 64.03%, SE = 3.18%, /r.C/: M = 65.85%, SE = 3.71%).

Concluding remarks

This study investigated the production of intrusive vowels in /Cr/ and /r.C/ consonant clusters in both child and adult HSs, comparing their speech to that of Spanish speakers raised in monolingual environments. The goal was to examine the potential influence of the majority language, American English, on the heritage language, Spanish, and to determine whether this influence is associated with a shift in language dominance from Spanish to English between primary school years and young adulthood. We compared

INTRUSIVE VOWELS IN SPANISH HERITAGE SPEAKERS

three groups of Spanish heritage speakers: younger children (ages 5 to 8), older children (ages 8 to 11), and young adults, each age-matched with non-heritage Spanish native speakers.

Our findings indicate that Spanish HSs produce intrusive vowels less frequently compared to Spanish speakers raised in monolingual environments. Moreover, when normalized by tap duration, the relative duration of intrusive vowels is shorter in Spanish HSs than in their non-heritage counterparts. For the Spanish HSs, we propose that English-specific coordination of laryngeal requirements and C1-C2 timing patterns may influence those in Spanish, potentially leading to greater consonantal overlap or periods of voicelessness during the intergestural plateau interval. It is important to note that the study findings cannot rule out the possibility that the lower frequency and shorter duration of intrusive vowels are due to the influence of English in the input that HSs receive from their English-influenced environment (e.g., family, peers, teachers).

Furthermore, we found that 5-to-8-year-olds produce fewer intrusive vowels than adults, aligning with a view in which segment-to-segment articulation shows greater overlap during early language development and persists until motor skills fully mature. An alternative explanation that we have considered is that children produce devoiced intergestural plateaux due to the complexity of initiating or maintaining voicing throughout the transition.

However, the absence of an interaction between age group and type of speaker suggests that a potential shift in language dominance from the majority language to the heritage language does not lead to attrition or protracted development. Instead, our results are consistent with the null hypothesis. Although our null results could be due to a small sample size, we tentatively suggest that parallel change may occur in both child and adult HSs, indicating that the potential influence of English persists throughout the HSs' lifespan.

Finally, in terms of phonetic effects, our data align with previous studies, showing that intrusive vowels are produced more frequently in groups with voiceless non-tap

consonants than in those with voiceless consonants (Colantoni and Steele, 2005; Schmeiser, 2006; Gibson et al., 2019).

We recognize the need for further research on speech gesture coordination during language development to better understand the differences in the rate of intrusive vowel realization and its duration relative to the tap between HSs and Spanish speakers raised in monolingual environments. Specifically, future studies should investigate whether these differences stem from interactions in intergestural timing or from crosslinguistic influence in the degree of voicing of the intergestural plateau.

References

- Alwan, A., Narayanan, S., and Haker, K. (1997). Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part II. The rhotics. *The Journal of the Acoustical Society of America*, 101(2).
- Amengual, M. and Chamorro, P. (2016). The Effects of Language Dominance in the Perception and Production of the Galician Mid Vowel Contrasts. *Phonetica*, 72(4).
- Au, T. K. F., Knightly, L. M., Jun, S. A., and Oh, J. S. (2002). Overhearing a language during childhood. *Psychological Science*, 13(3):238–243.
- Au, T. K. f., Oh, J. S., Knightly, L. M., Jun, S. A., and Romo, L. F. (2008). Salvaging a childhood language. *Journal of Memory and Language*, 58(4):998–1011.
- Barbier, G., Perrier, P., Payan, Y., Tiede, M. K., Gerber, S., Perkell, J. S., and Ménard, L. (2020). What anticipatory coarticulation in children tells us about speech motor control maturity. *PLoS ONE*, 15(4):e0231484.
- Bates, D., Maechler, M., Bolker, B., and Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7, <http://CRAN.R-project.org/package=lme4>. *R package version*.
- Bellik, J. (2018). An acoustic study of vowel intrusion in Turkish onset clusters. *Laboratory Phonology*, 9(1):1–23.
- Benmamoun, E., Montrul, S., and Polinsky, M. (2013). Defining an "ideal" heritage speaker: Theoretical and methodological challenges Reply to peer commentaries. *Theoretical Linguistics*, 39(3-4):259–294.
- Beristain, A. (2023). An Airflow Analysis of Spanish and English Anticipatory Vowel Nasalization among Heritage Bilinguals. *Languages*, 8(3).

INTRUSIVE VOWELS IN SPANISH HERITAGE SPEAKERS

- Blecua, B. (2001). *Las vibrantes del español: manifestaciones acústicas y procesos fonéticos*. PhD thesis, Universitat Autònoma de Barcelona, Bellaterra.
- Booij, G. (1995). *The Phonology of Dutch*, volume 72. Oxford University Press.
- Bradley, T. G. (2002). Gestural Timing and Derived Environment Effects in Norwegian Clusters. In Mikkelsen, L. and Potts, C., editors, *West Coast Conference on Formal Linguistics 21 Proceedings*, number May, pages 43–56. Cascadia Press.
- Bradley, T. G. (2004). Gestural Timing and Rhotic Variation in Spanish Codas. In Face, T. L., editor, *Laboratory Approaches to Spanish Phonology*, pages 195–220. Mouton de Gruyter, Berlin.
- Bradley, T. G. (2007). Spanish complex onsets and the phonetics–phonology interface.
- Buchwald, A. B., Rapp, B., and Stone, M. (2007). Insertion of discrete phonological units: An articulatory and acoustic investigation of aphasic speech. *Language and Cognitive Processes*, 22(6):910–948.
- Byrd, D. (1996). Influences on articulatory timing in consonant sequences. *Journal of Phonetics*, 24(2):209–244.
- Catford, J. C. (1977). *Fundamental problems in phonetics*. Indiana University Press, Bloomington and London.
- Chang, C. (2021). Phonetics and Phonology of Heritage Languages. In *The Cambridge Handbook of Heritage Languages and Linguistics*.
- Chitoran, I., Goldstein, L., and Byrd, D. (2002). Gestural overlap and recoverability: Articulatory evidence from Georgian. In *Laboratory Phonology*, volume 7, pages 419–447.
- Colantoni, L. and Steele, J. (2005). Phonetically-driven epenthesis asymmetries in French and Spanish obstruent-liquid clusters. In Gess, R. and Rubin, E., editors, *34th*

Linguistics Symposium on Romance Languages, page 77, Amsterdam/Philadelphia. John Benjamins.

Colantoni, L. and Steele, J. (2006). Native-Like Attainment in the L2 Acquisition of Spanish Stop-Liquid Clusters. In *Selected Proceedings of the 7th Conference on the Acquisition of Spanish and Portuguese as First and Second Languages*, pages 59–73. Cascadilla Proceedings Project.

Colantoni, L. and Steele, J. (2007). Voicing-dependent cluster simplification asymmetries in Spanish and French. In Prieto, P., Solé, M.-J., and Mascaró, J., editors, *Segmental and prosodic issues in Romance linguistics*, pages 109–129. John Benjamins, Amsterdam/Philadelphia.

Delattre, P. and Freeman, D. C. (1968). A dialect study of american R'S by x-ray motion picture. *Linguistics*, 6(44):29–68.

Gibson, M., Bunta, F., Johnson, C., and Huárriz, M. (2022). The production of English syllable-level timing patterns by bilingual English- and Spanish-speaking children with cochlear implants and their peers with normal hearing. *Journal of Phonetics*, 95.

Gibson, M., Goodin-Mayeda, E., Bunta, F., and Hernandez, A. (2018). The acquisition of syllable-level timing contrasts by English- and Spanish-speaking bilingual children with normal hearing and English- and Spanish-speaking bilingual children with cochlear implants. *Journal of Phonetics*, 71:98–112.

Gibson, M., Muxika, O., and Pouplier, M. (2021). Dissecting the aero-acoustic parameters of open articulatory transitions. In *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, volume 3, pages 1698–1702. International Speech Communication Association.

Gibson, M., Sotiropoulou, S., Tobin, S., and Gafos, A. I. (2019). Temporal Aspects of Word Initial Single Consonants and Consonants in Clusters in Spanish. *Phonetica*, 76(6).

INTRUSIVE VOWELS IN SPANISH HERITAGE SPEAKERS

- Gili Gaya, S. (1921). La r simple en la pronunciación española. *Revista de Filología Española*, VIII:271–280.
- Hall, N. (2006). Cross-linguistic patterns of vowel intrusion.
- Hardcastle, W. J. and Roach, P. (1979). An instrumental investigation of coarticulation in stop consonant sequences. *Current issues in the phonetic sciences*, 9:531–540.
- Henriksen, N. (2015). Acoustic analysis of the rhotic contrast in Chicagoland Spanish. *Linguistic Approaches to Bilingualism*, 5(3):285–321.
- Henriksen, N. and Willis, E. W. (2010). Acoustic characterization of phonemic trill production in Jerezano Andalusian Spanish. In *Selected Proceedings of the 4th Conference on Laboratory Approaches to Spanish Phonology*, pages 115–127, Somerville, MA. Cascadilla Proceedings.
- Kahn, D. (1976). *Syllable-based generalizations in English phonology*. PhD thesis, Massachusetts Institute of Technology.
- Kim, J. Y. (2024). The Phonetics and Phonology of Heritage Language Speakers. In Amengual, M., editor, *The Cambridge Handbook of Bilingual Phonetics and Phonology*, Cambridge Handbooks in Language and Linguistics, pages 560–583. Cambridge University Press.
- Kim, J. Y., Faytak, M., Repiso Puigdelliura, G., and Maufray, E. (2020). Articulation of non-normative variants of L1 and L2 Spanish trill /r/. *The Journal of the Acoustical Society of America*, 148(4).
- Kochetkova, U. (2019). Svarabhakti vowel occurrence and duration in rhotic clusters in French lyric singing. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 11658 LNAI.

INTRUSIVE VOWELS IN SPANISH HERITAGE SPEAKERS

- Kupisch, T., Kolb, N., Rodina, Y., and Urek, O. (2021). Foreign Accent in Pre- and Primary School Heritage Bilinguals. *Languages 2021, Vol. 6, Page 96*, 6(2):96.
- Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models . *Journal of Statistical Software*, 82(13).
- Ladefoged, P. and Maddieson, I. (1998). The Sounds of World’s Languages. *Language*, pages 73–77.
- Lancien, M. and Côté, M.-H. (2019). Intrusive vowels preceding /R/ in Quebec French. In Calhoun, S., Escudero, P., Tabain, M., and Warren, P., editors, *Proceedings of the 19th International Congress of Phonetic Sciences*, pages 3473–3477, Melbourne.
- Lenth, R., Singmann, H., Love, J., Buerkner, P., and Herve, M. (2020). Package ‘ emmeans ’ R topics documented :. *R package version 1.15-15*, 34(1).
- Lüdecke, D., Ben-Shachar, M. S., Patil, I., Wiernik, B. M., Bacher, E., Thériault, R., and Makowski, D. (2022). easystats: Framework for Easy Statistical Modeling, Visualization, and Reporting. *CRAN*.
- Malmberg, B. (1965). Estudios de fonética hispánica. Technical report, CSIC, Madrid.
- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., and Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, 94:305–315.
- Mayer, M. (1969). *Frog Where Are You?* Dial Press, New York.
- McAuliffe, M., Socolof, M., Mihuc, S., and Wagner, Michael Sonderegger, M. (2017). Montreal Forced Aligner: trainable text-speech alignment using Kaldi. In *Proceedings of the 18th Conference of the International Speech Communication Association*.
- Menke, M. R. (2018). Development of Spanish rhotics in Spanish-English bilingual children in the United States. *Journal of Child Language*, 45:788–806.

INTRUSIVE VOWELS IN SPANISH HERITAGE SPEAKERS

- Montrul, S. (2023). Heritage Languages: Language Acquired, Language Lost, Language Regained. *Annual Review of Linguistics*, 9(1):399–418.
- Natvig, D. (2021). Modeling heritage language phonetics and phonology: Toward an integrated multilingual sound system. *Languages*, 6(4):209.
- Nittrouer, S. (1993). The emergence of mature gestural patterns is not uniform: Evidence from an acoustic study. *Journal of Speech and Hearing Research*, 36(5):959–972.
- Nittrouer, S., Studdert-Kennedy, M., and McGowan, R. S. (1989). The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *Journal of Speech and Hearing Research*, 32(1).
- Noiray, A., Abakarova, D., Rubertus, E., Krüger, S., and Tiede, M. (2018). How do children organize their speech in the first years of life? Insight from ultrasound imaging. *Journal of Speech, Language, and Hearing Research*, 61(6):1355–1368.
- Oh, J. S., Jun, S. A., Knightly, L. M., and Au, T. K. F. (2003). Holding on to childhood language memory. *Cognition*, 86(3):B53–B64.
- Perry, S. J., Kelley, M. C., and Tucker, B. V. (2024). Documenting and modeling the acoustic variability of intervocalic alveolar taps in conversational Peninsular Spanish. *The Journal of the Acoustical Society of America*, 155(1):294–305.
- Polinsky, M. (2018). *Heritage Languages and their speakers*, volume 159. Cambridge University Press.
- Poupplier, M., Pastätter, M., Hoole, P., Marin, S., Chitoran, I., Lentz, T. O., and Kochetov, A. (2022). Language and cluster-specific effects in the timing of onset consonant sequences in seven languages. *Journal of Phonetics*, 93.
- Proctor, M. (2011). Towards a gestural characterization of liquids: Evidence from Spanish and Russian. *Laboratory Phonology*, 2(2):451–485.

INTRUSIVE VOWELS IN SPANISH HERITAGE SPEAKERS

- Ramírez, C. J. (2006). Acoustic and Perceptual Characterization of the Epenthetic Vowel between the Clusters Formed by Consonant + Liquid in Spanish. In *Selected Proceedings of the 2nd Conference on Laboratory Approaches to Spanish Phonetics and Phonology*, pages 48–61, Somerville, MA. Cascadilla Proceedings Project.
- Ramon-Casas, M., Cortés, S., Benet, A., Lleó, C., and Bosch, L. (2023). Connecting perception and production in early Catalan-Spanish bilingual children: language dominance and quality of input effects. *Journal of Child Language*, 50(1).
- Rao, R. (2019). The phonological system of adult heritage speakers of Spanish in the United States. In Martínez Gil, F. and Colina, S., editors, *The Routledge Handbook of Spanish Phonology*, pages 439–451. Routledge.
- Rao, R. (2024). *The Phonetics and Phonology of Heritage Languages*. Cambridge University Press.
- Recasens, D. (2018). *The production of consonant clusters: Implications for phonology and sound change*, volume 26. Walter de Gruyter GmbH & Co KG.
- Recasens, D., Fontdevila, J., Pallarès, M. D., and Solanas, A. (1993). An electropalatographic study of stop consonant clusters. *Speech Communication*, 12(4):335–355.
- Repiso-Puigdelliura, G. (2022). *The Development of Cross-linguistic Transfer: The Case of Word-External Repairs of Empty Onsets in Spanish Heritage Speakers*. PhD thesis, University of California, Los Angeles, Los Angeles.
- Repiso-Puigdelliura, G. (2023). Glottalizing at word junctures: Exploring bidirectional transfer in child and adult Spanish heritage speakers. *Bilingualism: Language and Cognition*, pages 1–13.

INTRUSIVE VOWELS IN SPANISH HERITAGE SPEAKERS

- Repiso-Puigdelliura, G. and Kim, J. Y. (2021). The missing link in Spanish heritage trill production. *Bilingualism: Language and Cognition*, 24(3):454–466.
- Romero, J. (2008). Gestural Timing in the Perception of Spanish r+C Clusters. In *Selected proceedings of the 3rd Conference on Laboratory Approaches to Spanish Phonology*, pages 59–71, Somerville, MA. Cascadilla Proceedings Project,.
- Rubertus, E., Abakarova, D., Tiede, M., Ries, J., and Noiray, A. (2015). Development of Coarticulation in German Children: Articulatory Locus Equations Contact Information. Technical report.
- Rubertus, E. and Noiray, A. (2020). Vocalic activation width decreases across childhood: Evidence from carryover coarticulation. *Laboratory Phonology*, 11(1).
- Savu, C.-F. (2013). Another look at the structure of [ɹ]: Constricted intervals and vocalic elements. In Spreafico, L. and Vietti, A., editors, *Rhotics. New Data and Perspectives*, pages 145–158. Bozen-Bolzano University Press, Free University of Bozen-Bolzano.
- Schmeiser, B. (2006). *On the durational variability of svarabhakti vowels in spanish consonant clusters*. PhD thesis, University of California, Davis.
- Schmeiser, B. (2009). An Acoustic Analysis of Intrusive Vowels in Guatemalan Spanish /rC/ Clusters. In *Proceedings of the 32nd Annual PennLinguistics Colloquium*, volume 15, pages 193–202. University of Pennsylvania Working Papers in Linguistics.
- Schmeiser, B. (2020). Prosodic and segmental effects on the durational variability of svarabhakti vowels in Spanish /CR/ clusters. In *Proceedings of the International Conference on Speech Prosody*, volume 2020-May, pages 16–20, Tokyo.
- Stoehr, A., Benders, T., van Hell, J. G., and Fikkert, P. (2019). Bilingual Preschoolers’ Speech is Associated with Non-Native Maternal Language Input. *Language Learning and Development*, 15(1):75–100.

INTRUSIVE VOWELS IN SPANISH HERITAGE SPEAKERS

- Tiede, M. K., Boyce, S. E., Holland, C. K., and Choe, K. A. (2004). A new taxonomy of American English /r/ using MRI and ultrasound. *The Journal of the Acoustical Society of America*, 115(5):2633–2634.
- Tobin, S. J., Gibson, M., Sotiropoulou, S., and Gafos, A. (2019). Articulatory Coordination in L2-Speakers of Spanish. In Calhoun, S., Escudero, P., Tabain, M., and Warren, P., editors, *Proceedings of the 19th International Congress of Phonetic Sciences*, number August, pages 3353–3357, Melbourne. Australasian Speech Science and Technology Association Inc., and International Phonetic Association.
- Zsiga, E. C. (2000). Phonetic alignment constraints: Consonant overlap and palatalization in English and Russian. *Journal of Phonetics*, 28(1):69–102.
- Zsiga, E. C. (2003). Articulatory timing in a second language. *Studies in Second Language Acquisition*, 25(3).

Table 1

Summary of mean duration of intrusive vowels (IV) in ms, tap duration in ms, and IV/tap duration ratio by group and type of speaker

Group	Mean IV (SD)	Mean tap (SD)	M dur. ratio (SD)
young. child. HSs	28.68 (14.42)	36.72 (18.78)	0.92 (0.50)
young. child. SpanMonoSs	35.39 (20.73)	32.21 (15.90)	1.31 (1.01)
old. child. HSs	29.15 (13.50)	31.10 (13.94)	1.05 (0.55)
old. child. SpanMonoSs	26.78 (9.59)	24.57 (10.72)	1.27 (0.68)
adult HSs	24.56 (9.55)	28.13 (9.79)	0.96 (0.46)
adult SpanMonoSs	24.57 (7.83)	25.29 (8.07)	1.05 (0.45)

Table 2*Generalized logistic regression: rate of intrusive vowel production*

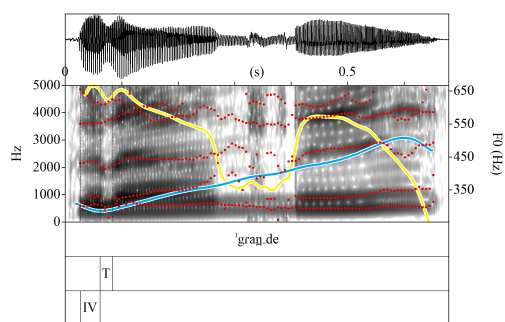
Model Coefficients				
Term	Estimate	Std. Error	z value	p-value
Intercept	1.105	0.131	8.461	<0.001 ***
Type speaker: HSs	-0.711	0.240	-2.956	0.0031 **
Age group: old. children	0.376	0.310	1.213	0.2252
Age group: adults	0.675	0.312	2.162	0.0306 *
Syll. affiliation: rC	-0.415	0.125	-3.327	<0.001 ***
Voicing:voiced	1.173	0.161	7.282	<0.001 ***
PoA: coronal	0.166	0.156	1.062	0.2883
PoA: dorsal	0.045	0.187	0.244	0.8074
HSs * old. children	0.222	0.622	0.358	0.7206
HSs * adults	0.172	0.623	0.275	0.7830

*Performance Metrics:*AIC: 2021.56, BIC: 2087.77, Log-Likelihood: -998.78

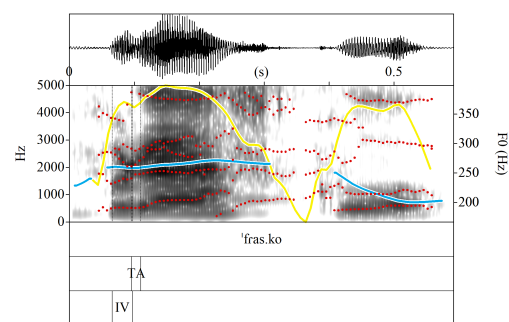
Table 3*Mixed effects linear model: IV/tap duration ratio*

Model Coefficients					
Term	Estimate	Std. Error	df	t value	p-value
Intercept	-0.073	0.061	120.060	-1.190	0.2366
Type speaker: HSs	-0.303	0.116	104.098	-2.601	0.0107 *
Age group: old. children	-0.002	0.069	84.977	-0.032	0.9748
Age group: adults	0.100	0.069	98.475	1.458	0.1481
Syll. affiliation: rC	-0.025	0.039	53.429	-0.639	0.5256
PoA: coronal	-0.008	0.051	54.174	-0.162	0.8718
PoA: dorsal	0.028	0.047	62.038	0.597	0.5527
Voicing: voiced	0.133	0.042	73.557	3.187	0.0021 **
HSs * old. children	0.186	0.137	84.341	1.363	0.1765
HSs * adults	0.114	0.138	98.617	0.828	0.4099
<i>Performance Metrics: AIC: 1717.51, BIC: 1784.81, Log-Likelihood: -845.76</i>					

INTRUSIVE VOWELS IN SPANISH HERITAGE SPEAKERS



(a) Segmentation of intrusive vowel
(i.e., IV) in the word grande ('big')

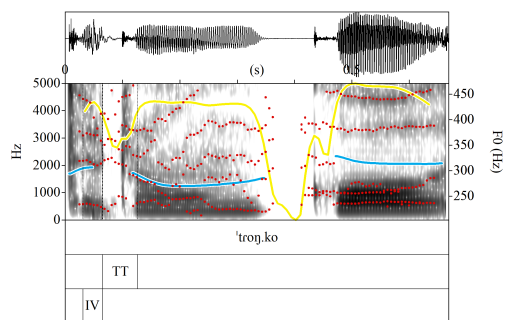


(b) Segmentation of intrusive vowel
(i.e., IV) in the word frasco ('jar')

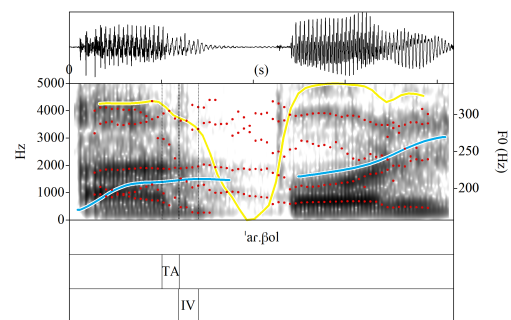
Figure 1

Segmentation of intrusive vowels produced by child Spanish-speakers raised in monolingual environments (an 8;11-years-old-child in Figure 1a and a 9;11-years-old-child in Figure 1b). In the textgrid, tier 1 contains the word transcription in IPA, tier 2 contains the tap annotation (TT for true tap, AT for approximant tap, and tier 3 contains the segmentation of the intrusive vowel.

INTRUSIVE VOWELS IN SPANISH HERITAGE SPEAKERS



(a) Tap labeled as true tap in tronco ('log')



(b) Tap labeled as approximant tap in árbol ('tree')

Figure 2

Labeling of tap segments produced by child HSs (a 9;1-years-old-child in Figure 2a and a 10;1-years-old-child in Figure 2b). Tier 1 contains the word in IPA, tier 2 contains the coding of the tap (TT for true tap and TA for approximant tap), and tier 3 contains the segmentation of the intrusive vowel.

INTRUSIVE VOWELS IN SPANISH HERITAGE SPEAKERS

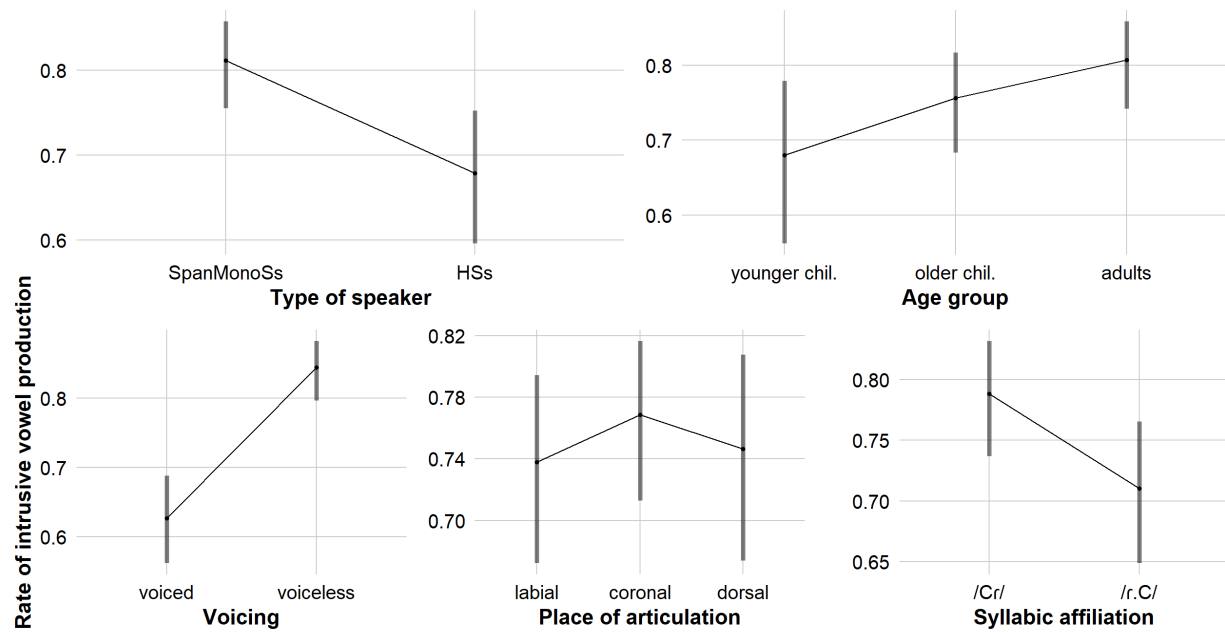


Figure 3

Estimate response rates for fixed covariates in the mixed-effects logistic regression model.

INTRUSIVE VOWELS IN SPANISH HERITAGE SPEAKERS

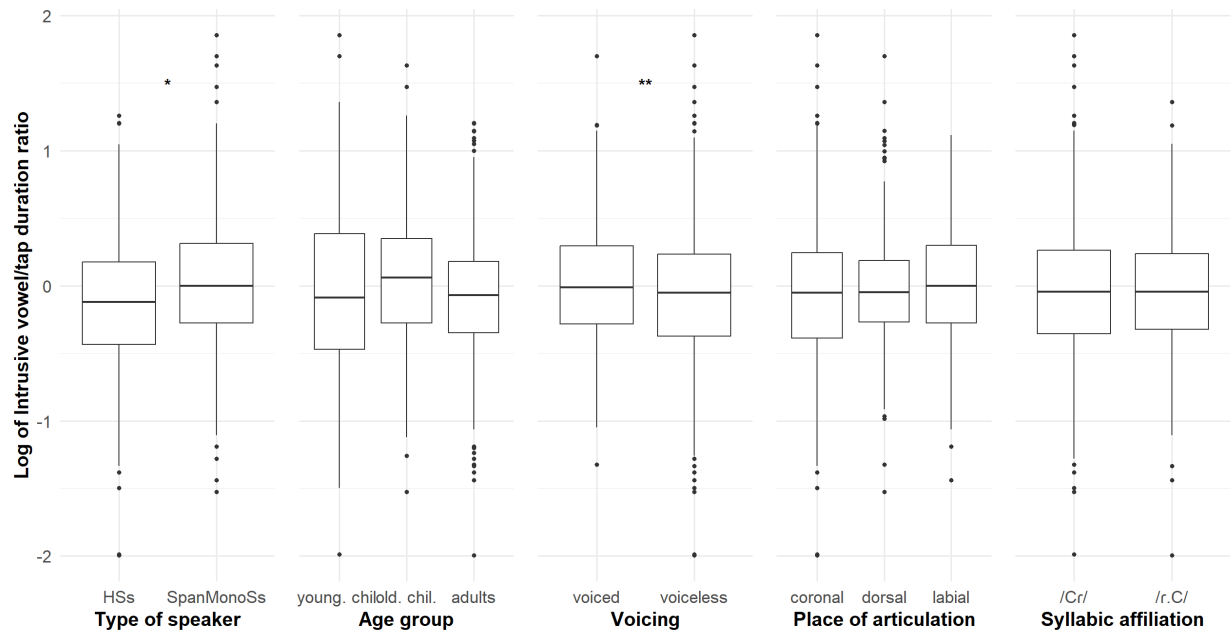


Figure 4

Log transformed IV/tap duration ratio, split by type of speaker, age group, voicing, place of articulation, and syllabic affiliation. Higher values indicate a greater intrusive vowel duration relative to the tap duration.

INTRUSIVE VOWELS IN SPANISH HERITAGE SPEAKERS

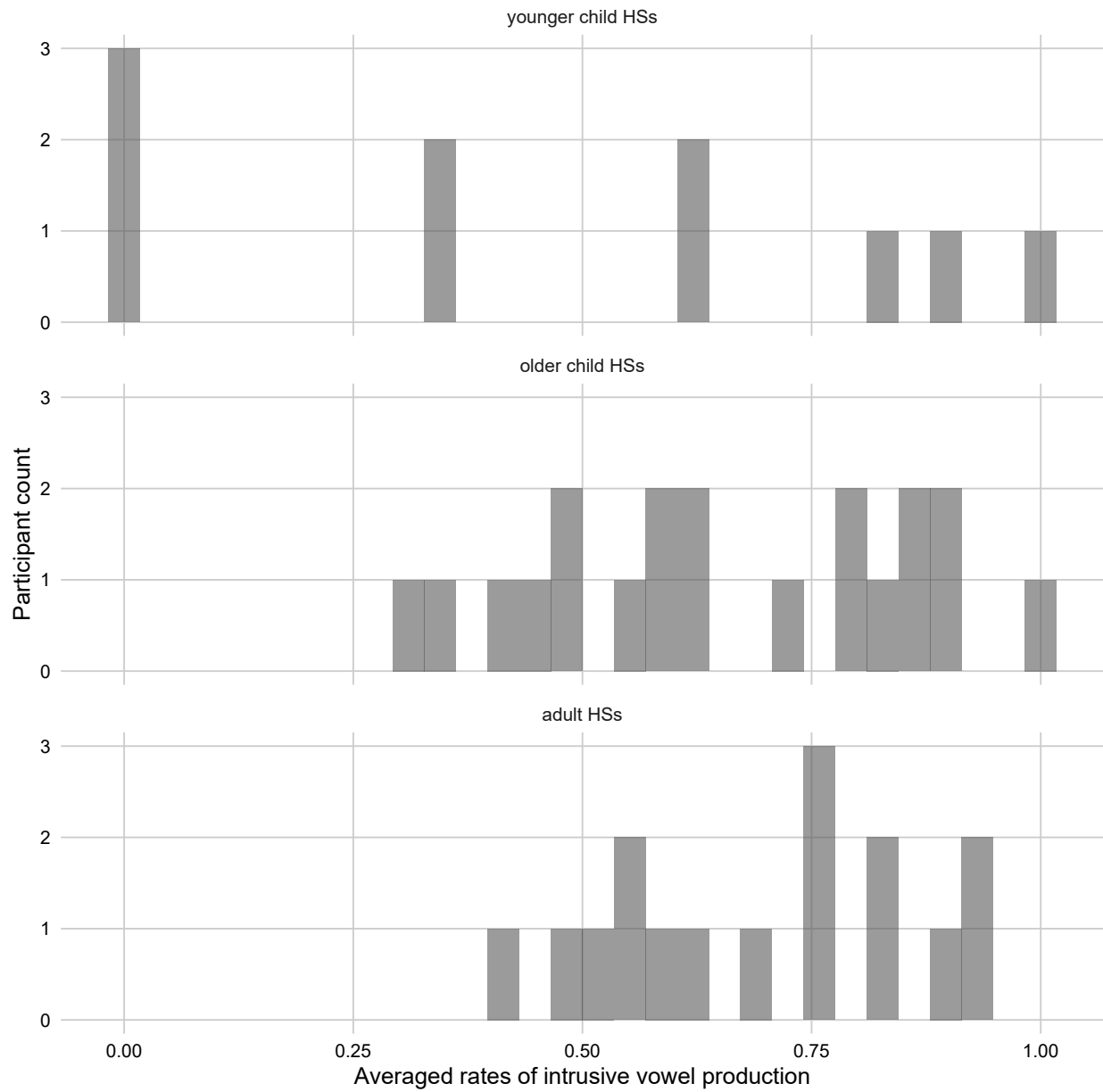


Figure 5

Distribution of aggregated percentages across tokens per speaker