2018

An Exploratory Study of Heritage Spanish Rhotics: Addressing Methodological Challenges of Heritage Language Phonetics Research

Elizabeth M. Kissling
University of Richmond, ekisslin@richmond.edu

This is a pre-publication author manuscript of the final, published article.

Recommended Citation

Kissling, Elizabeth. "An Exploratory Study of Heritage Spanish Rhotics: Addressing Methodological Challenges of Heritage Language Phonetics Research." *Heritage Language Journal* 15, no. 1 (2018): 25-70.

An Exploratory Study of Heritage Spanish Rhotics: Addressing Methodological Challenges of Heritage Language Phonetics Research

Elizabeth M. Kissling
University of Richmond

ABSTRACT
When speaking their heritage language, heritage speakers typically sound much like other "native speakers." However, recent studies have found that heritage speakers (HSs) are highly variable and produce a range of more and less "native-like" phonetic features. In an effort to stimulate productive new research in this area, this article addresses some of the methodological challenges of heritage language phonetics research, namely dealing with high variability and identifying the best predictors of that variability. A study on heritage Spanish rhotics is presented to elucidate those methodological challenges. The study took an exploratory, bottom-up approach to analyzing the rhotics produced by speakers of central Mexican and Salvadoran Spanish with different language profiles: HSs, traditional native speakers, long-term immigrants, and second language learners. The results suggested that overall between-group comparisons of means based on isolated acoustic features could be insufficiently informative. The study also evaluated the contribution of various linguistic (e.g., proficiency and use) and extralinguistic (e.g., cultural and ethnic identity) factors for identifying more homogeneous subgroups of HSs and found that the latter were useful for predicting phonetic variation.

Keywords: *heritage speakers, methods, phonetics, rhotics, Spanish, variation*

INTRODUCTION

As evidenced by the growth of this specialized journal over the past decade, research on heritage speakers (HSs) is bourgeoning. Yet definitions of HSs vary. Some researchers take a broad perspective informed by sociolinguistics, which leads them to focus on learners' historical and personal connection to the language (see Polinsky & Kagan, 2007). Others, including many foreign language educators and SLA researchers, take a psycholinguistic approach and focus more narrowly on language proficiency (see Leeman, 2015). SLA researchers studying heritage languages in the U.S. typically identify HSs based on both biographical information and language proficiency, defining HSs as those who are exposed to a home language from birth in their family environment and who maintain at least some language proficiency, even if only passive aural comprehension. Though their backgrounds can differ dramatically, HSs as defined in this way--those who learned an ethnolinguistic minority home language--typically share some common traits. The extant research suggests that HSs' morphology, syntax, and semantics can diverge from monolingual first language (L1) norms to varying degrees. The area of heritage grammar that seems least divergent is phonology. Many HSs are perceived as having very good accents, generally much more target-like than what most adult second language (L2) learners achieve (Montrul, 2010). Conversely, other HSs perceive and produce sounds differently than most native speakers (Rao & Ronquest, 2015).

Recently a special issue of this journal (Rao, 2016b) reported on investigations of a variety of phonetic targets (consonants, vowels, prosody, and global accent) in a variety of contexts and with seven heritage languages. This collection of research represents an important step forward to understanding heritage phonetics, which is relatively unexplored (Rao, 2016b). Yet the methodologies employed in those and other previous studies exhibit some challenges that should be addressed to ensure productivity of the field. This article explores new methodological approaches that could be useful for research in heritage phonetics, using an empirical study on Spanish rhotics as an illustrative case-in-point.

The structure of the article is as follows. First, the introduction examines challenges of the current methodologies with regard to addressing variability in the HS population. Research linking accent and identity is briefly reviewed to motivate consideration of two potentially important extralinguistic predictor variables that have not been considered in heritage phonetics research: cultural and ethnic identity. Then a brief introduction to heritage Spanish phonetics and acquisition of rhotics provides background for the exploratory empirical study on heritage Spanish rhotics. The study is intended to illustrate how making different methodological choices would lead to different conclusions about heritage phonetics. Finally, the discussion offers suggestions for how future research might address our current methodological challenges.

VARIABILITY IN HERITAGE PHONETICS

As evidenced by the special issue (Rao, 2016b), researchers working in this area have recruited participants with a range of HS profiles for their studies, such as those born in country or having immigrated, belonging to different generations and with different lengths of residence, and with varying amounts of exposure to the heritage language and varieties of that language. Though all these participants fit the commonly accepted definition of a HS (as someone exposed to an ethnolinguistic minority home language from birth in their family environment and who maintains at least some language proficiency), clearly they do not form a homogeneous group in terms of language experience. Casting such a wide net may be a practical necessity given the small numbers of HSs in some contexts, and because research in this area is so scant, any new study is a welcome contribution. Yet the methodological convention of lumping together HSs with different backgrounds may limit our ability to elucidate patterns in HSs' variability.

The HS population is often described as extremely heterogeneous in their language abilities at all levels of linguistic analysis. Polinsky and Kagan argue that the "illusion of endless variation comes from our neglecting to look closer and recognize [that] … their speaking abilities fall within a continuum … reminiscent of what has been proposed in Creole studies" (2007, p. 371). Chang and Yao discuss HS heterogeneity at length in their study of heritage Mandarin prosody, included in the special issue (2016). Their HSs were much more variable than the other participant groups. Raters also found them more difficult to classify demographically (i.e., as American-born Chinese, native Chinese, or non-Chinese American speakers). Chang and Yao argued:

> Although one could argue that HL [heritage language] speakers' demographic ambiguity is merely an artifact of the way the HL group was constituted (which resulted in the inclusion of a wider range of experience with the target language than in the NM and L2 groups), in many ways, this is exactly the point: the linguistic heterogeneity of HL Mandarin speakers makes it difficult to associate this population with a well-defined perceptual category. While some HL speakers may sound like native speakers, others sound more like L2 learners (and yet others, somewhere in between). (2016, p. 155)

Chang and Yao's point about the inherent heterogeneity of the HS population is well taken. Yet one could argue, given the high degree of variability exhibited by HSs compared to other native speaker and L2 learner groups, that the differences in group means so often reported are insufficiently informative. Indeed, others have argued that presenting group means can fail to reveal interpersonal variation and can obscure differences both within and between groups (e.g. Birdsong, 2007; Henriksen, 2015; Markham, 1997; Ronquest, 2016). Across-group comparisons of means could be especially problematic if the HS group is constituted a priori as highly heterogeneous and then compared to a more homogeneous group, such as monolingual L1 speakers who speak the same local variety (as in Flores & Rato, 2016; Łyskawa, Maddeaux, Melara, & Nagy, 2016), which is not an appropriate baseline for comparison in most heritage contexts (Ronquest & Rao, 2015).

At least some of the variation expressed in Spanish HSs' phonological competencies might be explained by differences in the various experimental tasks employed. As Face (2003) and Rao (2009) have argued, the continuum of data elicitation techniques ranges from highly controlled lab prompts, such as reading isolated words, to free flowing spontaneous speech. Unscripted but semi-controlled/semi-spontaneous tasks such as story retelling from pictures are somewhere in the middle of the continuum. Tasks across the continuum are likely to elicit different speech styles and rates, which can have an effect on the phonetic features of the speech produced. For instance, the HSs in Ronquest's (2012) study produced longer and more greatly dispersed vowels when the elicitation tasks were more controlled. Similarly, the HSs in Goodin-Mayeda's (2016) study produced rhotics ("r" sounds) like other L1 speakers when engaged in a picture identification task but not when engaged in a read-aloud task. Though read-aloud tasks are commonplace in phonetics studies (e.g. Amengual, 2016) because of their obvious advantages in controlling for features of interest, they may serve to exaggerate phonetic differences in the HS population, because HSs have varying degrees of experience reading in their home language. In fact, Colantoni, Cuza and Mazzaro (2016) caution that "metalinguistic tasks, such as reading aloud, should be implemented with caution, crucially among Spanish heritage speakers who are in a semi-diglossic situation in the US" (p. 3). Future studies should continue to explore suitability of various tasks for HSs and, until then, perhaps a variety of tasks should be employed to permit a more holistic evaluation of HSs' competencies.

The argument that high variability is a challenge to be addressed has also been made in studies of L1 attrition. These studies examine the L1 of individuals who migrate to a different language community in adulthood and experience attrition. While there are qualitative differences between adult migrants and HSs, they do share the common experience of living for long periods in contexts where their L1 has minority language status and is relegated to particular domains of use. Some of the first studies in L1 phonological attrition, Major's (1992) study of voice onset time (VOT) in /p t k/ produced by English L1 late consecutive bilingual immigrants in Brazil, and Flege's (1987) study of VOT of /t/ produced by late consecutive English/French bilinguals living in the U.S. and France, failed to account for the high degree of variability in their data, but De Leeuw, Mennen and Scobbie (2012) have offered ideas about how to do so. In their study (2012) of prosody in 10 late consecutive German–English bilinguals, they underscored the interpersonal variation evidenced by their participants, particularly two who performed within the monolingual German norm and one who demonstrated no L1 attrition, and they reported on which aspects of prosody exhibited more intrapersonal variation. They also searched for predictors of that variability, exploring multiple variables. In a subsequent study De Leeuw, Mennen and Scobbie (2013) investigated the lateral phoneme /l/ with the same population and again found a high degree of interpersonal and intrapersonal variation, noting both "cognitive and socially-mediated change" over the lifespan of consecutive bilinguals (- p. 698). The language acquisition process of HSs, too, is affected by a multitude of factors and characterized by highly variable outcomes. Thus this population is quite complex as an object of study, but it is precisely the phonetic variability exhibited by HSs, rather than overall between-group differences in means, that will be most interesting to explore, along with the factors that predict that variability.

Differences in HSs' language abilities at the phonological level and beyond have thus far been explained largely as a function of proficiency (Polinsky & Kagan, 2007). In the specific context of this study (Spanish as a heritage language in the U.S.), language proficiency and language dominance, which is relative proficiency between two or more languages, have been considered important predictor variables (Amengual, 2016; Goodin-Mayeda, 2016; Ronquest, 2016, p. 292). It seems entirely sensible to look to proficiency and the related construct of relative proficiency (dominance) as explanations for variability in phonetic performance, yet there are also some challenges to this approach. The first challenge is that proficiency is a holistic measure based on numerous aspects of language that may be largely irrelevant for phonetic competence, such as morphosyntactic accuracy. Secondly, it is unclear why proficiency should be sufficiently explanatory when language experience is not, given the strong correlation between proficiency and language experience. To date, studies attempting to predict HS phonetic outcomes with language experience variables have been relatively unsuccessful (e.g. Rao, 2016a; Chang & Yao, 2016; Flores & Rato, 2016). For instance, Rao's (2016a) study divided Spanish HSs into groups of regular speakers, childhood speakers, and those with minimal exposure, but concluded that that past and present exposure to the heritage language did not sufficiently explain their intonational patterns. Chang and Yao (2016) divided Mandarin HSs into high exposure and low exposure groups but found little difference between those groups in the acoustic analysis of tones they produced. Since language experience constitutes a multidimensional set of language practices, perhaps operationalizing it as a categorical variable with two or three levels is simply too reductionist. Yet Flores and Rato (2016) attempted to predict global accent among Portuguese HSs with several continuous variables measuring language experience, and still they did not find statistically significant correlations. In sum, though language experience correlates with proficiency, language experience in itself has not proven to be predictive of phonetic outcomes among HSs, whereas proficiency has. Finally, and most importantly, focusing solely on linguistic measures such as proficiency, dominance, and language experience ignores extralinguistic factors that might help predict phonetic outcomes of HSs, such as identity factors.

CONSIDERING CULTURAL AND ETHNIC IDENTITY AS POTENTIAL PREDICTORS OF HS VARIABILITY

There are at least three reasons why identity might be important to consider in heritage phonetics. First, the term "heritage" inherently reflects a sense of shared identity or heritage, so it seems illogical to ignore HSs' attachments to their heritage. Second, a wealth of research indicates that matters of phonetics cannot be separated from the social meaning that speakers attach to them. Though accent is in some ways arbitrary by definition (a subjective and relative judgment of comparison with a local norm) and often inconsequential for communication (as it may not impact intelligibility), the social relevance of accent has been extensively documented. Le Page and Tabouret-Keller (1985) describe speech as "a series of acts of identity" through which "the individual creates for himself the patterns of his linguistic behavior so as to resemble those of the group or groups with which from time to time he wishes to be identified, or so as to be unlike those from whom he wishes to be distinguished" (p. 181). A number of studies have shown that "very fine phonetic detail is used for the construction of social identity" (Hay & Drager, 2007). Human beings are both highly sensitive to and highly prejudiced about accent, basing many interpersonal evaluations on accent alone (e.g., Fuertes, Gottdiener, Martin, Gilbert, & Giles, 2012).

Speakers construct multiple intersecting identities as they move through the world, and all those identities are likely to manifest phonetic markers, but two types of identity seem particularly relevant to heritage phonetics: ethnic and cultural identity. Ethnic groups are in essence "imagined communities" (Anderson, 1983) that are more social constructions than biological realities, but ethnic group membership is highly salient and meaningful in society. So too is cultural identity, which is often understood as engagement in the cultural practices shared by an ethnic group. These identities are constructed and marked in various ways, including accent. In a study testing the importance of accent using the "Who Said What?" paradigm, which presents participants with visual and audio stimuli from multiple talkers, Rakić, Steffens, and Mummendey (2011) found that participants relied as much on accent as on physical appearance to make ethnic categorizations of the speakers. Tseng (2014) found that a backed /æ/ vowel (as in the word "bad") in English both marked ethnic identity as Latino and served various stylistic functions for speakers in Washington, DC. Bailey's (2000) ethnographic study of second-generation Dominican American teenagers in Providence, Rhode Island, found that they used language to construct their multiple identities as Hispanic, American, and of African descent. They employed both prescriptivist standard American English forms and forms associated with African American Vernacular English to index different aspects of their identities. They also employed linguistic features of Caribbean Spanish such as elision of syllable-final /s/ (e.g., in *e[s]toy*, 'I am') to index an ethnolinguistic identity as Dominican or Hispanic, in resistance to the common black-white dichotomization that reduces racial identities to phenotypes. Giles and Johnson's (1987) Ethnolinguistic Identity Theory would explain this selective retention and use of distinctive markers as indicating that the speakers both align with an ethnic identity and view that identity as having vitality. Ethnicity and culture are likely important identity constructs for many speakers of ethnolinguistic minority languages. For example, Sánchez-Muñoz (2013) reported that college students enrolled in a U.S. Spanish Heritage language class claimed maintaining and developing Spanish was integral to their identities as Latinos (ethnic identity) and to doing their "duty" to pass on their culture to future generations (cultural identity).

A third reason why it might be important to consider identity in heritage phonetics is that there is some empirical evidence for a link already, in different contexts. Hirson and Sohail (2007) studied group affiliation among second-generation Punjabi speakers in southeast Britain. They found that speakers' self-identification as either "British Asian" or "Asian" correlated with their realization of /r/. Those who identified as "Asian" were more likely to retain features of Punjabi rhotic realizations and postvocalic /r/ when speaking in English. Schmid and Dusseldorp's (2010) study of study of L1 German speakers residing in Canada and the Netherlands found that identification and affiliation with L1 (German) culture was a unique predictor of the amount of attrition they experienced in German over time. Though they looked at language performance indicators (C-tests, GJTs, etc.), not phonetics specifically, an important finding of their study was the unique contribution of ethnic and cultural identity measures as separate from linguistic variables. That is, they conducted principal component analyses (PCAs) of many potentially relevant predictor variables and came to a two-component solution. One component contained variables relating to identification and affiliation with L1 *culture*, and the other component contained variables relating to exposure to and attitudes towards the *language*. For all these reasons, it seems important for investigations of HS phonetics to account for the role of ethnic and cultural identity. Yet surprisingly little attention is paid to identity in recently published studies on HS phonetics. The language background questionnaires typically used in the field (e.g., Asherov, Fishman & Cohen (2016), pp. 129–130; Kang, George & Soo (2016), pp. 215–216) focus on language use and proficiency almost exclusively. They rarely probe participants' identification or affiliation with particular speech communities. The Bilingual Language Profile used to assess language dominance (e.g., Amengual, 2016) includes just one such item: "I identify with a Spanish-speaking culture." It does include four items that measure attitudes towards language, but as Schmid and Dusseldorp (2010) found, attitudes towards language are not equivalent to measures of ethnic or cultural identity. One study that has attempted to account for cultural identity is Oh and Au (2005). They sought to identify sociocultural background variables that predicted mastery of accent and grammar in the heritage language. Their participants were Latino college students studying Spanish. Oh and Au found that use of Spanish outside the classroom was positively correlated with global accent ratings and ratings of stop consonant pronunciation. They also measured self-identification with and participation in Latino culture with The Multigroup Ethnic Identity Measure (MEIM; Phinney, 1992). They reported weak to moderate positive correlations between accent and identification with Latino culture, which led them to argue for more research to understand the role of multiple background variables. The current study follows that line of inquiry.

U.S. Heritage Spanish Phonetics and the Case of the Spanish Rhotics

This section offers a very brief summary of the research on U.S. heritage Spanish phonetics and acquisition of Spanish rhotics in order to provide background for the exploratory study presented here. Research with Spanish HSs in the U.S. suggests that they have a phonological advantage over L2 learners but are not identical to other native speakers. Adults who overheard Spanish as children, even if they did not use Spanish after childhood, have better pronunciation than late adult learners (Au, Knightly, Jun & Oh, 2002). HSs' advantage in production compared to L2 learners appears at both the segmental level, with stop consonants and vowels, and the suprasegmental level (e.g., a presentational focus as found in Hoot, 2012; see Rao & Ronquest, 2015, for a review). Yet though they may sound "native-like," Spanish HSs' perception (Mazzaro, Cuza & Colantoni, 2016) and production do not always reflect idealized monolingual norms, including intonation patterns, unconstricted approximant allophones of /b, d, g/, and consistent production of five tense, unreduced vowels in a symmetrical vowel space (Ronquest, 2013; Rao & Ronquest, 2015; Rao, 2016a).

The phonemes investigated in the current study are rhotics, which are the sounds that correspond to the grapheme "r." Spanish has two such rhotic phonemes, the trill /r/ and the tap /ɾ/. They are in contrastive distribution in intervocalic position (as in *pero*, 'but' and *perro*, 'dog'). In other positions, they are complementary in distribution: /r/ occurs in word initial position (as in *rojo*, 'red') and in syllable initial position after the heterosyllabic consonants /n, l, s/ (as in *sonreir*, 'to smile,' *alrededor*, 'around', and *Israel*, 'Isreal'), whereas /ɾ/ occurs after tautosyllabic consonants to form complex onsets (as in *cuatro*, 'four' and *drama*, 'drama') and in word final position before a vowel (as in *estar emocionado*, 'to be excited'). In syllable final (as in *carta*, 'letter') and word final (as in *amor*, 'love') positions before a consonant or pause, a tap is typically produced, but careful or emphatic speech may employ a trill.

The canonical Spanish tap is described phonetically as having one lingual contact, a quick tap of the tongue against the alveolar ridge just behind the upper front teeth. The canonical Spanish trill is described phonetically as having two or more lingual contacts in quick succession. However, rhotics are expressed with vast sociophonetic variation throughout the Spanish-speaking world (e.g. Henriksen, 2015; Lipski, 2008). This is especially true of the phonemic trill, which speakers may produce phonetically as a fricative (turbulent air like the English "sh" sound), a lateral (like "l"), an approximant (similar to the English "r"), a vowel, or a sound that combines more than one of these properties, for instance a fricative followed by a lingual contact.

Socially, rhotics are of interest to U.S. Spanish HSs because they contribute to the percept of accent and thus can index in-group or out-group status. The trill in particular is highly salient to listeners and can constitute a marked segment in HS speech. Rhotics are also of special interest to SLA researchers because of their acquisitional patterns. Rhotics are typically the last sounds that normally developing monolingual Spanish-speaking children acquire (Carballo & Mendoza, 2000), especially the articulatorily difficult trill. The trill is acquired after age 5, the age when most U.S. Spanish HSs start to be exposed to English more than Spanish. Early exposure to an L2 might explain why bilingual children produce rhotics less accurately than monolingual Spanish speakers of similar ages, though this difference may not emerge in vibrant Spanish-speaking communities (Fabiano-Smith & Goldstein, 2010).

Several recent empirical studies have documented variation in U.S. Spanish HSs' rhotics and posited various explanations for it. Henriksen (2015) attributed the non-canonicity of rhotics produced by second-generation Mexican Spanish speakers in Chicago (n = 8) to the variability inherent in the input HSs received rather than language contact or neutralization of the phonemes. Goodin-Mayeda (2016) studied Mexican and Central American Spanish HSs in

Houston, Texas, (*n* = 18) and reported that proficiency predicted performance: low-proficiency (*n* = 6) HSs' phonemic trills diverged from the trills of intermediate-proficiency (*n* = 12) HSs and NSs (*n* = 12). The effect was found in the reading task but not the picture-naming task, recalling Colantoni, Cuza and Mazzaro's (2016) recommendation to use reading tasks with caution in HS research. Amengual (2016) also used a read-aloud task to compare HSs (*n* = 40) to L2 learners (*n* = 20) in northern California. He found that while most L2 learners and English-dominant HSs produced noncanonical trills (fewer than 2 occlusions) and maintained the tap/trill phonemic contrast through duration, the Spanish-dominant HSs produced mostly canonical trills. Amengual concluded that HSs are a heterogeneous group, but language dominance helps explain the variation in their production of rhotics. In addition to the challenges of explaining HSs' phonetic variability and of selecting appropriate elicitation tasks, these studies also have faced the challenge of how to compare groups' productions of phonemes that vary across multiple phonetic features. These studies selected particular features of interest (duration, lingual contacts, and categorical ratings) to be studied in isolation as dependent variables. Yet given that all those phonetic features are interconnected for any particular realization, it seems that different methods could allow for comparisons across more than one feature of interest at a time.

In sum, previous studies have suggested that (a) Spanish HSs in the US do acquire two rhotic phonemes differentiated by duration, (b) HSs' rhotics are distinguishable from those produced by other Spanish NSs and L2 learners, (c) HSs' production of rhotics is characterized by high variability and (d) proficiency and dominance help explain this variability. However, previous studies have also exhibited the methodological challenges discussed in this introduction, which the current study attempted to address in several ways. The current study took an exploratory approach to analyzing a wider range of both dependent variables (rhotic productions) and independent variables (linguistic and extralinguistic predictor variables) and assessing their unique contributions with novel methods. The study was motivated by the following research questions:

RQ 1 – What does an exploratory analysis of heritage speakers' (HSs') rhotics suggest about HSs' rhotics vis-à-vis other speaker groups?
It is expected that HSs should be highly variable in their phonetic outcomes. It has also been argued that between-group comparisons of means can be problematic, especially if some groups are more variable than others. An exploratory analysis that included multiple acoustic measures of rhotics was employed here to determine how HSs' rhotics differed from other groups and if overall between-group mean comparisons would be justified.

RQ 2 – What does an exploratory analysis of linguistic and extralinguistic predictor variables suggest in terms of which could be most useful for identifying subgroups of HSs that might better account for HSs' phonetic variability?

Given that our current definition of HS is a broad umbrella term and that HSs as a broadly defined category of speakers typically express high variability in phonetic outcomes, it seems logical to explore methods for circumscribing the HS group and identifying subgroups more useful for elucidating the patterns in their variability. Previous studies have explored language proficiency, dominance, and experience variables. This study also included ethnic and cultural identity measures and explored the relative importance of these linguistic and extralinguistic predictor variables for identifying more narrow categories of different HS profiles.

RQ 3 - Does varying the approach to identifying HSs (from RQ 2) change our understanding of their rhotics vis-à-vis other groups?
The bottom-up (exploratory) approach to identifying HS subgroups based on a range of background measures was compared to a more conventional analysis—top-down (confirmatory) analysis comparing group means on isolated acoustic measures—to determine whether the approaches led to qualitatively different results.

METHODS

Context and Participants

The study took place in Richmond, Virginia, a city with a population of 223,170 (1.2 million in the greater metropolitan area) that is 6.5% Hispanic or Latino (U.S. Census Bureau, 2016). All participants (*n* = 38) were college-educated, proficient speakers of Spanish between the ages of 18 and 28 (average age = 20). Participants had one of four language background profiles, which will be referred to as heritage speakers, long-term immigrants, "traditional" native speakers, and second language learners. The heritage speakers (HSs, *n* = 15, 13 F) were lifelong residents of the U.S. and had used Spanish from birth with their families, who had immigrated from central Mexico or El Salvador, areas in which rhotics are realized canonically as alveolar taps and trills.[1] All the HSs self-reported English as their dominant language. The bilingual, long-term immigrant native speakers (IMs, *n* = 5, 4 F) immigrated from Mexico or El Salvador before adolescence. Two reported being dominant in English, as would be expected for immigrants with long lengths of residence. The "traditional" native speakers (NSs, *n* = 10, 4 F), meaning nearly monolingual native speakers, were lifelong residents of central Mexico (*n* = 6), or recent adult immigrants (*n* = 4) who had immigrated to the U.S. after age 17 and spent fewer than five years in the U.S. The term "native speaker" is used as a convention only, and its use is not meant to suggest that HSs and IMs were not also native speakers or even that the term "native speaker" is particularly meaningful. Indeed, HSs are, by any reasonable definition of this imprecise term, also "native speakers" (Rothman & Treffers, 2014). The advanced second language learners (L2s, *n* = 8, 6 F) were lifelong residents of the U.S. who used English from birth with their families and were pursuing undergraduate degrees in Spanish. No participant had training in phonetics. Except for the traditional NSs, all participants were attending college in Virginia during the study, but they had lived in other areas of the U.S. before moving to the state.

Five instruments assessed participants' backgrounds. (1) A questionnaire created for this study, (2) The Language Experience and Proficiency Questionnaire (LEAP-Q; Marian, Blumenfeld, & Kaushanskaya, 2007), and a (3) questionnaire from Oh and Au (2005) were used to gauge participants' language experiences and other relevant demographic features. (4) The Multigroup Ethnic Identity Measure (MEIM; Phinney, 1992) was used to measure ethnic identity. It is a reliable instrument that has been used in dozens of previous studies (Phinney, 1992), including Oh and Au's (2005) study of heritage Spanish. Both English and Spanish versions of the MEIM have been validated and proven suitable for use among adolescents and adults of various racial and ethnic groups, including Hispanics, as well as different immigrant statuses and generations (e.g. Avery, Tonidandel, & Thomas, 2007; Yap et al., 2016). The MEIM includes 15 Likert-scale items targeting exploration, commitment, cognitive clarity, affective pride, and behavioral engagement with self-reported ethnic identity. (5) The Abbreviated Multidimensional Acculturation Scale (AMAS; Zea, Asner-Self, Birman, & Buki, 2003) was also used; it measures identification with a culture of origin and U.S.-American culture. All participants in this study identified their culture of origin as Hispanic, Latino, or Mexican American. The AMAS has been validated with Latino college students (Zea et al., 2003) as well as older adults (Yamada, Valle, Barrio & Jeste, 2006). Of the 15 tools reviewed by Yamada et al. (2006), the AMAS is the only one that has been validated with a population like that of the current study (descendants of Mexican and Central American immigrants residing in the Mid-Atlantic U.S. The AMAS targets three factors: cultural identity, language competence, and cultural competence, which is operationalized as knowledge of cultural artifacts such as national heroes, history, and popular television shows. When items relating to both cultures are averaged, the AMAS measures acculturation, or the process of incorporating mainstream culture into one's minority culture. Isolating just the items pertaining to culture of origin, as the current study did, provides a measure of enculturation, which is the process of learning the culture of one's own group. Prior studies have often equated ethnic to cultural identity or understood them as two sides of the same coin. Oh and Au (2005), for instance, interpreted the MEIM to measure "identification with Latino culture." Indeed, some of the underlying psychometric properties of the MEIM and AMAS do correlate (Miyoshi, Asner-Self, Yanyan, & Koran, 2017), but the instruments are not identical, and given the exploratory goals of the current study, both were included here.

Participants' background information is summarized in Table 1. Years of Spanish immersion indicate full school years in which at least two content courses were taught in Spanish. The life stages participants transitioned to English were coded as 1 (before elementary school), 2 (during elementary school), 3 (during middle school), 4 (during high school), 5 (after high school), or 6 (never), based on a questionnaire from Oh and Au (2005). The next four variables were based on participants' responses on the LEAP-Q. *Prefer to speak Spanish* indicates the percentage of time participants would choose to speak Spanish with a person who was equally fluent in Spanish and English. Relative proficiency was calculated by subtracting participants' self-rating of English speaking proficiency from their Spanish rating, on a scale of 0 (none) to 10 (perfect); positive scores reflect higher relative speaking proficiency in Spanish. The Ethnic identity score reported in Table 1 reflects an average of all 15 items on the MEIM, and the cultural identity score reflects an average of the 21 items on the AMAS relating to culture of origin; the authors of the instruments do not recommend further subdividing the items (Phinney, 1992; Zea et al., 2003).[2]

Table 1 also reports speech rate because it has been argued that speech rate represents a "promising method of identifying and classifying heritage speakers" (Benmamoun, Montrul, & Polinsky, 2013, p. 135). Speech rate was calculated as the average words per minute produced during fluent spurts of at least 5 seconds during the middle of each experimental task.

Table 1.

Group Averages of Language Background Measures

Group	Years of Spanish immersion in formal instruction	Age of learning English	Life stage transitioned to English (1-5, 6 = never)	Live in Spanish-speaking country (% of life)	Current use of Spanish (% of day)	Prefer to speak Spanish (% of time)	Relative speaking proficiency (Spanish-English) (-10 – 10)	Ethnic identity (MEIM) (1-4)	Cultural identity (AMAS) (1-4)	Speech rate (WPM)
Heritage Speakers (*n* = 15)	2.73 (3.37)	4.33 (1.72)	3.33 (1.50)	15.60 (22.59)	22.27 (17.26)	35.27 (13.85)	-1.33 (1.72)	3.24 (0.45)	3.10 (0.43)	188.51 (22.23)
Long-term Immigrants (*n* = 5)	5.80 (1.92)	9.80 (2.95)	3.50 (1.91)	43.38 (4.09)	32.50 (18.93)	35.00 (24.83)	0.75 (1.71)	2.98 (0.45)	3.21 (0.29)	201.44 (9.73)
Traditional Native Speakers (*n* = 10)	15.10 (2.51)	8.50 (4.50)	5.56 (0.73)	96.41 (6.79)	70.00 (35.97)	85.33 (22.30)	3.78 (1.56)	3.74 (0.45)	3.84 (0.32)	172.28 (17.99)
L2 Learners (*n* = 8)	0.00 (0.00)	0.00 (0.00)	1.00 (0.00)	1.21 (1.26)	12.13 (9.05)	14.38 (8.63)	-2.63 (1.41)	1.77 (0.25)	2.00 (0.16)	132.43 (20.54)

Table 2.

Distribution of Tokens across Phonological Context and Task (All Groups)

Task	Inter-vocalic	Word initial	After hetero-syllabic consonant	After tautosyl-labic consonant	Before con-sonant	Word final before vowel	Word final before consonant	Word final before pause
Trills (n = 1608)								
Reading (n = 452)	49%	44%	0%	0%	3%	0%	0%	3%
Story Retell (n = 950)	51%	46%	1%	0%	1%	0%	0%	1%
Conversa-tion (n = 206)	23%	67%	0%	0%	8%	0%	0%	2%
Taps (n = 4264)								
Reading (n = 1244)	31%	0%	0%	27%	20%	6%	5%	10%
Story Retell (n = 1336)	31%	0%	0%	29%	22%	5%	8%	5%
Conversa-tion (n = 1648)	30%	0%	0%	26%	24%	8%	8%	4%

Data Collection

Participants performed three tasks meant to represent the continuum of data elicitation techniques: spontaneous, semi-spontaneous, and controlled. The participants first completed a spontaneous conversation task, which was to talk for at least 3 minutes but no more than 5 minutes about the place where they grew up. They were encouraged to talk about whatever was most comfortable for them, including their childhood activities, family, friends, school, home, neighborhood, attractions, or weather. The researcher made this task conversational by reacting appropriately and asking for more information if participants stopped talking before the 3 minute minimum. The second task was semi-spontaneous. Participants were asked to retell a story based on pictures. The story was the children's book *Frog, where are you?* (Mayer, 1969), which is freely available to researchers and suitable for participants with a range of language abilities (e.g., Reilly, Losh, Bellugi, & Wulfeck, 2004). The book has been used in Spanish HS phonetics studies previously (Henriksen, 2015) because its characters are a boy, a dog (in Spanish, *perro*), and a frog (*rana*) and thus will require numerous tokens of the phonemic trill if narrated in Spanish. Participants were given time to preview the entire story before beginning the retell. They were asked to spend at least 3 but not more than 5 minutes on this task. At participants' request, the researcher supplied vocabulary words when needed. All participants produced the words *perro* and *rana* without assistance. Finally, participants completed a read-aloud task in which they read a list of 22 idioms (see Appendix A). Recordings were made in a quiet room with a studio quality microphone, digitized into a wav format (44kHz, 16 bit quantization).

Acoustic Analysis

The recordings were analyzed with the acoustic analysis software PRAAT (Boersma & Weenink, 2015). All tokens of the phonemic trill that participants produced (range of 26-70 tokens per participant across 3 tasks, average of 42) were submitted to analysis. Participants produced more taps than trills, on average 40 tap tokens per task. This average number of tokens was considered a large enough sample to be reliable, and only the first 40 tokens produced by a participant on each task were submitted to analysis. On the rare occasion when a token was too noisy to analyze, it was excluded. The final dataset included 1608 trill measurements and 4264 taps. The tokens in the dataset were distributed across tasks and phonological contexts but not balanced perfectly (see Table 2).

Each measurement, or token, was coded for: (a) duration; (b) number of lingual contacts based on acoustic cues, and (c) number of lingual contacts based on auditory perception alone. Lingual contacts are brief taps of the tip of the tongue against the alveolar ridge. They can be detected by ear and can also usually be seen visually in PRAAT as marked by an apical occlusion, indicated through reduced waveform amplitude and lack of formant structure in the spectrogram. The canonical tap is described as having one lingual contact and the canonical trill two or more lingual contacts. Examples of each are provided in Appendix B. For some tokens, the auditory and visual evidence were not identical, and in these instances both measures were included for analysis.

However, Spanish speakers produce many non-canonical rhotics. In an effort to capture the variation exhibited by speakers in this study, each token was also coded categorically as either a canonical tap, canonical trill, approximant, fricative, perceptual, or elided phone. These categories reflect all the major categories used in previous investigations of rhotic variation among Mexican and Mexican American speakers (e.g. Bradley & Willis, 2012; Henriksen, 2015). Examples of tokens and coding are provided in Appendix B. Tokens that presented auditory evidence of a lingual contact but lacked reliable acoustic landmarks in the waveform and spectrogram were coded as perceptual. Tokens that presented no auditory or acoustic evidence of rhotic production were coded as elided. Tokens with no audible lingual contact and with clear formant structures present throughout in the spectrogram were classified as approximants; these often sounded like the English /r/, a retroflex consonant. Tokens exhibiting non-periodic noise throughout most of their duration were classified as fricatives. Portions of tokens that exhibited features of other categories were also measured. These included fricative portions as well as vowel-, approximant-, or “r”-like portions. Many of these acoustic properties should correlate in principle (e.g., tokens with more lingual contacts are longer in duration), but all contribute to the overall quality of any given token, which is why they were all included in the exploratory analysis of rhotics.

Two raters blindly coded all the tokens of the first HS. They compared their codes and discussed their differences until reaching consensus on all tokens. They followed the same procedure for the next 4 HSs. Subsequently, the first rater completed all of the coding for the rest of the speakers. The second rater selected a random sample of 10% of the data and coded it independently. Inter-rater agreement on this subset of the data was 89%, which was considered sufficiently reliable, given the large number of tokens being coded, and so the first rater's codes were used in all cases.

Statistical Analyses

All statistical analyses were conducted with SPSS (Version 24.0), beginning with principal components analyses (PCA). PCA is an exploratory data reduction technique uncommonly seen in the field of applied linguistics, so it requires further explanation. PCA is often confused with factor analysis, but differs in that it is a solely exploratory technique and does not estimate any underlying latent factors, but instead linear components. PCA decomposes the data into linear "components," which are composite variables, and determines how particular independent variables contribute to those components. It uses a correlation matrix of the variables of interest to compute the components, which can number between one and the total number of variables in the analysis. Each variable is assumed to have a common variance (or communality) of one, which is contributed to the analysis. Each component in a PCA will have an associated eigenvalue; the higher the eigenvalue, the more variance the component explains. In a PCA that includes all possible components, the sum total of eigenvalues will always equal the total number of variables. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy (Kaiser, 1970; 1974) is one method to ensure that PCA is appropriate and would yield distinct and reliable components from the data. Values can range from 0 to 1, and values below .5 are unacceptable, between .7 and .8 good, and between .8 and .9 great (Hutcheson & Sofroniou, 1999).

PCA is an exploratory data reduction technique, which means that the goal of PCA is to explain a majority of the variance in a dataset with fewer variables than when the analysis was begun. Each successive component in a PCA tries to explain as much linear variance as it can in the data that remains after each previous component. Many different benchmarks are used for selecting the number of components out of a PCA. The criteria used here include: (1) stopping after there are enough components to explain at least 70% of the variance in the data; (2) visual inspection of a scree plot (component number on the x-axis and the eigenvalue of the component on the y-axis; Cattell, 1966) and retaining only those components that appear before a sudden drop, and (3) Kaiser's rule (1960), which is to retain all components that have an eigenvalue greater than one, the logic being that a component should explain more variance than any one variable contributed.

Similar to factor analysis, PCA output provides a loadings matrix. A rule of thumb of .4 was applied in this study, meaning that a variable (that loaded greater than or equal to .4 on a particular component) contributed meaningfully to that component and should contribute to the interpretation of that component. Also similar to factor analysis, PCA solutions can be rotated to improve fit between the variables and components and, as a result, the interpretation of the components. All PCAs in this study with more than one component were performed with Oblimin rotation, which allows components to correlate with one another. For further information on the PCA as a method, see Field (2005). For an example of PCA of a range of correlated predictor variables that combine into components, see Schmid and Dusseldorp's (2010) study of the impact of extralinguistic factors in L1 attrition among L1 German emigrants.

To address RQ 1, first PCAs of the acoustic data were conducted to see the separability of participants across the variability in their acoustic output. The advantage of PCA to approach this RQ was that it allowed all the potentially relevant phonetic features of rhotics, which have been analyzed separately in previous studies, to be analyzed in one model. Though many of those features should correlate, PCA technique allows for such correlation.

To address RQ 2, a PCA of potential predictor variables was conducted to investigate the separability of participants across the common variance of the measured predictor variables. Again, the advantage of PCA to approach this RQ was that it allowed all the potentially relevant predictor variables to be analyzed together, even though many correlated, and to model at once how participants from all groups differed on all these predictor variables. Heritage speakers (HSs) were split into different groups based on the spread of participants from the results of this PCA, which constituted an inductive method for assigning group membership. In order to verify which predictor variables contributed most to distinguishing the HSs subgroups that emerged in the inductive approach, a repeated measures analysis of variance (RM ANOVA) with follow up planned independent samples *t*-tests were conducted. The assumptions of normality and heterogeneity of residuals were met for this and all other RM ANOVA analyses. When the statistical assumption of sphericity (that the differences between the variances of a single participant's data are equal) was not met via Mauchly's test, the Greenhouse-Geisser correction was used (Howell, 2002). This correction was chosen because it is the one most commonly used in the field, though different corrections rarely produce different results.

A series of RM ANOVAs was then run to determine if this inductive method for grouping based on a variety of linguistic and extralinguistic variables was indeed advantageous (RQ 3) for predicting variability in HS's production of rhotics. First, two RM ANOVAs were run like in prior studies: deductive RM ANOVAs, in which group assignments were made *a priori* based on limited linguistic background variables. One RM ANOVA tested group differences in terms of rhotic Duration, and the second tested group differences in terms of number of Lingual contacts (detected acoustically). Then, identical RM ANOVAs were run with one change: group assignment made inductively. The relative adequacy of these models was compared directly via comparison of effect size (partial eta-squared), which is an indication of how much variance in the data is explained by the model.

RESULTS

RQ 1. Exploratory Analysis of Rhotic Acoustics

Descriptive acoustic measurements of trill and tap tokens, averaged for each participant group and across all three tasks, are reported in Table 3. Recall that canonical taps are perceived as having just one lingual contact, and canonical trills are perceived as having two or more lingual contacts. The majority of trills and the vast majority of taps produced by heritage speakers (HSs), bilingual long-term immigrant native speakers (IMs) and traditional native speakers (NSs) fit these definitions of canonical rhotics (see Table 3). Second language learners (L2s) produced many more noncanonical rhotics than the other groups. HSs' trills differed from traditional NSs in that they were shorter on average and had fewer lingual contacts and more frication (with some HSs producing many fricatives while others did not), but on most of these measures HSs' trills appeared similar to those produced by IMs. All these acoustic measures were potentially relevant, yet interrelated to varying degrees, and so they were submitted to a principal component analysis (PCA) to tease apart on what dimension(s), if any, the participant groups differed. HSs produced taps descriptively similar to traditional NSs on all measures, and for this reason the PCA explored the trill data only, not taps.

Table 3.

Group Averages (SDs) on Acoustic Measures

Group	Duration (ms)	Lingual contacts detected acoustically	Lingual contacts detected auditorily	Approx-imant *	Fricative *	Percep-tual *	Elided *	Fricative portion (ms)	Vowel or approximant portion (ms)
Trills									
L2 (*n* = 8)	55.44 (27.07)	0.92 (0.48)	1.16 (0.50)	0.26 (0.24)	0.04 (0.06)	0.19 (0.22)	0.00 (0.00)	5.96 (6.86)	10.25 (5.83)
HS (*n* = 15)	80.78 (10.88)	1.39 (0.45)	1.71 (0.50)	0.07 (0.08)	0.23 (0.55)	0.05 (0.06)	0.00 (0.00)	16.45 (18.74)	11.67 (4.53)
IM (*n* = 5)	79.14 (10.69)	1.34 (0.45)	1.62 (0.44)	0.03 (0.03)	0.09 (0.06)	0.20 (0.41)	0.00 (0.00)	15.20 (9.18)	17.31 (3.02)
NS (*n* = 10)	89.26 (10.11)	1.83 (0.33)	2.07 (0.26)	0.05 (0.05)	0.06 (0.04)	0.02 (0.02)	0.00 (0.00)	9.04 (8.33)	11.43 (6.39)
Taps									
L2	38.30 (17.87)	0.56 (0.31)	0.77 (0.25)	0.26 (0.26)	0.12 (0.26)	0.30 (0.27)	0.05 (0.04)	6.32 (6.68)	10.18 (6.46)
HS	40.54 (5.11)	0.69 (0.14)	0.92 (0.04)	0.13 (0.07)	0.05 (0.10)	0.11 (0.07)	0.06 (0.04)	3.32 (3.66)	11.91 (2.96)
IM	38.78 (6.93)	0.68 (0.13)	0.90 (0.04)	0.11 (0.05)	0.04 (0.03)	0.11 (0.06)	0.06 (0.04)	3.82 (2.98)	11.64 (4.04)
NS	41.51 (4.89)	0.69 (0.05)	0.94 (0.04)	0.14 (0.02)	0.02 (0.02)	0.12 (0.04)	0.04 (0.03)	3.55 (2.16)	12.91 (2.53)

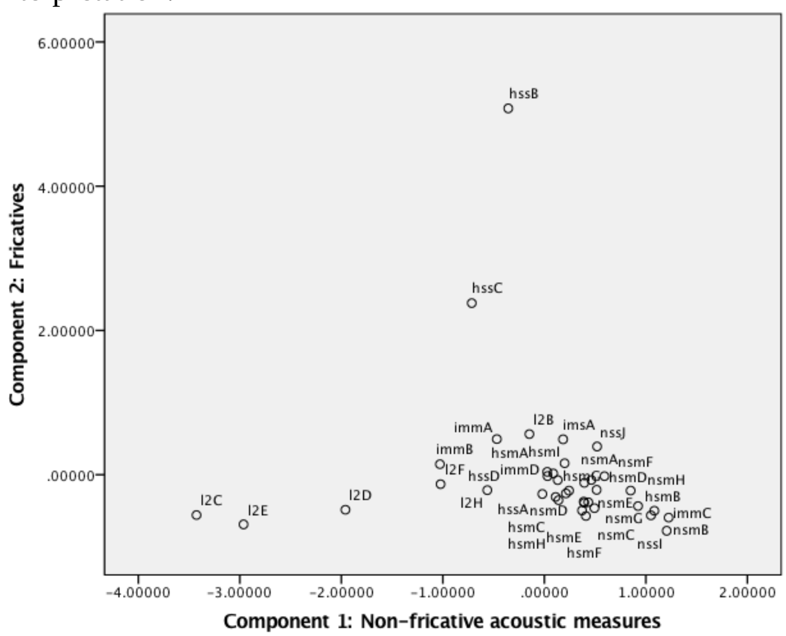
* Each token rated categorically, 0 or 1

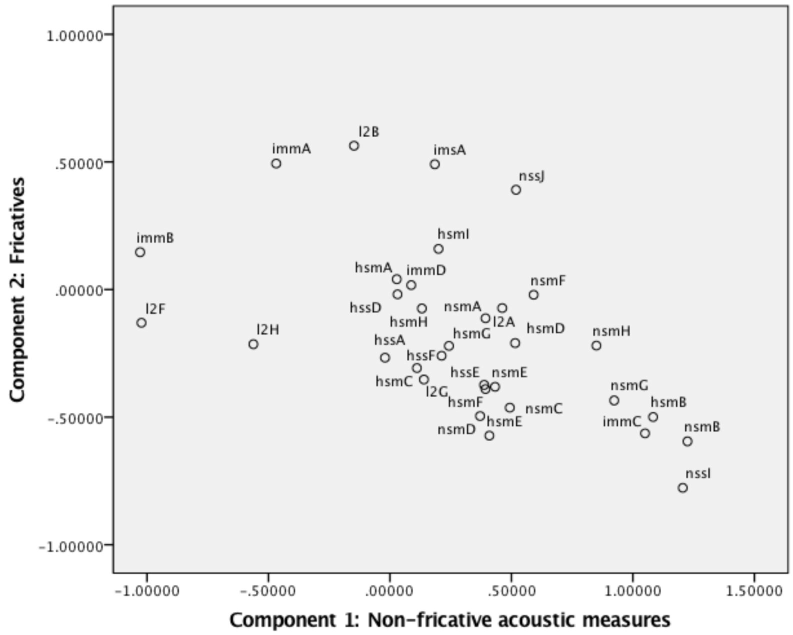
Table 4.

Loading Matrix of PCA of Acoustic Measures

	Non-fricative acoustic measures (Component 1)	Fricatives (Component 2)
Duration	0.87	0.24
Lingual contacts (acoustic)	0.83	-0.39
Lingual contacts (auditory)	0.83	-0.38
Approximants	-0.84	-0.28
Fricatives	0.02	0.95
Perceptuals	-0.70	0.04

Figure 1. Scatterplot of the First Two Principal Components for Acoustic Measures of Trills of All Participants (above) and a Zoomed-in View of the Central Data Cluster (below). Correlated Components Are Depicted Orthogonally for Ease of Interpretation.





Previous studies (Amengual, 2016; Goodin-Mayeda, 2016; Henriksen, 2015) have described HSs' rhotics in three ways: with categorical ratings, measures of duration, and number of occlusions (as markers of lingual contacts). The methods employed in those studies necessitated that these three types of acoustic measures be analyzed in isolation as distinct dependent variables along which HSs could be compared to other groups. However, in reality they all contribute to what listeners hear for any given rhotic**.** All three are implicated in each production of a rhotic, and they correlate and interact to varying degrees. Ideally, future studies that compare HSs with other groups will develop more sophisticated analytical methods to account for the interplay among many potentially relevant phonetic features. To take a step in that direction, in the present study an exploratory principal component analysis (PCA) with Oblimin rotation was conducted with all the acoustic measures of trills that demonstrated a good sampling adequacy for analysis (KMO = .7). These variables were: Duration, Lingual contacts (acoustic), Lingual contacts (auditory), Approximants, Fricatives, and Perceptuals. For this analysis, measures were averaged across all three tasks.

Recall that the objective of PCA is to reduce a number of interrelated variables into fewer composite variables, or components. In this PCA, a two component solution was reached that accounted for 77.95% of the variance in the data and, following Kaiser's rule, only the first two components of the PCA had eigenvalues greater than one. The two components' loadings appeared to roughly correspond to (1) Non-fricative acoustic measures and (2) Fricatives, with a correlation of -0.91. Their loading matrix is reported in Table 4. Figure 1 is a scatterplot of these components, with individuals labeled by participant code. The plot shows that HSs were indeed highly variable as predicted, with some producing trills identically to NSs and others producing trills quite unlike NSs. Figure 1 also suggests that most of the HSs were indiscernible from the distribution of the NSs and IMs along the dimensions of these two components, with the exception of only two (hssC and hssB).[3] This suggests that to simply compare overall group means between these HSs and the other groups, especially when multiple acoustic measures are accommodated into one analysis, would not be the most appropriate way to elucidate meaningful patterns in the data.

RQ 2. Exploratory Analysis of Background Measures

In an attempt to identify useful subgroups of HSs based on a variety of linguistic and extralinguistic predictor variables, another exploratory analysis was conducted on the background information collected via survey instruments. This PCA included all the background measures that had sufficient variance among the HS sample, which are those listed in Table 5. There was great sampling adequacy for this analysis (KMO = .86). Only one component had an eigenvalue greater than one, and so a one-component solution was reached that accounted for 77.30% of variance. This component appeared to correspond to all the background measures, as shown in Table 5, with Cultural identity and Prefer to speak Spanish having the highest absolute values of loadings. With just one component generated, the solution could not be rotated. The component scores produced for each participant in this analysis ranged from -1.50 to 1.88, with a median score of -0.15. All NSs had positive scores (0.36 – 1.88), all L2s had negative scores (-1.50 – -0.91), and IMs and HSs fell along a continuum between L2s and other NSs, as shown in Figure 2.

In order to allow for subsequent between-group comparisons, this continuum of HSs was subdivided into groups. Most HSs appeared to cluster in two discernable subgroups on either side of the mean component score, and henceforth are referred to as HS+ID and HS-ID. This shorthand notation reflects the weight of identity measures relative to the various other background measures that loaded onto the component (Table 5). The HS+ID (n = 6) group included HSs whose component scores (-0.04 – 0.31) were above the overall median. The HS-ID (n = 8) group included HSs whose component scores (-0.72 – -0.20) were below the median. HssB was considered an outlier by virtue of falling within the L2 range and was thus excluded from further analyses of these HS subgroup.

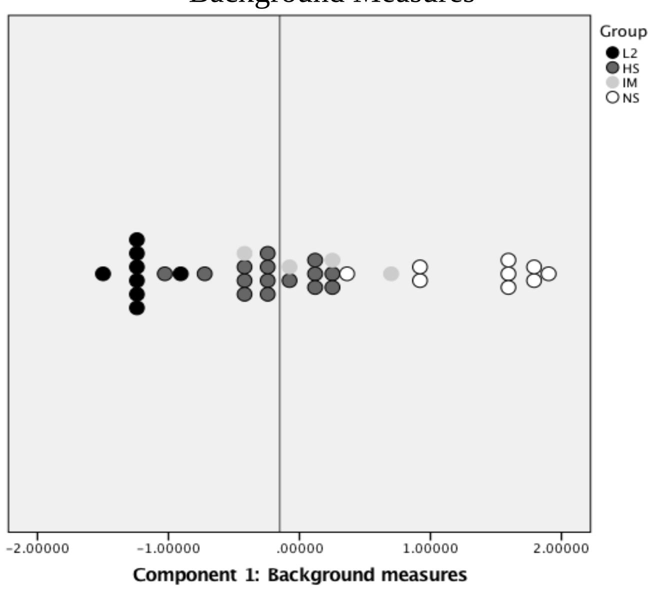
The overall goal of the present study was to find new approaches for accounting for variability in HS phonetics. The PCA of background measures was one step in the process. Its aim was to discover whether, based on a variety of linguistic and extralinguistic background variables, distinct subgroups of HSs could be discerned as separable both from L2 and NS backgrounds. The results (Figure 2) suggested that for all but one participant, whose background was more L2-like, indeed two subgroups of HSs could be identified: HS+ID and HS-ID.

Table 5.

Loading Matrix of PCA of Background Measures

	Background Measures (Component 1)
Years of Spanish immersion in formal instruction	0.87
Age of learning English	0.80
Life stage transitioned to English	0.86
Live in Spanish-speaking country	0.90
Current use of Spanish	0.88
Prefer to speak Spanish	0.92
Relative speaking proficiency	0.91
Ethnic identity	0.86
Cultural identity	0.92

Figure 2. Scatterplot of the Principal Component for Background Measures



RQ 2a. Background measures most relevant for discerning HS subgroups

The loadings onto the component were suggestive (Table 5), but to confirm which primary background variable(s) were driving the observed clustering of HSs into two subgroups (Figure 2), the background data of just the HSs were analyzed again. The variables were converted to zscores to get them on a common scale and were then submitted to a repeated-measures analysis of variance (RM ANOVA) with a between-groups factor of Group (two levels: HS+ID and HS-ID) and within-groups factors of various Background measures. There was a significant main effect for Group with a large effect size, $F(1, 12) = 44.94$, $p < .001$, $\eta_p^2 = .79$, observed power 1.00, suggesting that the groups (HS+ID and HS-ID) were quite different in terms of these background measures. Follow-up planned independent samples *t*-tests indicated that the variables for which the groups differed significantly were Current use of Spanish, Live in Spanish-speaking country, Ethnic identity, and Cultural identity, as reported in Table 6. Effect sizes were large, and the largest effect size was found for Cultural identity, Cohen's $d = 2.1$.

In sum, the analyses described up to this point (both PCAs and RM ANOVA) suggested that (a) it was not in fact easy to discern many HSs from other NSs and IMs in an exploratory analyses of their rhotics that considered multiple acoustic features (Figure 1), (b) in terms of their background, HSs could be usefully divided into two groups (Figure 2), and these groups (c) differed as much or more on the extralinguistic measures of Cultural identity and Ethnic identity as linguistic measures like Proficiency (Table 6). These results inspired further analyses to compare the utility of the present study's approach to grouping HSs with methods more commonly used to date in the field.

Table 6.

Language Background Differences between HS Subgroups

Group	Years of Spanish immersion in formal instruction	Age of learn-ing English	Life stage transi-tioned to English (1-5, 6=never)	Live in Spanish-speaking country (% of life)	Current use of Spanish (% of day)	Prefer to speak Spanish (% of time)	Relative speaking profi-ciency (Spanish-English) (-10 – 10)	Ethnic identity (MEIM) (1 – 4)	Cultural identity (AMAS) (1 – 4)
HS-ID (*n* = 8)	2.63 (3.96)	3.88 (1.36)	3.13 (1.73)	6.08 (17.20)	14.25 (15.28)	33.75 (17.06)	-1.38 (1.30)	3.05 (0.48)	2.88 (0.34)
HS+ID (*n* = 6)	3.33 (2.80)	5.00 (2.19)	3.83 (1.17)	30.89 (23.26)	35.83 (11.14)	38.33 (10.33)	-0.50 (0.84)	3.56 (0.19)	3.48 (0.22)
t statistic	0.37	1.19	0.86	2.30	2.92	0.62	1.43	2.47	3.72
p-value	.72	.26	.41	.04	.01	.55	.18	.03	.003
95% CI	(3.44-4.87)	(0.94-3.19)	(1.08-2.50)	(1.33-48.28)	(5.45-37.71)	(11.51-20.68)	(0.46-2.21)	(0.06-0.97)	(0.25-0.95)
Cohen's *d*	0.20	0.61	0.47	1.21	1.61	0.32	0.80	1.40	2.10
Power	0.06	0.18	0.13	0.55	0.79	0.09	0.28	0.67	0.95

HS-ID members: hsmA, hsmB, hsmC, hsmG, hssA, hssC, hssD, hssE

HS+ID members: hsmD, hsmE, hsmF, hsmH, hsmI, hssF

* *p*-values are not corrected for multiple comparisons. If a Bonferroni correction is used, only the Cultural identity result survives ($p = .027$). However, note the small sample size, the methodological focus of the current paper, and claims that the Bonferroni correction is overly conservative (e.g., Riazi, 2016, p. 23). Further, without correction, given the alpha level of .05 (and so a false discovery rate of 5%) for nine *t*-tests, one would only expect maximally one spuriously significant result if all null hypotheses were true, and there are four significant results here.

RQ 3. Comparing Deductive and Inductive Approaches

Between-group comparisons of participants' rhotics were made again, with methods more typical of the field, but with two different approaches to classifying participant groups. The goal was to investigate whether different methods of classifying HSs might impact the results obtained. To that end, RM ANOVAs were conducted to compare the effect of Group on Duration and Lingual contacts detected acoustically, which are the features of rhotics most often investigated in prior studies (e.g., Amengual, 2016; Henriksen, 2015). These analyses included the between-groups factor of Group and within-groups factors of Task and Phone because these variables have been included in prior studies (e.g., Goodin-Mayeda, 2016). The analyses were completed in two stages, with two different approaches to group assignment. The deductive approach was to assign group membership a priori by virtue of the participant using Spanish in the home from birth, the characteristic conventionally used to identify HSs, which resulted in one HS group.[4] The inductive approach was to assign group membership based instead on observing overall patterns in their background, via the PCA of all collected background variables as described in the previous section, which resulted in two subgroups: HS-ID and HS+ID (Table 6).

The present study compared the Group variable, defined differently (the traditional deductive approach and a more novel inductive approach), across two otherwise identical RM ANOVAs, and noted their resulting effect sizes. Effect size is a measure of how much variance is explained by the model. Thus comparing effect size is a comparison of goodness of fit of the model to the data. In the deductive analysis of Duration, all main effects were significant ($p < .05$). Fisher's least significant difference (LSD) post hoc tests indicated that only the L2 group differed significantly from the other groups; HSs were not significantly different from NSs or IMs. The only interaction effect that reached significance was Phone × Group.[5] Importantly, though, when the inductive approach was taken to identify HS subgroups, the variance explained by the Phone × Group interaction increased by 10% (deductive approach $F(3, 32) = 5.77$, $p = .003$, $\eta_p^2 = .35$, observed power .92; inductive approach $F(4, 30) = 6.10$, $p = .001$, $\eta_p^2 = .45$, observed power .97). Similarly, in the analysis of Lingual contacts, all main effects were significant in both the deductive and inductive analyses, and the only interaction effect that reached significance was Phone × Group. The effect size of this interaction increased by 4% when the inductive approach to grouping was used ($F(4, 30) = 4.20$, $p = .008$, $\eta_p^2 = .36$, observed power .88) instead of the deductive approach ($F(3, 32) = 5.08$, $p = .005$, $\eta_p^2 = .32$, observed power .88). The effect size of the main effect of Group also increased by 5% when the inductive approach was used ($F(4, 30) = 5.17$, $p = .003$, $\eta_p^2 = .41$, observed power .94) instead of the deductive approach ($F(3, 32) = 5.89$, $p = .003$, $\eta_p^2 = .36$, observed power .93). These changes in effect size were small, but the consistency of the pattern was intriguing and suggests that the inductive approach lent itself to increased explanatory power. Perhaps more importantly, in the deductive analysis, LSD post hoc tests (Tables 7 and 8) indicated that the differences between HSs and NSs were significant. However, in the inductive analysis, only the HS-ID subgroup was found to be different from NSs, whereas the HS+ID subgroup was not. In sum, in the analyses of Duration and Lingual contacts, adopting an inductive approach instead of a deductive approach to classifying HSs, both increased explanatory power and produced different results, and thus the approach to grouping changed the conclusions one would likely draw about HSs' rhotics vis-à-vis other groups.[6]

Table 7.

LSD Post hoc Tests of Group Differences in Lingual Contacts (Deductive Approach)

Groups		*Mean difference*	*Std. Error*	*p*	*95% CI*	
HS	L2*	0.34	0.12	.010	0.09	0.59
	IM	0.01	0.13	.938	-0.26	0.28
	NS*	-0.22	0.10	.048	-0.43	0.00
IM	L2*	0.33	0.16	.041	0.01	0.65
	NS	-0.23	0.14	.117	-0.51	0.06
NS	L2*	0.56	0.13	< .001	0.29	0.83

* $p < .05$

Table 8.

LSD Post hoc Tests of Group Differences in Lingual Contacts (Inductive Approach)

Groups		*Mean difference*	*Std. Error*	*p*	*95% CI*	
HS-ID	L2*	0.29	0.13	.036	0.02	0.57
	HS+ID	-0.17	0.13	.206	-0.45	0.10
	IM	-0.04	0.14	.801	-0.32	0.25
	NS*	-0.26	0.12	.033	-0.50	-0.02
HS+ID	L2*	0.47	0.14	.003	0.18	0.76
	IM	0.14	0.15	.368	-0.17	0.44
	NS	-0.09	0.13	.491	-0.35	0.17
IM	L2*	0.33	0.15	.036	0.02	0.64
	NS	-0.23	0.14	.106	-0.50	0.05
NS	L2*	0.56	0.13	< .001	0.30	0.82

* $p < .05$

DISCUSSION

The objective of this article has been to address some of the methodological challenges presented by research on heritage phonetics, the most important of which is the inherent heterogeneity of the heritage speaker population, as it is currently broadly defined. An exploratory study of heritage Spanish rhotics was conducted to illustrate these challenges and attempt to identify some productive methods for dealing with them. The findings suggested several avenues that could be explored in future studies.

Variability

Researchers have found time and again that heritage speakers (HSs) are highly variable in their backgrounds and in their language competencies, including in the area of phonetics. Studies in heritage phonetics also report how HSs on average, as a group, pattern differently from other first language speakers and from second language speakers. While these studies make important contributions to the research on heritage languages, focusing mainly on group comparisons of central tendencies may be less productive than exploring intra-speaker variability, if the main goal is to better understand the HS population. Such comparisons are even less illuminating when the HSs recruited are highly heterogeneous but the other native speakers (NSs) recruited are monolingual speakers who share a local variety of the target language. These across-group comparisons are problematic in other ways, too; recruiting monolingual non-U.S. residents for NS controls fails to compare HSs' production to the source variety they actually learned (Polinsky & Kagan, 2007), fails to recognize long-term U.S. residents as legitimate "native" speakers, and may perpetuate the "linguistic hegemony of monolingual (foreign born) Spanish varieties in the US" (Mrak, 2011).

In this study, a Principal Component Analysis (PCA) explored the rhotics produced by advanced second language learner (L2s), heritage speakers (HSs), bilingual long-term immigrant native speakers (IMs), and traditional (i.e., nearly monolingual) native speakers (NSs) of Spanish. The variables included were those for which the speaker groups have been shown to exhibit significant differences in prior research (duration, lingual contacts, etc.). Here a traditional top-down analysis (comparison of group means) found the HSs to be significantly different from NSs and L2s (Table 7). However, when the same data were submitted to exploratory, bottom-up analysis, it was difficult to discern many HSs from other NSs and IMs (Figure 1), whereas some HSs were starkly different. These divergent results suggest that our across-group comparisons of central tendencies should be complemented with other methods that can elucidate the intra-speaker variability exhibited by HSs.

The present study also suggested that the heterogeneous HS group might be better understood as made up of subgroups more narrowly circumscribed based on differing background profiles. The present exploratory study cannot offer definitive answers yet about precisely how those background profiles should be defined, but the results suggest that it is worthwhile to explore more inductive grouping mechanisms and include a wider variety of potentially important background variables. Here, models that grouped HSs inductively explained more phonetic variability (e.g., fit the data better) than models that assigned group membership a priori based on a broad definition of "heritage speaker" (see Results RQ 3). Future research might explore other techniques for identifying distinct profiles of HSs that can better account for their apparent endless heterogeneity.

Holistic Approaches to Phonetic Variation

Prior studies of U.S. Heritage Spanish rhotics have made across-group comparisons utilizing ANOVAs, mixed effects models, and regression analyses. These analyses require that groups be compared on one dependent variable at a time. For rhotic phonemes, the dependent variables selected have been duration, number of occlusions (lingual contacts), and categorical ratings. Yet in reality, these variables are not independent of one another. That is, for any given token all these acoustic features, along with others, interrelate to various degrees and contribute to the perception of relative canonicity of the token and relative accentedness of the speaker. Therefore our methods should ideally allow us to make comparisons across multiple acoustic features simultaneously. This study was able to explore the rhotics produced by L2s, HSs, IMs, and traditional NSs taking into account multiple, interrelated and correlated acoustic features by employing a PCA. The results of the PCA suggested that it was difficult to discern many HSs from other NSs and IMs (Figure 1) when a variety of acoustic features was taken into account, as opposed to the relatively greater differences found in prior studies that have compared groups on one isolated acoustic feature at a time (for example, Table 7). The relatively new and understudied field of heritage phonetics could probably benefit from adopting more exploratory techniques such as PCAs to identify precisely which acoustic feature(s) will be of most interest in future research.

Another challenge of heritage phonetics research has been to estimate the effect of tasks. Some prior studies have employed only read-aloud tasks or have found across-group differences with read-aloud tasks but not less controlled tasks. Read-alouds have the distinct advantage of eliciting a controlled number of tokens in particular phonological environments, but such tasks “should be implemented with caution, crucially among Spanish heritage speakers” (Colantoni, Cuza & Mazzaro, 2016). This study took a different approach, which was to employ a variety of tasks along the continuum ranging from spontaneous to controlled tasks, in an effort to glean a more holistic view of HSs’ phonology with respect to rhotic phonemes. Again, the results of the PCA suggested that it was difficult to discern many HSs from other NSs and IMs (Figure 1) when data from a variety of tasks were taken into account, as opposed to the relatively greater differences found in prior studies that have compared groups on read-aloud tasks. It is possible that those apparent group differences are owed more to differences in L1 literacy among participants than to underlying phonological abilities. In the present study task effects were investigated only obliquely, in the RM ANOVAs that found a main effect for task and a few significant interactions of task and group. However, the results of the various analyses overall suggest that task effects warrant much more investigation in future heritage phonetics research.

Identity and Other Extralinguistic Factors

It has been argued that in addition to reporting differences between HSs and other groups, we should be looking within the HSs, seeking out factors that better explain their high degree of variability, because what appears to be inexplicable heterogeneity could be an “illusion of endless variation” (Polinsky & Kagan, 2007) owing to insufficient identification of the precise underlying variables that predict outcomes in HS phonetics. In this study, multiple instruments were used to capture a range of linguistic and extralinguistic factors involved in participants’ unique backgrounds. An exploratory PCA of these data found that a variety of factors contribute to participants’ backgrounds. In this analysis, the composite variable/component of all background data suggested that two clusters of HS background profiles were distinct from L2s and other NSs, though they were not distinct from IMs (Figure 2). The distinction between the two clusters of HS backgrounds was as much related to cultural and ethnic identities as variables relating to language experience (Table 6). Prior studies have found that differences in proficiency and language dominance are linked to variability in HS rhotic production. In the present study, proficiency and dominance were not good predictors, though both were self-reported and perhaps inaccurate. Rather, here HSs who produced rhotics more like traditional NSs tended to self-identity with Latino/Hispanic ethnicity or culture more than the HSs who produced rhotics more like L2s, whereas on other measures of proficiency and language experience they were quite similar. Two particular cases serve to illustrate this general pattern. Speaker hsmF produced rhotics like traditional NSs. Her language history (schooling, age of learning English, length of residence in the U.S., life stage transitioned to using more English), current use, and relative proficiency measures were close to the overall HS mean, yet her Cultural identity score was the highest of all HSs.[7] In contrast, speaker hssC's rhotics diverged greatly from traditional NSs. Her language history and proficiency were also similar to other HSs, but her Ethnic identity and Cultural identity scores were the lowest of the group.

Prior studies have found that Latinos in the U.S., like all speakers of all languages, employ particular phonetic features to construct multiple identities (e.g., Tseng, 2014; Bailey, 2000). Given that rhotics manifest vast sociophonetic variation across the Spanish-speaking world (e.g., Henriksen, 2015; Lipski, 2008), rhotics seem to be particularly relevant phonemes for expressing identities in Spanish. Heritage speakers may selectively retain and use rhotics as distinctive markers that indicate they both align with an ethnic identity and view that identity as having vitality, as Ethnolinguistic Identity Theory would suggest (Giles & Johnson, 1987). If indeed ethnic and/or cultural identities are important factors for understanding heritage Spanish phonology/ phonetics, the nature of the relationship merits exploration. Cultural ties could be the cause or the result of language usage behaviors or perhaps relate to speech production indirectly. For instance, stronger cultural identity could drive someone to seek more contact with other speakers or receive richer input. Stronger cultural identity could be related to a greater desire to sound "native" and could motivate a HS to pay close attention to phonetic features, perhaps particularly the salient and socially relevant features of the trill. Conversely, it is also possible that non-canonical production of rhotics could lead to weaker cultural identity rather than be a result of it. A HS who produces a noticeably nonstandard trill may be seen by others as a less legitimate member of the community or may internalize that articulatory difficulty as an outward sign of not being an authentic speaker. Helmer's ethnography of high school U.S. Spanish HSs demonstrated that native-like accent was perceived as a necessary element of cultural authenticity (2011; 2013), lending credence to this hypothesis. A HS whose nonstandard accent weakens her cultural identity could withdraw from the community over time and use Spanish less often. In this case, her language use (and subsequent proficiency, dominance, etc.) would not be the cause of accent but rather the effect. Oh and Au's (2005) study of U.S. Spanish HSs analyzed global accent ratings and ratings of stop consonants, finding positive but weak correlations with cultural identity and participation, but stronger correlations with language experience and use. The present study found a stronger effect of cultural identity, perhaps owing to differences in phonetic targets, instruments, and participants. However, the present study supports their argument that HSs' "mastery of the language may be more complicated than simple bivariate relationships" and that we must better understand the nature and relevance of particular background variables before we can construct a suitable sociocultural model of HS phonological acquisition (Oh & Au, 2005, p. 238). The present exploratory study cannot offer definitive answers yet, but it suggests that matters of identity should not be ignored.

In much recent research on Spanish heritage phonetics, identity factors are either not considered or are included in a cursory way and framed narrowly as a matter of *language* identity and use. Attitudes towards language are often measured, but as Schmid and Dusseldorp (2010) found in their PCA of L1 attrition, exposure to and attitudes towards language (L1) constituted a factor quite distinct from identification and affiliation with L1 culture. Future research could probe the potential role of identity in heritage phonetics in various ways. The instruments used here (AMAS and MEIM) are just two of dozens of instruments available to researchers (Center for substance Abuse Treatment, 2014), many of which have been validated for first and second generation Spanish speakers in the U.S. These instruments warrant further exploration in heritage phonetics research. Furthermore, it will be important not only to identify better instruments for understanding HSs' experiences but also to carefully consider how particular survey items are framed and how they might be interpreted by HSs. For instance, the item of the Bilingual Language Profile that related to culture states, "I identify with a Spanish-speaking culture." This phrasing is likely not equivalent to asking if and how one identifies as a member of a self-selected cultural or ethnic heritage group (as on the MEIM or AMAS). The present study also employed the LEAP-Q, an instrument used in many similar prior studies (e.g. Henriksen, 2015) and found that some items could be interpreted differently by HSs. For instance, an item about the number of years participants had lived in a Spanish-speaking country, which seems to be an objective measure on its surface, in fact revealed some participants' subjective perceptions of their imagined speech communities. Speaker hssB reported having never lived in a Spanish-speaking country. Speaker hsmF, who was indistinguishable from NSs in rhotic production, reported having lived half of her life in a Spanish-speaking country. Yet both participants had in fact lived in Brooklyn, New York, until approximately age 18. The questionnaire item seemed for these participants not to elicit an objective measure of language experience but rather a construction of community identity. Indeed, it may be that many instruments currently used to describe HSs' experiences warrant closer item analyses. In their L1 attrition study, Schmid and Dusseldorp (2010) argued that "very global statements about L1 use may not be valid indications of actual language use" (p. 139) and that L1 use is more complex than generally recognized. They reported that use of L1 in contexts that seem equivalent (e.g., use of L1 with partner and with friends) in fact exhibited low correlations (p. 149). Perhaps the instruments being used and our often coarse grained analysis of the responses are limiting our ability to find patterns in the relationships between language experience and phonetic outcomes. The same can be argued of proficiency, which is interrelated with input, exposure, and use of the heritage language "in ways that may not yet be fully understood" (Polinsky & Kagan, 2007). Yet all these factors should be important in predictive models of heritage phonetics.

Limitations

Though the current study attempted to advance methods used in the field of HS phonetics, it left many questions unanswered and presented a number of important limitations. The sample size was adequate for the analyses undertaken but was relatively small, owing to the difficulty of recruiting speakers at the location of the study. In an attempt to limit dialectal variation, most traditional NSs recruited lived in the same state of Mexico, which may have contributed to that group's relative homogeneity in ways other than dialect. Like most studies on heritage phonetics, speech samples were not collected from HSs' families and thus the study did not compare HSs' production to the Spanish they were actually exposed to, which is the source variety that ideally should be considered the baseline for HSs' speech (Henriksen, 2015; Polinsky & Kagan, 2007; Pascual y Cabo & Rothman, 2012; Rao, 2016a; Rao & Ronquest, 2015, p. 407). The nature of the experimental tasks employed required that all the HSs recruited for this study be proficient enough in Spanish to tell a story, read a list of idioms, and speak extemporaneously for at least three minutes; the results may have been different had the participants been less proficient.

Rhotics were selected as the phonetic targets because they are typically the last phones acquired by Spanish-speaking children, but they are also some of the most salient and socially relevant phones in Spanish, meaning that the same study carried out on different phones may have produced different results. The study did not include ratings of global accent or ratings of phoneme accent, and so the acoustic analysis presented should not be interpreted as indicating how these rhotics would be perceived by others. In this study, the number of lingual contacts detected through auditory analysis was greater overall than those detected through acoustic analysis, but the acoustic data were included in subsequent analyses in keeping with conventions in the field. Future research should explore the actual percept of accent related to the production of rhotics. Since accent rating is a function of speaker and listener skills, such research should be careful to select listener/raters appropriate for the context (Spanish in the U.S.).

Though identity is a fluid construct, the AMAS and MEIM treat cultural and ethnic identity as relatively stable characteristics. In this way the present study is similar to the first wave of variationist studies, which operationalized identity as a pre-existing and stable construct and understood sociophonetic variation as signifying group membership. Second wave studies recognized the fluidity of identities, and third wave studies recognized that sociophonetic variation could index stances in context-specific ways (see Drager, 2015, for a review). The present study is just a first step toward understanding the role of identity in HS phonetics. Future research should aim to elucidate how sociophonetic variation among HSs can serve all these functions simultaneously, investigating, for instance, which features of rhotics carry more overt and covert prestige in U.S. Spanish, as well as to what extent some HSs who acquire canonical /r/ may selectively produce it in particular contexts.

Future research might also expand the set of potentially relevant predictor variables beyond the set that was included in the present study. For instance, studies of L1 attrition suggest that frequency of code switching, both in terms of an individual's usage (Łyskawa et al., 2016) and in terms of community norms (De Leeuw, Schmid, & Mennen, 2010), can predict L1 accent. Another potential variable is aptitude for pronunciation. Aptitude has been found to have a predictive role in the L1 grammatical proficiency maintained by L1 Spanish-L2 Swedish consecutive bilinguals (Bylund, Abrahamsson, & Hyltenstam, 2010).

CONCLUSION

Returning now to the issues noted in the introduction, research on heritage language speakers is important for theory building because it has the potential to increase our understanding of possible sensitive periods and the purported native/non-native divide. What the present study suggests is that the mainstream SLA definition of HSs--those who have used an ethno-linguistically minority home language since birth and maintain proficiency into adulthood--may not be an identifier incisive enough to best elucidate meaningful patterns in heritage acquisition of phonology/phonetics. The present study suggests that the language experience measures typically employed in the field to explain heritage phonological acquisition with "simple bivariate relationships" (Oh & Au, 2005, p. 238) might in fact obscure underlying sociocultural factors, perhaps especially in contexts where HSs are not surrounded by a vibrant bilingual speech community. The HSs in the present study constituted two subgroups that differed in cultural identity. If one aimed to circumscribe a core "heritage" group here, it would seem reasonable to characterize those with stronger cultural identity as *the* heritage speakers, given that the term "heritage" inherently evokes notions of cultural connection. Those with little cultural affiliation might be better characterized with more linguistically descriptive and less culturally loaded terms, such as "early, consecutive, subtractive bilinguals." In the present study the (true?) "heritage" speakers produced rhotics that were indistinguishable from other native speakers and long-term immigrants, when multiple acoustic features were analyzed simultaneously. This result presents a challenge to the narrative emerging about U.S. Spanish heritage speakers in phonetics research, which describes them as not quite like "native speakers."

Acknowledgements

This work was supported by a grant from the Faculty Research Committee of the University of Richmond's School of Arts and Sciences. The study would not have been possible without the assistance of Maya Barzilai and all the participants. Thanks also are due to Nick Pandža and several anonymous reviewers of the manuscript for their helpful comments. All shortcomings are my own.

References

Amengual, M. (2016). Acoustic correlates of the Spanish tap-trill contrast: Heritage and L2 Spanish speakers. *Heritage Language Journal*, *13*(2), 88–112. Available from http://www.heritagelanguages.org

Anderson, B. (1983). *Imagined communities.* New York, NY: Verso.

Asherov, D., Fishman, A., & Cohen, E. (2016). Vowel reduction in Israeli heritage Russian. *Heritage Language Journal*, *13*(2), 113–133. Available from http://www.heritagelanguages.org

Au, T. K. F., Knightly, L. M., Jun, S. A. & Oh, J. S. (2002). Overhearing a language during childhood. *Psychological Science*, *13*(3), 238–243.

Avery, D. R., Tonidandel, S., & Thomas, K. M. (2007). Assessing the Multigroup Ethnic Identity Measure for measurement equivalence across racial and ethnic groups. *Educational and Psychological Measurement*, *67*(5), 877–888.

Bailey, B. (2000). The language of multiple identities among Dominican Americans. *Journal of Linguistic Anthropology*, *10*(2), 190–223.

Benmamoun, E., Montrul, S., & Polinsky, M. (2013). Heritage languages and their speakers: Opportunities and challenges for linguistics. *Theoretical Linguistics*, *39*(3–4), 129–181.

Birdsong, D. (2007). Nativelike pronunciation among late learners of French as a second language. In O. S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning* (pp. 99–116). Philadelphia, PA: John Benjamins.

Boersma, P. & Weenink, D. (2015). Praat: doing phonetics by computer [Computer program], Version 6.0.05. Retrieved from http://www.praat.org/

Bradley, T. G. & Willis, E. W. (2012). Rhotic variation and contrast in Veracruz Mexican Spanish. *Estudios de fonetica experimental*, *21*(1), 43–74.

Bylund, E., Abrahamsson, N., & Hyltenstam, K. (2010). The role of language aptitude in first language attrition: The case of pre-pubescent attriters. *Applied Linguistics, 31*(3), 443–464.

Carballo, G. & Mendoza, E. (2000). Acoustic characteristics of trill productions by groups of Spanish children. *Clinical Linguistics and Phonetics, 14*(8), 587–601.

Cattell, R. B. (l966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*, 245–276.

Chang, C. B., & Yao, Y. (2016). Toward an understanding of heritage prosody: Acoustic and perceptual properties of tone produced by heritage, native, and second language speakers of Mandarin. *Heritage Language Journal*, *13*(2), 134–160. Available from http://www.heritagelanguages.org

Colantoni, L., Cuza, A., & Mazzaro, N. (2016). Task related effects in the prosody of Spanish heritage speakers. In M. Amstrong, N., Henrikssen & M. Vanrell, (Eds.), *Intonational grammar in Ibero-Romance* (pp. 3–24). Amsterdam, Netherlands: John Benjamins.

De Leeuw, E., Mennen, I., & Scobbie, J. M. (2012). Singing a different tune in your native language: First language attrition of prosody. *International Journal of Bilingualism*, *16*(1), 101–116.

De Leeuw, E., Mennen, I., & Scobbie, J. M. (2013). Dynamic systems, maturational constraints and L1 phonetic attrition. *International Journal of Bilingualism*, *17*(6), 683–700.

De Leeuw, E., Schmid, M. S. & Mennen, I. (2010). The effects of contact on native language pronunciation in an L2 migrant setting. *Bilingualism: Language and Cognition*, *13*(10), 33–40.

Drager, K. K. (2015). Linguistic Variation, Identity Construction and Cognition. Berlin, Germany: Language Science Press.

Fabiano-Smith, L. & Goldstein, B. A. (2010). Early-, middle-, and late-developing sounds in monolingual and bilingual children: An exploratory investigation. *American Journal of Speech-Language Pathology*, *19*(1), 66–77.

Face, T. (2003). Intonation in Spanish declaratives: Differences between lab speech and spontaneous speech. *Catalan Journal of Linguistics*, *2*, 115–131.

Field, A. (2005). *Discovering statistics using SPSS* (2nd ed.). London, England: Sage.

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.

Flores, C. & Rato, A. (2016). Global accent in the Portuguese speech of heritage returnees. *Heritage Language Journal*, *13*(2), 161–183. Available from http://www.heritagelanguages.org

Fuertes, J. N., Gottdiener, W. H., Martin, H., Gilbert, T. C., & Giles, H. (2012). A meta-analysis of the effects of speakers' accents on interpersonal evaluations. *European Journal of Social Psychology*, *42*, 120–133.

Giles, H., & Johnson, P. (1987). Ethnolinguistic identity theory: A social psychological approach to language maintenance. *International Journal of the Sociology of Language, 68*, 69–99.

Goodin-Mayeda, C. E. (2016). Acoustic analysis of the tap/trill contrast by heritage and native speakers of Spanish. Paper presented at the Current Approaches to Spanish and Portuguese Second Language Phonology Conference. Columbus, OH.

Hay, J. & Drager, K. (2007). Sociophonetics. *Annual Review of Anthropology*, *36*, 89–103.

Helmer, K. A. (2011). "'Proper' Spanish is a waste of time": Mexican-origin student resistance to learning Spanish as a heritage language. In L. Scherff, & K. Spector (Eds.), *Culturally relevant pedagogy: Clashes and confrontations* (pp. 135–164). Blue Ridge Summit, PA: Rowman and Littlefield Education.

Helmer, K. A. (2013). A twice-told tale: voices of resistance in a borderlands Spanish heritage language class. *Anthropology and Education Quarterly*, *44*(3), 269–285.

Henriksen, N. C. (2015). Acoustic analysis of the rhotic contrast in Chicagoland Spanish: An intergenerational study. *Linguistic Approaches to Bilingualism*, *5*(3), 285–321.

Hirson, A., & Sohail, N. (2007). *Variability of rhotics in Punjabi-English bilinguals*. Paper presented at the 16th International Congress of Phonetic Sciences (ICPhS), Saarbrücken, Germany. Available at: http://www.icphs2007.de/conference/Papers/1623/1623.pdf

Hoot, B. (2012). *Presentational focus in heritage and monolingual Spanish* (Unpublished doctoral dissertation). University of Illinois at Urbana Champaign.

Howell, D. C. (2002). *Statistical methods for psychology*. Pacific Grove, CA: Duxbury/Thomson Learning.

Hutcheson. G., & Sofroniou. N. (1999). *The multivariate social scientist*. London, England: Sage.

Center for Substance Abuse Treatment (U.S.). (2014). Instruments to Measure Identity and Acculturation. Appendix B of *Improving Cultural Competence. Treatment Improvement Protocol* (TIP) Series No. 59. HHS Publication No. (SMA) 14–4849. Rockville, MD: Substance Abuse and Mental Health Services Administration. Available from https://www.ncbi.nlm.nih.gov/books/NBK248425/

Kang, Y., George, S., & Soo, R. (2016). Cross-language influence in the stop voice contrast in heritage Tagalog. *Heritage Language Journal*, *13*(2), 184–218. Available from http://www.heritagelanguages.org

Kaiser. H. F. (1960). The application of electronic computer to factor analysis. *Educational and Psychological Measurement*, *20*, 141–151.

Kaiser. H.F. (1970). A second-generation little jiffy. *Psychometrika*, *35*, 401–415.

Kaiser. H. F. (1974). An index of factorial simplicity. *Psychometrika*, *39*, 31–36.

Leeman, J. (2015). Heritage language education and identity in the United States. *Annual Review of Applied Linguistics*, *35*, 100–119.

Le Page, R. B., & Tabouret-Keller, A. (1985). *Acts of identity: Creole-based approaches to language and ethnicity*. Cambridge, UK: Cambridge University Press.

Lipski, J. M. (2008). *Varieties of Spanish in the United States*. Washington, DC: Georgetown University Press.

Łyskawa, P. Maddeaux, R., Melara, E. & Nagy, N. (2016). *Heritage Language Journal, 13*(2), 219–244. Available from http://www.heritagelanguages.org

Major, R. C. (1992). Losing English as a first language. *The Modern Language Journal*, *76*, 190–208.

Markham, D. (1997). *Phonetic imitation, accent, and the learner*. Lund, Sweden: Lund University Press.

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research, 50*(4), 940–967.

Mazzaro, N., Cuza, A., & Colantoni, L. (2016). Age effects and the discrimination of consonantal and vocalic contrasts in heritage and native Spanish. In C. Tortora, M. den Dikken, I. L. Montoya, & T. O'Neill (Eds.), *Romance Linguistics 2013: Selected papers from the 43rd Linguistic Symposium on Romance Languages (LSRL) New York, 17–19 April 2013*, vol. 9 (pp. 277–300). Amsterdam, Netherlands; John Benjamins.

Mayer, M. (1969). *Frog, where are you*? New York, NY: Dial Press. Available from https://www.phil-fak.uni-duesseldorf.de/fileadmin/Redaktion/Institute/Allgemeine_Sprachwissenschaft/Frogstory-2_01.pdf

Miyoshi, M., Asner-Self, K., Yanyan, S., & Koran, J. M. (2017). Psychometric properties of the Abbreviated Multidimensional Acculturation Scale and the Multigroup Ethnic Identity Measure with Japanese sojourners. *Assessment*, *24*(1), 104–114.

Montrul, S. (2010). Current issues in heritage language acquisition. *Annual Review of Applied Linguistics*, *30*(1), 3–23.

Mrak, N. A. (2011). Heritage speakers and the standard: Fighting linguistic hegemony. In L. A. Ortiz-López (Ed.), *Selected proceedings of the 13th Hispanic linguistics symposium* (pp. 161–168). Somerville, MA: Cascadilla Press.

Oh, J. S. & Au, T. K. F. (2005). Learning Spanish as a heritage language: The role of sociocultural background variables. *Language, Culture and Curriculum, 18*(3), 229–241.

Pascual y Cabo, D., & Rothman, J. (2012). The (Il)logical problem of heritage speaker bilingualism and incomplete acquisition. *Applied Linguistics, 33*(4), 450–455.

Phinney, J. (1992). The Multigroup Ethnic Identity Measure: A new scale for use with adolescents and young adults from diverse groups. *Journal of Adolescent Research, 7*, 156–176.

Polinsky, M., & Kagan, O. (2007). Heritage languages: In the 'wild' and in the classroom. *Language and Linguistics Compass, 1*(5), 368–395.

Quesada Pacheco, M. A. (2010). *El español hablado en América Central: Nivel fonético*. Madrid, Spain: Iberoamericana Vervuert.

Rakić, T., Steffens, M. C., & Mummendey, S. (2011). Blinded by the accent! The minor role of looks in ethnic categorization. *Journal of Personality and Social Psychology, 100*(1), 16–29.

Rao, R. (2016a). *Growing yet relatively unexplored: The case of heritage Spanish phonology*. Paper presented at the Current Approaches to Spanish and Portuguese Second Language Phonology Conference. Columbus, OH.

Rao, R. (2016b). Introduction to special issue on heritage phonetics and phonology. *Heritage Language Journal*, *13*(2), i-ii. Available from http://www.heritagelanguages.org

Rao, R. (2009). Deaccenting in spontaneous speech in Barcelona Spanish. *Studies in Hispanic and Lusophone Linguistics*, *2*(1), 31–75.

Rao, R., & Ronquest, R. (2015). The heritage Spanish phonetic/phonological system: Looking back and moving forward. *Studies in Hispanic and Lusophone Linguistics, 8*(2), 403–414.

Reilly, J., Losh, M., Bellugi, U., & Wulfeck, B. (2004). "Frog, where are you?" Narratives in children with specific language impairment, early focal brain injury, and Williams syndrome. *Brain and Language, 88*, 229–247.

Riazi, A. M. (2016). *The Routledge encyclopedia of research methods in applied linguistics*. New York, NY: Routledge.

Ronquest, R. (2016). Stylistic variation in heritage Spanish vowel production. *Heritage Language Journal*, *13*(2), 275–297. Available from http://www.heritagelanguages.org

Ronquest, R. (2013). An acoustic examination of unstressed vowel reduction in heritage Spanish. In C. Howe, S. E. Blackwell & M. L. Quesada (Eds.), *Selected proceedings of the 15th Hispanic Linguistics Symposium* (pp. 151–171). Somerville, MA: Cascadilla Press.

Ronquest, R. (2012). *An Acoustic Analysis of Heritage Spanish Vowels* (Unpublished doctoral dissertation). Indiana University. Retrieved from http://www.academia.edu/13285559/An_Acoustic_Analysis_of_Heritage_Spanish_Vowels

Rothman, J. & Treffers, J. (2014). A prolegomenon to the construct of the native speaker: Heritage speakers are natives too! *Applied Linguistics, 35*(1), 93–98.

Sánchez-Muñoz, A. (2013). Identidad y confianza lingüística en jóvenes latinos en el sur de California. In D. Dumitrescu y G. Piña Rosales (Eds.), *El español en los Estados Unidos: E Pluribus Unum?* (pp. 217–232). New York, NY: North American Academy of the Spanish Language.

Schmid, M. S. & Dusseldorp, E. (2010). Quantitative analyses in a multivariate study of language attrition: The impact of extralinguistic factors. *Second Language Research, 26*(1), 125–160.

Tseng, A. (2014). Vowel variation, style, and identity construction in the English of Latinos in Washington, D.C. (Unpublished doctoral dissertation). Georgetown University, Washington, DC. Retrieved from https://repository.library.georgetown.edu/handle/10822/1029902

US Census Bureau. (2016). Quick Facts [web page]: Richmond, Virginia. Available at https://www.census.gov/quickfacts/

Yamada, A. M., Valle, R., Barrio, C., & Jeste, D. (2006). Selecting an acculturation measure for use with Latino older adults. *Research on Aging*, *28*(5), 519–561.

Yap, S. C. Y., Donnellan, M. B., Schwartz, S. J., Zamboanga, B. L., Kim, S. Y., Huynh, Q. L., Vazsonyi, A. T., Cano, M. A., Hurley, E. A., Whitbourne, S. Krauss, Castillo, L. G., Donovan, R. A., Blozis, S. A., & Brown, E. J. (2016). Evaluating the Invariance of the Multigroup Ethnic Identity Measure Across Foreign-Born, Second-Generation and Later-Generation College Students in the United States. *Cultural Diversity and Ethnic Minority Psychology*, *22*(3), 460–465.

Zea, M. C., Asner-Self, K. K., Birman, D., & Buki, L. P. (2003). The abbreviated multidimensional acculturation scale: Empirical validation with two Latino/Latina samples. *Cultural Diversity and Ethnic Minority Psychology*, *9*(2), 107–126.

Appendix A. Idioms for Read-Aloud Task, with Target Features Underlined

	Spanish Idiom	Approximate English equivalent
1	*Piensa el ladrón que todos son de su condición.* 'Think the thief that all are of his condition'	A thief believes everyone steals.
2	*A palabras necias oídos sordos.* 'To words foolish ears deaf'	Don't listen to foolish words.
3	*Con paciencia y la maña el elefante se comió la araña.* 'With patience and skill the elephant ate the spider'	Little strokes fell great oaks.
4	*Lo barato cuesta caro.* 'The cheap costs expensive'	If you pay cheaply, you pay dearly.
5	*No es rico el que más tiene sino el que menos quiere.* 'Not is rich he who more has but he who least wants'	The rich are not those who have the most but those who need the least.
6	*Perro que ladra no muerde.* 'Dog that barks no bites'	All bark and no bite.
7	*Quien a buen árbol se arrima buena sombra le cobija.* 'Who to good tree comes closer good shadow him covers'	Whoever leans close to a good tree is covered by a good shadow.
8	Con el viento se limpia el trigo y los vicios con castigo. 'With the wind cleans the wheat and the vices with punishment'	Vices are only killed with punishment.
9	*A burro muerto la cebada al rabo.* 'To donkey dead the barley to the tail'	No good being wise after the event.
10	*De la mar el mero y de la tierra el cordero.* 'From the sea the grouper and from the land the lamb'	The best fish and the best meat.
11	*A río revuelto ganancia de pescadores.* 'To river rough profit of fishermen'	Troubled waters, fisherman's gain.
12	*En el país de los ciegos el tuerto es el rey.* 'In the country of the blind the one-eyed person is the king'	The blind leading the blind
13	*Cada uno sabe dónde le aprieta el zapato.* 'Each one knows where him squeezes the shoe'	Each person knows where his problems lie.
14	*A todos les llega su momento de gloria.* 'To all them arrives their moment of glory'	Every dog has its day.
15	*Ni tanto ni tan poco… marido loco.* 'Not as much nor as little … husband crazy'	Neither one extreme nor the other.
16	*Menea la cola el can no por ti sino por el pan.* 'Waves the tail the dog not for you rather for the bread'	Affection may be more self-interest than love.
17	*Cuando uno tiene hambre todo sabe bueno.* 'When one has hunger all tastes good'	Hunger is the best sauce.
18	*A buena obra buen pagar.* 'To good work good pay'	Good work is rewarded.
19	*A beber y a tragar que el mundo se va a acabar.* 'To drink and to swallow that the world itself ends'	Eat, drink, and be merry (for tomorrow we die).
20	*Camarón que se duerme se lo lleva la corriente.* 'Shrimp that falls asleep it is taken the current'	You snooze, you lose.
21	*Pica, lica y califica.* 'Peck, verify and qualify'	Look before you leap.
22	*Cielo aborregado suelo mojado.* 'Sky mackerel ground wet'	Mackerel sky means rain will come.

APPENDIX B. EXAMPLE RHOTICS WITH WAVEFORM AND SPECTROGRAM

Figure 3. Canonical intervocalic phonemic tap production containing one lingual contact based on acoustic cues (occlusion, marked by "*") and auditory cues in the word *araña* 'spider' (tap duration = 33 ms)

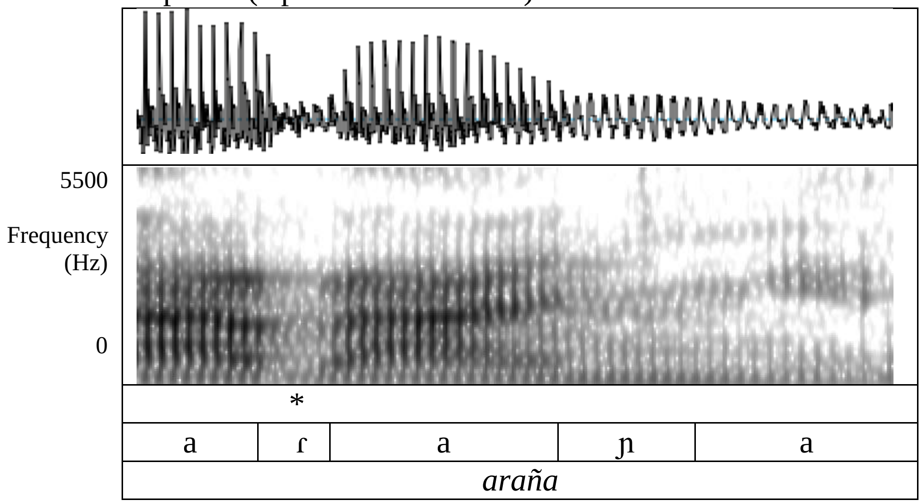


Figure 4. Canonical intervocalic phonemic trill production containing two lingual contacts based on acoustic cues (occlusions, marked by "*"), and three lingual contacts based on auditory cues, followed by a vowel-like portion (marked by "^") in the word *corrió* 'he ran' (trill duration = 94 ms)

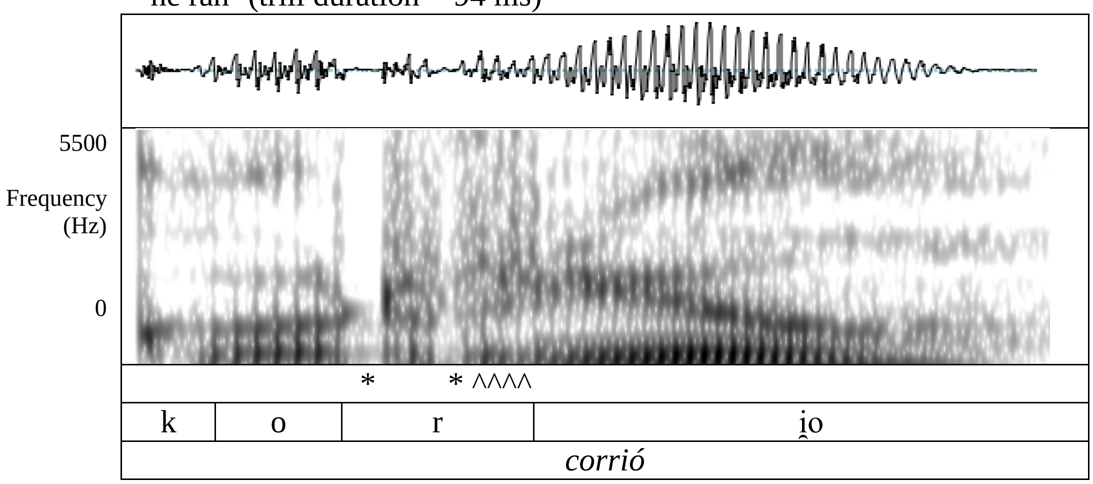


Figure 5. Fricative word-initial phonemic trill production containing zero lingual contacts and non-periodic noise throughout in the word *ranas* 'frogs' (fricative duration = 84 ms)

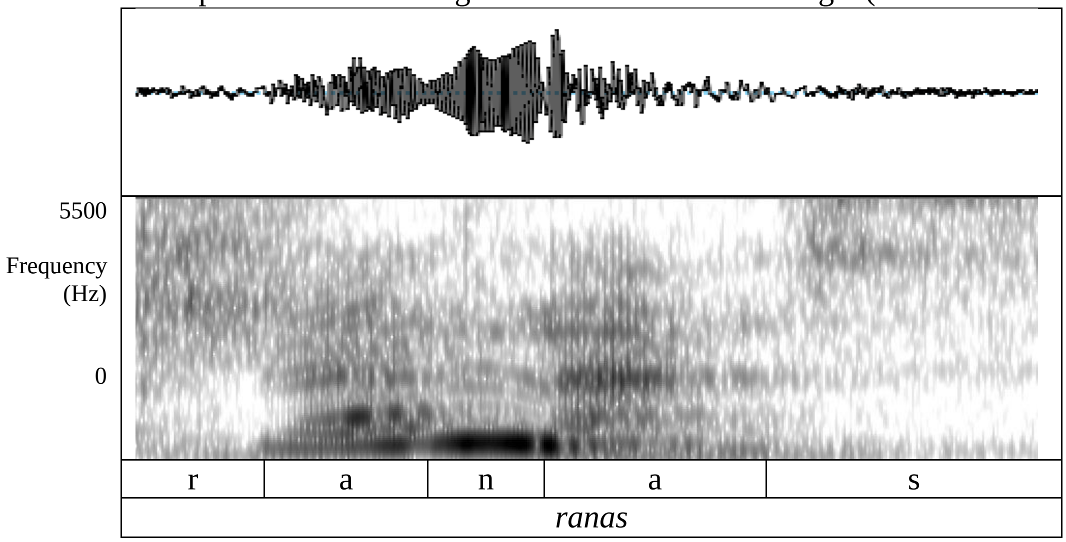


Figure 6. Approximant word-initial phonemic trill production containing zero lingual contacts and with clear formant structures present throughout in the word *regresaba* 'he was returning' (approximant duration = 90 ms)

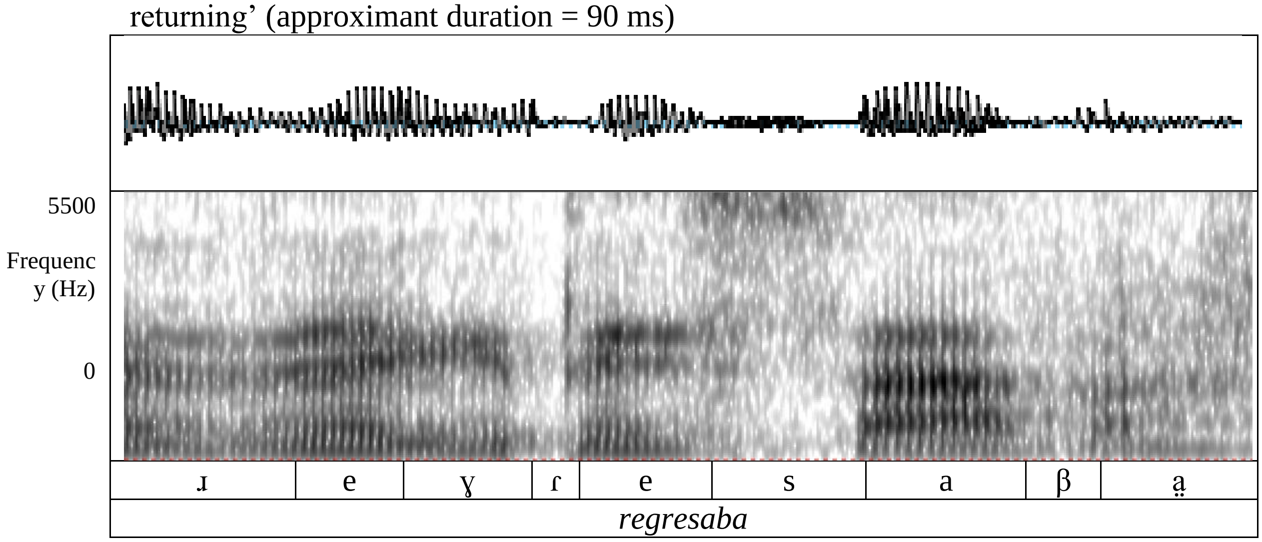


Figure 7. Perceptual intervocalic phonemic tap production containing one lingual contact based on auditory cues and zero reliable acoustic landmarks in the word *salieron* 'they left' (duration cannot be measured)

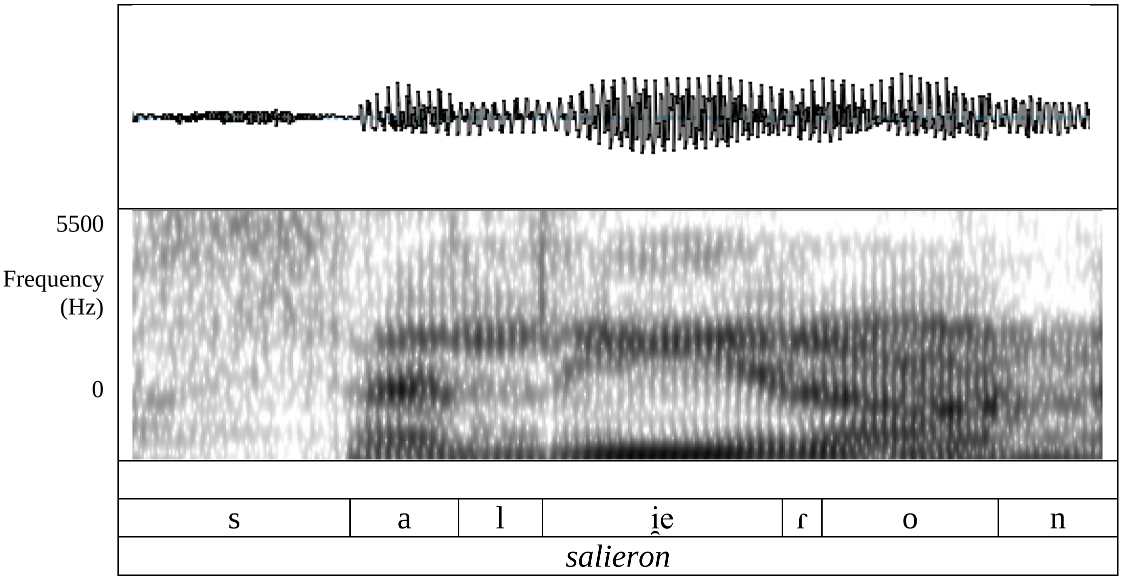


Figure 8. Elided intervocalic phonemic tap production containing zero auditory or acoustic evidence of rhotic production in the word *camarón* 'shrimp' (duration cannot be measured)

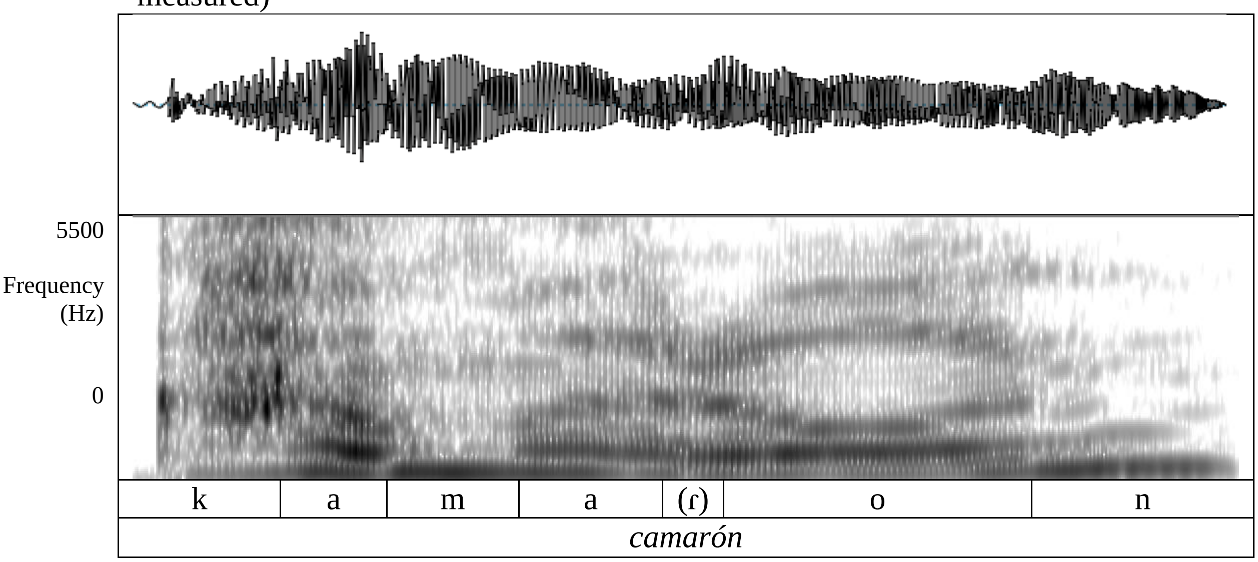


Figure 9. Noncanonical word-final phonemic trill production containing one lingual contact based on acoustic cues (occlusion, marked by “*”) and auditory cues, as well as 69 ms of aperiodic noise (frication, marked by “~”) in the word *dormir* ‘to sleep’ (trill duration = 79 ms)

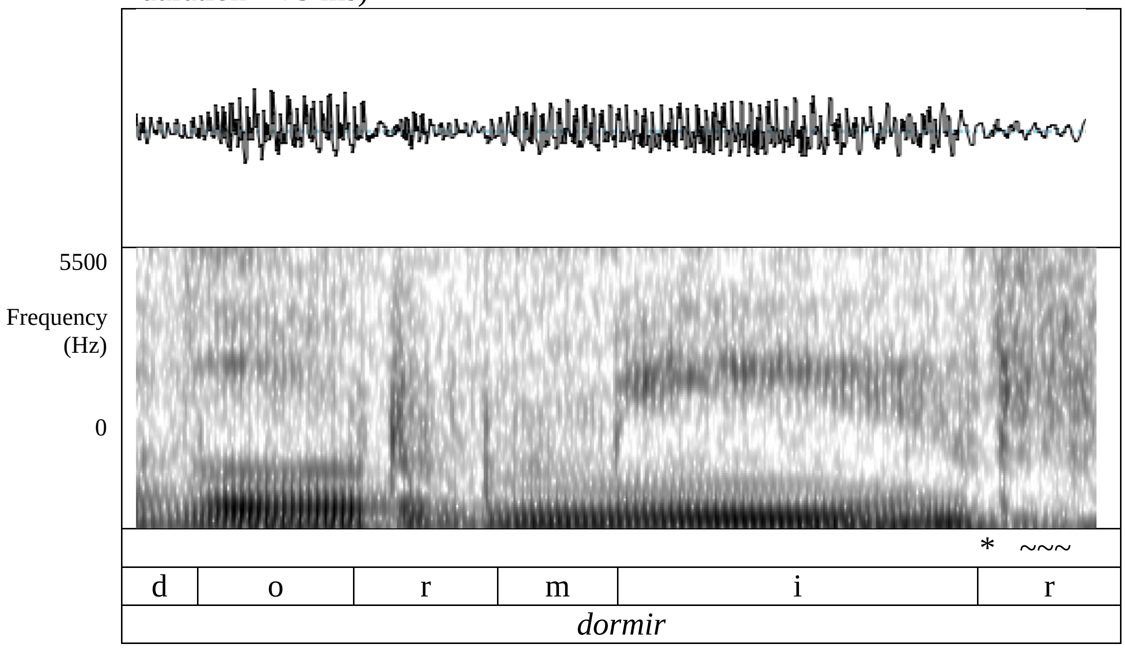
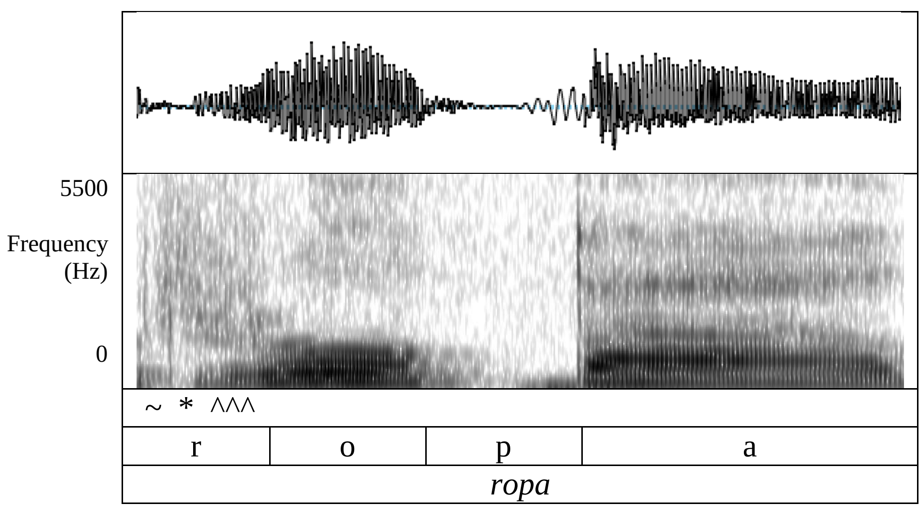


Figure 10. Noncanonical word-initial phonemic trill production containing 36 ms of aperiodic noise (frication, marked by “~”) and one lingual contact (occlusion, marked by “*”) followed by a 71 ms vowel-like portion (marked by “^”) in the word *ropa* ‘clothes’ (overall duration = 128 ms)



Appendix C. Tables

Table 9.

RMANOVA of Duration, Groups Determined Deductively

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2	*Power*
Between subjects							
Intercept	577680.73	1	577680.73	648.83	< .001	.95	1.00
Group (L2, HS, IM, NS)	10289.51	3	3429.84	3.85	.018	.27	.77
Error	28491.16	32	890.35				
Within Subjects							
Phone (Trill, Tap)	55330.90	1	55330.90	217.83	< .001	.87	1.00
Phone × Group	4399.41	3	1466.47	5.77	.003	.35	.92
Error (Phone)	8128.24	32	254.01				
Task (1, 2, 3)	709.66	2	354.83	3.22	.046	.09	.60
Task × Group	1234.66	6	205.78	1.87	.100	.15	.65
Error (Task)	7044.88	64	110.08				
Task × Phone	404.38	2	202.19	1.84	.167	.05	.37
Task × Phone × Group	347.38	6	57.90	0.53	.786	.05	.20
Error (Task × Phone)	7038.74	64	109.98				

Table 10.

LSD Post Hoc Tests of Group Differences in Duration (Deductive Approach)

Groups		*Mean Difference*	*Std. Error*	*p*	*95% CI*	
HS	L2	15.19	5.88	.015	3.21	27.18
	IM	1.99	6.29	.753	-10.82	14.81
	NS	-5.98	4.97	.238	-16.11	4.15
IM	L2	13.20	7.38	.083	-1.83	28.22
	NS	-7.97	6.67	.241	-21.56	5.62
NS	L2	21.17	6.29	.002	8.36	33.99

Table 11.

RMANOVA of Duration, Groups Determined Inductively

Source		*SS*	*df*	*MS*	*F*	*p*	η_p^2	*Power*
Between subjects								
	Intercept	644542.78	1	644542.78	690.43	< .001	.96	1.00
	Group (L2, HS-ID, HS+ID, IM, NS)	10770.64	4	2692.66	2.88	.039	.28	.71
	Error	28006.29	30	933.54				
Within Subjects								
	Phone (Trill, Tap)	60238.15	1	60238.15	261.53	< .001	.90	1.00
	Phone × Group	5616.21	4	1404.05	6.10	.001	.45	.97
	Error (Phone)	6909.89	30	230.33				
	Task (1, 2, 3)	1221.29	2	610.65	5.51	.006	.16	.83
	Task × Group	1118.34	8	139.79	1.26	.281	.14	.53
	Error (Task)	6656.01	60	110.93				
	Task × Phone	545.73	2	272.87	2.48	.093	.08	.48
	Task × Phone × Group	577.71	8	72.21	0.66	.728	.08	.28
	Error (Task × Phone)	6610.00	60	110.17				

Table 12.

LSD Post hoc Tests of Group Differences in Duration (Inductive Approach)

Groups		*Mean difference*	*Std. Error*	*p*	*95% CI*	
HS-ID	L2	17.30	6.74	.015	3.54	31.06
	HS+ID	4.85	6.74	.477	-8.91	18.61
	IM	4.10	7.11	.569	-10.42	18.62
	NS	-3.87	5.92	.518	-15.96	8.21
HS+ID	L2	12.45	7.20	.094	-2.26	27.16
	IM	-0.75	7.55	.921	-16.18	14.67
	NS	-8.72	6.44	.186	-21.88	4.43
IM	L2	13.20	7.55	.091	-2.23	28.62
	NS	-7.97	6.83	.252	-21.93	5.98
NS	L2	21.17	6.44	.003	8.02	34.33

Table 13.

RMANOVA of Lingual Contacts, Groups Determined Deductively

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2	*Power*
Between subjects							
Intercept	435.29	1.00	435.29	550.73	< .001	.95	1.00
Group (L2, HS, IM, NS)	13.96	3.00	4.65	5.89	.003	.36	.93
Error	25.29	32.00	0.79				
Within Subjects							
Phone (Trill, Tap)	41.37	1.00	41.37	102.27	< .001	.76	1.00
Phone × Group	6.17	3.00	2.06	5.08	.005	.32	.88
Error (Phone)	12.94	32.00	0.40				
Evidence (Acoustic, Auditory)	5.45	1.00	5.45	186.38	< .001	.85	1.00
Evidence × Group	0.05	3.00	0.02	0.57	.641	.05	.15
Error (Evidence)	0.94	32.00	0.03				
Phone × Evidence	0.04	1.00	0.04	1.25	.272	.04	.19
Phone × Evidence × Group	0.09	3.00	0.03	0.87	.465	.08	.22
Error (Phone × Evidence)	1.04	32.00	0.03				
Task (1, 2, 3)	4.55	1.62	2.82	21.17	.000	.40	1.00
Task × Group	0.36	4.84	0.07	0.55	.730	.05	.19
Error (Task)	6.88	51.67	0.13				
Task × Phone	2.40	1.69	1.42	12.92	.000	.29	.99
Task × Phone × Group	0.27	5.08	0.05	0.49	.784	.04	.17
Error (Task × Phone)	5.95	54.21	0.11				
Task × Evidence	0.07	2.00	0.04	2.12	.128	.06	.42
Task × Evidence × Group	0.05	6.00	0.01	0.48	.819	.04	.18
Error (Task × Evidence)	1.08	64.00	0.02				
Task × Phone × Evidence	0.01	1.66	0.00	0.20	.783	.01	.08
Task × Phone × Evidence × Group	0.02	4.98	0.00	0.20	.960	.02	.09
Error (Task × Phone × Evidence)	1.07	53.08	0.02				

For LSD Post hoc tests of group differences in lingual contacts (deductive approach), see Table 7 in text.

Table 14.

RMANOVA of Lingual Contacts, Groups Determined Inductively

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2	*Power*
Between subjects							
Intercept	498.41	1.00	498.41	676.21	< .001	.96	1.00
Group (L2, HS-ID, HS+ID, IM, NS)	15.25	4.00	3.81	5.17	.003	.41	.94
Error	22.11	30.00	0.74				
Within Subjects							
Phone (Trill, Tap)	47.97	1.00	47.97	130.71	< .001	.81	1.00
Phone × Group	6.17	4.00	1.54	4.20	.008	.36	.88
Error (Phone)	11.01	30.00	0.37				
Evidence (Acoustic, Auditory)	6.27	1.00	6.27	210.35	< .001	.88	1.00
Evidence × Group	0.09	4.00	0.02	0.75	.569	.09	.21
Error (Evidence)	0.90	30.00	0.03				
Task (1, 2, 3)	5.11	1.62	3.16	22.45	< .001	.43	1.00
Task × Group	0.37	6.47	0.06	0.40	.885	.05	.16
Error (Task)	6.84	48.52	0.14				
Task × Phone	2.77	2.00	1.39	14.24	< .001	.32	1.00
Task × Phone × Group	0.28	8.00	0.04	0.36	.939	.05	.16
Error (Task × Phone)	5.84	60.00	0.10				
Task × Evidence	0.08	2.00	0.04	2.38	.101	.07	.46
Task × Evidence × Group	0.08	8.00	0.01	0.58	.788	.07	.25
Error (Task × Evidence)	1.02	60.00	0.02				
Phone × Evidence	0.09	1.00	0.09	2.69	.111	.08	.36
Phone × Evidence × Group	0.14	4.00	0.03	1.05	.399	.12	.29
Error (Phone × Evidence)	0.98	30.00	0.03				
Task × Phone × Evidence	0.01	1.66	0.01	0.29	.711	.01	.09
Task × Phone × Evidence × Group	0.03	6.65	0.00	0.21	.979	.03	.10
Error (Task × Phone × Evidence)	1.03	49.86	0.02				

For LSD Post hoc tests of group differences in lingual contacts (inductive approach), see Table 8 in text.

NOTES

1. Rhotics are realized canonically throughout most of Mexico and Mexican communities in the US, "except among semifluent speakers" (Lipski, 2008, p. 85). Similarly, Spanish in El Salvador (Quesada Pacheco, 2010) and Salvadoran communities in the U.S. (Lipski, 2008) exhibits non-canonical rhotics very infrequently.

2. Though L2 learners did not self-identify as Latino, their ethnic and cultural identity was also measured using the MEIM and AMAS questionnaires so that they could be included in subsequent analyses of background measures.

3. Hierarchical cluster analyses were performed on the acoustic variables as well, and they indicated that the first clusters to form included both HSs and NSs. Speaker hsmF in particular always clustered with NSs. However, the small sample size precluded cluster analyses from being definitive, and PCA was used instead.

4. Though all were born in the U.S., five of the HSs lived abroad at some point during childhood (mean 4.5 years, SD 3.74) and thus some researchers might not consider them heritage speakers. The deductive approach selecting groups for RMANOVAs as described here included the HSs who had lived abroad. However, the same tests were run with those speakers excluded, and the differences in results were nominal.

5. The full results of the RM ANOVAs and Post Hoc tests are reported in the Appendices.

6. To verify that the increasing effect sizes were not just an artifact of including more groups in the analysis, additional RM ANOVAs were run without the IMs' data, thus equalizing the number of groups across models, and yet the effect sizes still increased when groups were classified inductively.

7. The AMAS did include items about proficiency. As explained in the Methods section, to the author's knowledge the AMAS subscales have not been validated independently, and so items related to proficiency should not be eliminated from the measure. However, an item analysis revealed that the patterns reported here clearly did not result from differences in proficiency.