The Dissertation Committee for Cynthia Patricia Blanco Certifies that this is the approved version of the following dissertation:

Cross-language speech perception in context: Advantages for recent language learners and variation across language-specific acoustic cues

Committee:

Rajka Smiljanic, Supervisor

Colin Bannard, Co-Supervisor

Richard P. Meier

David Quinto-Pozos

Catharine H. Echols

Bharath Chandrasekaran

Cross-language speech perception in context: Advantages for recent language learners and variation across language-specific acoustic cues

by

Cynthia Patricia Blanco, B.A.; B.A.; M.A.; M.A.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May 2016

Acknowledgements

I am lucky to have worked with excellent scholars and mentors in the Department of Linguistics and in the larger academic community throughout my time at UT, and I would like to thank my dissertation committee in particular. Bharath, for challenging discussions both at conferences and on campus. David, for nurturing these dissertation research questions in their infancy in one of my first classes at UT. Cathy, for introducing me to new ways of considering language as a social and cultural tool. Richard, for many conversations that encouraged me to think beyond UT and to the larger community of linguists and speech scientists. Thanks also for help in finding financial support throughout my time in your department. I am also most grateful to my co-advisors, Rajka and Colin, for exceptional guidance as I developed this project and for providing pep talks on demand. Colin, thank you for valuable lessons in experimental design, interdisciplinarity, patience, and perseverance. Rajka, you have set an example for me of a model researcher, colleague, and mentor, and I am so thankful to you for the unwavering emotional and academic support that you have offered me far beyond your role as advisor.

Thanks are also due to my parents, my first teachers, who accidentally made me a linguist and who have endured over three decades of me analyzing their speech. You have supported me even when my plans were unclear, mysterious, or involved yet more schooling. Thanks also for letting me get away with brief, vague responses to questions about my research and for Googling additional details as necessary.

Thanks to my colleagues and mentors at Penn State – Giuli, Jorge, Rosa, and most especially Chip – who shaped my expectations of scholarship, collaboration, and academic community. The lessons I learned from you have contributed to me reaching this milestone.

I am grateful to all the undergraduate research assistants who helped with data collection: thanks especially to Michelle, Karen, Andrea, Gaby, and Maddie in the UT Sound Lab, and to Haley, Amanda, Rui, and Amie with Austin Language Development Labs. Thanks also to Amanda Meeks, who shared in my first attempts at collecting data from children and who readily joined me for the happy hours necessitated by said attempts.

Many thanks to all the participants and parents who volunteered their time to help me complete this project. Dr. Cathy Malerba and Principal Belinda Cini at Wells Branch Elementary School were instrumental in this process, and I could not have made this contribution without their cooperation. Thanks also to the friends, roommates, friends of friends, and roommates of friends of friends who helped in recruiting participants.

I am indebted to my wonderful friends, whose support has taken various forms over the years. To Barbara, Jacqueline, and Suzanne, who have been friends as much as mentors. Thanks especially to Jenny, Jamie, Sarah, Niamh, Monique, Marta, and Mauricio. You lent me your cars, cats, and voices; shared in cupcake happy hours and many margaritas; and took my mind off work with your Skype calls, visits, babies, and lunch dates. To Kirsten, may I be the positive, comforting force during your dissertation that you have been during mine! To Dr. Geer and Dr. Jerro, for commiserating, celebrating, procrastinating, and motivating, since our earliest days, and for the gifs, Google docs, surprise tacos, and playlists that made an impossible task achievable. To Stephanie, for letting me cry, laugh, and think out loud without judgment. Thanks also for pushing me, propping me up, and providing me ice cream as necessary. And to Mark, for all the lessons about linguistics, life, living, and adapting. I hope to learn from you for years to come.

Finally, I am thankful for the financial support offered to me these last years through a Dissertation Fellowship from the Donald D. Harrington Endowment and an American Fellowship from the American Association of University Women.

Cross-language speech perception in context: Advantages for recent language learners and variation across language-specific acoustic cues

Cynthia Patricia Blanco, PhD

The University of Texas at Austin, 2016

Supervisor: Rajka Smiljanic

Co-Supervisor: Colin Bannard

This dissertation explores the relationship between language experience and sensitivity to language-specific segmental cues by comparing cross-language speech perception in monolingual English listeners and Spanish-English bilinguals. The three studies in this project use a novel language categorization task to test language-segment associations in listeners' first and second languages. Listener sensitivity is compared at two stages of development and across a variety of language backgrounds. These studies provide a more complete analysis of listeners' language-specific phonological categories than offered in previous work by using word-length stimuli to evaluate segments in phonological contexts and by testing speech perception in listeners' first language as well as their second language. The inclusion of bilingual children also allows connections to be drawn between previous work on infants' perception of segments and the sensitivities of bilingual adults.

In three experiments, participants categorized nonce words containing different classes of English- and Spanish-specific sounds as sounding more English-like or Spanish-like; target segments were either a phonemic cue, a cue for which there is no analogous sound in the other language, or a phonetic cue, a cue for which English and Spanish share the category but for which each language varies in its phonetic implementation. The results reveal a largely consistent categorization pattern across target segments. Listeners from all groups succeeded and struggled with the same subsets of language-specific segments. The same pattern of results held in a task where more time was given to make categorization decisions. Interestingly, for some segments the late bilinguals were significantly more accurate than monolingual and early bilingual listeners, and this was the case for the English phonemic cues. There were few differences in the sensitivity of monolinguals and early bilinguals to language-specific cues, suggesting that the early bilinguals' exposure to Spanish did not fundamentally change their representations of English phonology, but neither did their proficiency in Spanish give them an advantage over monolinguals. The comparison of adult listeners with children indicates that the Spanish-speaking children who grow to be early bilingual adults categorize segments more accurately than monolinguals – a pattern that is neutralized in the adult results. These findings suggest that variation in listener sensitivity to language-specific cues is largely driven by inherent differences in the salience of the segments themselves. Listener language experience modulates the salience of some of these sounds, and these differences in cross-language speech perception may reflect how recently a language was learned and under what circumstances.

Table of Contents

List of Tables

List of Figures

1. Introduction

Listeners make judgments about talkers and their speech after only brief exposure. Considerable work has investigated the suprasegmental and subsegmental acoustic cues most important for listeners in their decisions about talker-specific characteristics like region of origin, age, and gender (Clopper & Pisoni, 2004, 2007; Harnsberger et al., 1997; Klatt & Klatt, 1990; Strand, 1999; Strand & Johnson, 1996; Tracy et al., 2015). For example, talkers from the Northern region of the U.S. produce the diphthong /oʊ/ with more rounding than talkers from other regions, and listeners are very sensitive to this difference (Clopper & Pisoni, 2004). Other acoustic cues may indicate that a talker grew up using a language other than the one being spoken, yielding a foreign accent (e.g., Flege, 1991; Flege et al., 1997a, 1997b; Flege & Munro, 1994). Foreign-accented speech reflects the transfer of first language (L1) phonological properties to the second language (L2) (Flege, 1995); for example, a Spanish-speaker may produce the English /e/ phone with F1 and F2 values like those used to produce the /e/ in Spanish. At times it may even be necessary for listeners to identify which language a talker is using; this would be the case when a bilingual maps a new word to the appropriate language and such language identification can facilitate a bilingual's access of a known word in one of their languages (Flege, 2007).

Differences in the phonological systems of the L1 and L2 can also lead to listeners "hearing" non-native languages with an accent (Jenkins et al., 1995). Previous work on cross-language speech perception has investigated the sensitivity of native and non-native listeners to phonological contrasts specific to one of the listener's languages. The cues studied in this literature are typically limited to voice onset time in stops (e.g., /b/ versus

/p/) and neighboring vowels (e.g., /i/ versus /ɪ/) in the non-native listeners' L2 – that is, subsegmental cues that differ along a continuum (e.g., the VOT continuum for stop consonants). However, languages can differ in other ways as well: for example, one of a bilingual's languages may use phones that simply do not exist in the other language. The previous findings of listener sensitivity to subsegmental differences may not generalize to other kinds of contrasts, e.g. to the new L2 phones that are theorized to be more easily perceived and acquired (Flege, 1987, 1995). Furthermore, very few studies have considered whether listeners show the same sensitivities when listening to cues embedded in longer stimuli like a word as they do when discriminating isolated segments, and the latter has been the focus of most work thus far. Tasks in which listeners identify and categorize language-specific segments presented in isolation show many similarities in monolinguals' and early bilinguals' sensitivities to particular cues. However, evidence from recent studies that present language-specific contrasts in the phonological context of a word suggests there may be dramatic differences in early bilingual listeners' ability to identify isolated cues on the one hand, and their sensitivity to these same cues in word-length stimuli on the other (Amengual, 2015). These earlier approaches have not fully explored whether listeners attend to all language-specific acoustic cues equally, particularly in more naturalistic stimuli, and so it is unclear how language experience may modulate the perception of language-specific categories in context.

This dissertation uses a novel language categorization task to gauge bilingual listeners' sensitivities to segments from both of their languages for the first time. Furthermore, the project tests the strengths of language-segment associations at two stages of development and across a variety of language backgrounds. The results offer some of the first evidence of the categorization of language-specific segments from both of a bilingual's languages, thus extending previous findings which have focused on listeners' perception of their L2. This approach also provides a means of assessing how listeners use language-specific sounds in context. Recent studies have suggested discrimination of contrasts in word-length stimuli is less reliable than in isolation, and so it is necessary to probe this finding further with new sounds and in a different bilingual population. Finally, previous work in cross-language speech perception has largely omitted listeners between infancy and adulthood, so the current study is also innovative in making explicit connections between bilingual six-year-olds' developing representations and the state of these categories in adults.

This dissertation project presents three experimental studies that examine variations in sensitivity to language-specific segments among listeners of varying ages and language backgrounds. In order to focus on specific segmental cues, nonce word stimuli were developed; these contained sounds common to both Spanish and English except for a particular target segment. These target sounds represented either phonemic cues, categories unique to each language, or phonetic cues, categories shared by the languages but which differ in their phonetic implementation. This contrast between phonemic and phonetic cues was modeled on the distinctions made in Best (1991, 1995) and Flege (1987, 1995) to describe patterns in cross-language speech perception and the acquisition of L2 sounds by second language learners: phones new to an L2 are like the phonemic cues used here, and L2 sounds similar to existing L1 categories are similar to phonetic cues.

In the three experiments described here, listeners categorized nonce-word stimuli with a phonemic or phonetic cue as sounding more like English or Spanish. In Experiment 1, the categorization accuracy and responses times of English monolingual and early and late Spanish-English bilingual adult listeners are compared. Experiment 2 further probes the relationship between accuracy and reaction time in monolinguals and bilinguals by forcing listeners to respond faster and more slowly than average reaction times observed in Experiment 1. The results provide insight about the nature of the mental representations of bilinguals and about their access to language-specific phonological information. Experiment 3 uses this novel task to examine the language-cue associations of six-year-old children, and three listener groups were compared: monolingual English children, children enrolled in a Spanish-English dual-language immersion program who heard only English at home, and children in the same program who heard Spanish at home. The findings demonstrate that listener sensitivity to segmental cues varies more widely than might be predicted by existing models of cross-language speech perception, and this variation is related to two factors. First, segments vary in their inherent salience or in how easily they are distinguished from other sounds perceptually, so even for a class of sounds that are predicted to be easily perceived there may be drastic differences across members of the class. Second, the recency with which a listener has learned a language highlights language-specific phonological differences and leads to better language-segment associations for learners for whom the L2 is more newly acquired. The conclusions describe the interactions of listeners' mental representations of each language, and implications for foreign accent perception and phonological development are discussed.

2. A review of cross-language speech perception

The current project seeks to evaluate listener sensitivity to language-specific segmental cues and will consider how a listener's language experience influences the use of these cues to categorize controlled stimuli as more English-like or more Spanish-like. This chapter has two goals: 1) to summarize previous findings on cross-language speech perception and the role of language experience, and 2) to motivate the current project.

2.1 Cross-language speech perception

Speech perception is shaped by a listener's native language(s), but it also reflects changes in the input and the frequency of exposure to talker- and language-specific properties. This section describes how phonological categories develop in the native language, how other-language phones are processed, and how listener experience influences the perception of language-specific categories.

2.1.1 The development of language-specific phonological categories

Listeners become attuned to the phonological contrasts particular to their ambient language(s) before the end of the first year of life (Albareda-Castellot et al., 2011; Bosch & Sebastián-Gallés, 2003; Werker & Tees, 1984). This perceptual tuning defines phonological categories according to language-specific boundaries such that listeners recognize sounds from different sides of the boundary as distinct categories and sounds from the same side of the boundary as the same category (Eimas et al., 1971; Lisker & Abramson, 1970). Acquisition of the category boundaries reflects distributional properties of the input (Maye et al., 2002). Much of the work on phonological categories has concerned the continua defining the vowel space and voiced and voiceless stops. For example, six-month-old infants can distinguish stops with voicing before the stop closure is released (pre-voiced stops) from stops with a short lag between stop closure and the onset of voicing (short-lag VOT), and short-lag stops from stops with a long lag between stop closure and the onset of voicing (long-lag VOT), irrespective of ambient language (Lasky et al., 1975). By twelve months, however, English-learning infants only distinguish short-lag from long-lag, which reflects the English difference between /b/ and /p/, and can no longer discriminate pre-voiced stops from those with short-lag VOT (Werker & Tees, 1984; although see Aslin et al., 1981, for conflicting results). Similarly, Spanish-learning infants at twelve months can discriminate pre-voiced stops from stops with short-lag VOT, representative of the Spanish /b/ vs. /p/ contrast, but recognize short-lag and long-lag stops as members of the same category. This means that a single auditory stimulus, like an alveolar stop with a short-lag VOT of 15ms will be recognized by an English-speaker as an exemplar of /d/ but by a Spanish-speaker as an exemplar of /t/. These sensitivities to language-specific category boundaries persist into adulthood (e.g., Pisoni & Tash, 1974)

There are two important consequences of such language-specific categories. The first is that listeners begin hearing with an accent (Jenkins et al., 1995), shaped by their native language, before their first birthday. That is, listener sensitivity to segments and acoustic cues reflects the listener's exposure to these cues in their native language, potentially resulting in mismatches between acoustic properties or differences in input and how the input is mapped to a listener's phonological categories. Secondly, infants exposed to two languages that differ in category boundaries (e.g., Spanish and English) must develop separate, even competing, categories across their languages. A Spanish-English bilingual will be exposed to the short- vs. long-lag distinction in English, but the pre-voicing vs. short-lag contrast in Spanish, and so this bilingual will have exposure to two distinct boundaries for the identification of voiced and voiceless stops. The mapping of a word or sound to the appropriate phonological category (voiced or voiceless) may then rely on the language in which the word was uttered. Additionally, a language may contain a sound that is so unlike any category in another language that listeners who lack exposure to this sound do not process it as speech at all, and so do not map it to a category in their own language. This may be the case for native speakers of English who hear a language with clicks, which may be perceived as non-linguistic (Best, 1991), so the acoustic salience of a sound also influences how distinctly a segment is perceived, in addition to a listener's language experience.

Models of native-language perception provide a framework for explaining how sounds across a continuum can be classified into phonological categories. For example, the perceptual magnet effect proposed by Kuhl (1991) explains that the best exemplar of a phonological category, e.g., of the vowel /i/, defines the category and that sounds in the surrounding perceptual space are judged to be variants of the perceptual magnet, that is, of the best exemplar. Infant and adult listeners are better able to generalize from the magnet to neighboring sounds, than from neighboring sounds to the magnet, even though neighboring sounds that differ from the magnet are identified as instances of the same category.

Iverson et al. (2003) account for such effects through the warping of the perceptual space, which happens through exposure to a certain type of input, e.g., many instances of the /i/ category in the ambient language. In the distorted or warped perceptual space, some acoustic dimensions are highlighted and others damped in accordance with cue weighting in the language. For example, Japanese listeners rely on F2 to distinguish English /ɹ/ and /l/, since this is the salient difference for this contrast in Japanese, even though F3 is more important in English. Lively & Pisoni (1997) and Lotto et al. (1998) provide evidence that the best exemplar of a category is likely context-dependent and changes in response to the other phones with which it is presented to listeners. This leaves unclear how "prototypical" any sound can be, if its prototypicality changes with context. However, results from these studies confirm that listeners identify some exemplars as more representative of a given category than others. (This, despite the fact that the stimuli identified as most prototypical of /i/ were not consistent across listeners and also did not replicate the values associated with Kuhl's (1991) best /i/ prototype.) Miller (1997) summarizes finding from a number of her projects that corroborate differences in the representativeness of stimuli for their phonological category. These studies of native-language categories suggest that while listeners can recognize certain sounds as good or bad examples of a category, category boundaries themselves may be more flexible and sensitive to acoustic context than the previous literature on VOT would suggest. Listeners thus retain a good deal of flexibility for phonological representations into adulthood, at least for familiar sounds that vary along familiar continua.

These distinctions among exemplars of a category also help describe how native listeners perceive other-language contrasts. Best's perceptual assimilation model (Best, 1991, 1995) proposes that non-native contrasts can be identified as more or less difficult for listeners to distinguish depending on how non-native contrasts relate to the phonological categories in the native language. According to the model, there are four potential relationships between non-native contrasts and categories in the native language, and these relationships can be ranked in order of predicted ease of discrimination for a non-native listener. The most easily perceived non-native contrasts are those that map to distinct phonemes in the listener's native language and those that are so unlike any sounds in the native language as to be perceived as unspeechlike. The most difficult non-native contrasts to discriminate are contrasts that map to a single native category; in one case, both non-native variants are equally poor exemplars of the single native category, while in the other case one of the variants matches the native category well and the other variant is divergent. This ranking allows for specific predictions about the sensitivity of native listeners to different kinds of segments and contrasts from other languages.

Differences in the discriminability of non-native contrasts and how they map to native language categories influence how these sounds are acquired by second language learners as well. Flege's Speech Learning Model (SLM; 1987, 1995) makes a similar division between phones in the second language (L2) that seem "new" versus L2 phones that seem "similar" to categories in the learner's first language (L1). An L2 sound perceived as distant from the nearest L1 category will be treated as a new sound and will be well-discriminated. Listeners create a new phonological category for the new sound, which allows the learner to begin acquiring the articulatory gestures necessary for native-like production. If an L2 sound is perceptually similar to an existing L1 phone, a learner will not initially treat the L2 category as different from the L1 category, even if there are acoustic differences between the articulations of each variant. Since the new sounds are better discriminated than the similar sounds, listeners establish separate categories for new phones earlier than for similar phones; listeners thus learn new sounds sooner and better than similar L2 phones. Similar sounds are subject to more perceptual confusion with the similar sound's L1 analog, but phonetic training (Francis & Nusbaum, 2002) and continued exposure and language use (Flege, 1995) can facilitate the development of distinct categories even for difficult similar sounds.

The SLM makes further predictions about how a listener's age of acquisition influences the development of L2 categories and how the L2 categories themselves may affect related L1 phones. The SLM hypothesizes that the later a listener begins learning the L2, the more difficult it will be to discriminate non-contrastive phones across the languages, like the similar ones described above, which differ in some subsegmental cue(s) from the L1 production of the category. Thus earlier learners are expected to have success discriminating both new and similar phones in the L2, whereas later learners will distinguish the acoustically distant new L2 phones better than similar L2 phones. The SLM also proposes that second language learners accommodate new categories in the same phonological space by keeping L1 and L2 categories maximally separate, if or once the sounds are recognized as members of separate categories.

No specific prediction is made for how age of learning affects the separation of these categories, since the SLM describes the same process for perceptual learning for listeners of all languages, but subsequent work attributes differences in L2 achievement to factors correlated with age of acquisition, including how established L1 categories were at the time of L2 learning and continued exposure to and use of the L1 (Flege & MacKay, 2004; Flege, 2007). It also follows from the other postulates and hypotheses of the SLM that earlier learners may keep language-specific categories more separate than later learners. These questions, and the SLM, have inspired a great deal of research into how listeners who vary according to age of acquisition and continued use of the L1 learn and maintain separate phonological categories for each language.

2.1.2 Language experience in category discrimination

This section presents research describing how phonological category boundaries can change as the result of experience with a particular kind of input and summarizes the results of experiments that have tested the SLM’s postulates and hypotheses. Even after phonological categories in the first language have been established, category boundaries are susceptible to subsequent input and so may move and change, even in the absence of a second language. Phonological categories represent consistencies in the language – e.g., American English contrasts stops with short versus long VOT lags – but listeners' phonological representations also include information about when and how these boundaries might change, for example, depending on a particular talker's difference in VOT between voiceless and voiced stops. These speaker and contextual characteristics are encoded in listeners' representations even after brief exposure in the lab (e.g. Allen & Miller, 2004; Kraljic & Samuel, 2006, 2007; Nygaard & Pisoni, 1998), and so listeners can manage their linguistic expectations if they have previous experience with a particular talker or context. For example, listeners shift the boundaries of a phonological category in response to the distribution of an ambiguous sound heard during a brief exposure period in the lab, depending on whether the talker used the ambiguous sound in place of /f/ or /s/ in known words (Kraljic & Samuel, 2006). The mapping of a sound to listeners' mental representations is also subject to boundary adjustments based on speaking rate (Miller & Volaitis, 1989), carrier phrases (Repp & Lin, 1991), and other co-occurring sounds (Lively & Pisoni, 1997; Lotto et al., 1998).

These talker- and context-specific shifts can be explained with exemplar theories of speech perception (Johnson, 1997; Pierrehumbert, 2002). Listeners use stored exemplars, e.g., those from an exposure period or simply from other stimuli in the task, to inform their expectations about unheard productions and word forms. This happens when listeners generalize over the stored exemplars; for example, exposure to many examples of a talker using very long VOTs to produce voiceless stops in English allows a listener to move the voiced-voiceless category boundary to longer VOTs for the particular speaker. Listeners are also able to associate indexical properties with production variants; for example, they store information about whether a pronunciation is associated with speakers from a particular language or regional background, even without knowing the language itself or necessarily having lived in the specific region, so listeners with exposure to particular accents show improved processing and categorization of familiar accents (Clopper & Pisoni, 2004, 2007; Vieru et al., 2011; Witteman et al., 2013). Listeners like bilinguals who have experience with a particular feature across two languages must associate productions with each language in order to make the appropriate conclusions about the phonological categories in each language. A Spanish-English bilingual who hears a word produced with a /t/ will store with this exemplar whether the sound was produced in English or Spanish, and information about how it was produced (e.g. the VOT of the stop) will be added to the listener's representation for the production of /t/ in the language. This linking of language-specific segments with the respective language may benefit the listener by "facilitat[ing] lexical access, especially when bilinguals search both the L1 and L2 lexicons in the process of accessing a word" (Flege, 2007). Spanish-English bilinguals will therefore have developed detailed phonological representations for English and Spanish, and their sensitivity to the distribution of sounds particular to each language might be expected to be greater than English monolinguals', who have only English productions on which to base their language representations. However, due to existence of multiple (language-specific) categories in the same phonological space, Spanish-English bilinguals' representations might also be unlike English monolinguals' in other ways: bilinguals might use categories more extreme than monolinguals to maximize differences between languages (cf. Flege, 1995), or bilinguals' categories may show evidence of cross-linguistic transfer and be less like the monolinguals', especially for later-acquired sounds (Flege, 2007).

Previous investigations into cross-language speech perception have compared monolinguals' and bilinguals' abilities to categorize or discriminate isolated segments or subsegmental cues. In these studies, bilingual listeners hear pairs or triplets of sounds ranging along a continuum, most often the VOT continuum (e.g. between /t/ and /d/) or formant continua between neighboring vowels in the L2 (e.g. /i/ and /ɪ/). The stimuli in these experiments test listeners' sensitivities to contrasts particular to the L2, typically English, but less frequently do they also compare the same listeners' ability to identify or discriminate sounds particular to their first language.[1] Most findings indicate that monolingual English listeners and early bilinguals make similar distinctions among English categories (e.g., Flege et al., 1999; Mack, 1989), and that this is especially true for bilinguals who have lower rates of continued use of or exposure to their L1 (Flege & MacKay, 2004; Flege, 2007). In some vowel discrimination tasks, even late bilinguals who acquired English after puberty pattern like English monolinguals (Flege et al., 1994), although other work indicates that late bilinguals are less sensitive to L2 contrasts (Flege et al., 1999). Since the five Spanish vowels also exist in English, Spanish-English bilinguals might be expected to use a single (Spanish) category for both languages, but at least some results suggest that proficient bilinguals represent categories from both languages in the same psychoacoustic vowel space, for example, with separate categories for Spanish /i/ and English /i/ and an additional category for English /ɪ/. This creates a greater density of phonological categories in bilinguals' perceptual space, so bilingual

---

[1] This pattern of investigation is in direct contrast to the extensive exploration of bilinguals' productions in both languages, especially in code-switching. For recent work in this area, see Amengual, 2012; Balukas & Koop, 2015; Bullock & Toribio, 2009; Goldrick et al., 2014; Olson, 2013, *inter alia*, and see Flege (2007) for a summary of earlier work.

listeners may be obligated to show greater sensitivity to the cues distinguishing the language-specific categories (Flege et al., 1994).

While it is clear that both early and late bilinguals can achieve native-like proficiency in the identification and discrimination of L2 phones, the results of these tasks may tell us more about general auditory abilities and less about listeners' linguistics representations. Presenting isolated segments may not adequately account for what cues listeners are sensitive to when perceiving continuous, naturalistic speech. In fact, differentiating stops with long and short VOTs may not require accessing linguistic representations at all, as is the case when listeners make similar judgments between continua of non-speech sounds (Diehl & Walsh, 1989; Pisoni, 1977). Some animal species can also make human-like distinctions between isolated segments or syllables (Kluender et al., 1987; Kuhl & Miller, 1978). It is therefore possible that listeners use non-linguistic and general auditory strategies to make decisions about the isolated segments and acoustic cues used in these tasks (Flege, 1987; see also Repp, 1982), and more naturalistic speech stimuli may provide additional insights into the organization of phonological information in bilinguals' mental representations. Furthermore, these studies typically only evaluate listener sensitivity to cues in the L2, most often English, so very little is known about how they process segments particular to their first language.[2]

2.2 LANGUAGE-SPECIFIC CATEGORIES IN CONTEXT

The discrimination of isolated segments may not tap listeners' language representations, and so the findings of studies with longer, more naturalistic stimuli are reviewed to assess how much phonetic detail is encoded in listeners', especially bilinguals',

[2] However, see Carlson et al. (2015) for recent work on early bilinguals' use of L1 (Spanish) phonotactics in speech perception.

representations of their language(s). While categorical perception and discrimination tasks like those described in the previous sections have been employed with adult and infant listeners, children older than infants have been absent from the research on segmental contrasts. However, the sensitivities of this age group to language-specific categories have been studied with longer stimuli, so the results of both adult and child perception literature are described here.

2.2.1 Adults

Relatively few studies have attempted to extend the findings of isolated-segment tasks to the perception of accented or language-specific speech, which both include language-specific segments in longer utterances. Flege (1984) posed the question of stop burst and VOT discrimination as one of distinguishing between native and non-native speech, and the results demonstrate that native English listeners can use even 30ms of a stop burst to differentiate productions from native- and French-accented talkers. In the same series of studies, Flege (1984) also found that listeners could distinguish native and non-native talkers after hearing CV syllables, single words, and three-word phrases. However, the fact that listeners mostly accurately categorized stimuli from across this range of input does not indicate that the strategies used in one task are the same as those employed for another. For the longer utterances, listeners may not have necessarily made use of stop burst differences at all, even though they can identify these differences in other tasks. Instead, listeners may pay more attention to other segmental and suprasegmental cues present in the longer speech, especially since the longer stimuli contained multiple co-occurring cues. For example, evidence from a perceptual-similarity task using phrase-length stimuli from 17 languages suggests that marked back consonants and front vowel rounding might be particularly salient dimensions for non-native listeners (Bradlow et al., 2010). However, there remains some question about the interpretation of at least the vowel dimension in the perceptual-similarity study, so the number of cues present in even short phrases makes it difficult to identify the most influential acoustic factors.

Flege & Munro (1994) attempted to test listener sensitivity to the multiple cues available in word-length stimuli by asking monolingual English listeners to categorize productions of *taco* as having been produced in English or in Spanish. The length of VOT associated with the initial /t/ explained more variance in listeners' responses than any other acoustic cue, but this language-specific difference is confounded with having occurred so early in the word – listeners may not have attended to the whole word if they could confidently make a decision based on the first segment or syllable. Since the four segmental cues in *taco* were all Spanish-like or English-like in any single production, the results also do not reveal how listeners might use any one of the cues in language judgment tasks. The VOT of /t/ was the strongest cue, but it is unclear if the other cues would have been sufficient for listeners to categorize productions accurately. The sensitivity of listeners to language-specific stops in Flege (1984) and Flege & Munro (1994) suggests that listeners may indeed be able to compare the VOT of the stimulus to their stored representations of what is an acceptable or atypical VOT for stops in their language, even if the strategies used to evaluate VOT differences in isolation may be non-linguistic. However, it is still unclear how important any single cue is, especially since the VOT cue in Flege & Munro (1994) was always presented in the first segment, and little is known about how listeners use other language-specific cues.

Other work with monolingual listeners has taken a broader approach to listeners' sensitivity to acoustic cues by testing accented speech and not targeting specific segments. In studies of accent perception and categorization, monolinguals can make above-chance accent categorization judgments (Clopper & Pisoni, 2004, 2007; Vieru et al., 2011) and are primed by accented productions (Witteman et al., 2013). Since some of this work was on regional varieties of American English (Clopper & Pisoni, 2004, 2007), and variations in VOT were unlikely to be cues to accent, monolingual listeners are probably sensitive to a number of other acoustic cues, in addition to VOT, which may be used in other tasks to make language judgments. However, in these studies, listeners with more exposure to the accent(s) made more accurate and more fine-grained classification decisions and were primed even by more strongly accented words, so inexperienced listeners seem less able than experienced listeners to capitalize on the same range of acoustic cues available in the input. This suggests that listeners will vary in their sensitivity to different segments, so while the extensive work on VOT is informative and the continuum is well understood, the results from this literature may be limited in their generalizability. Testing other segments, especially new L2 phones and those that represent the ends of other continua (e.g., of other vowels of continuously varying consonants like /l/ variants), is necessary to more completely understand the structure of phonetic detail in lexical representations.

The previous studies assessed only monolinguals' sensitivity to language-specific cues in context, but work from mispronunciation studies indicates that bilingual listeners who can easily discriminate segmental cues in isolation might be less able to identify those same differences in word-length stimuli. This disparity across tasks is true even for early, highly-proficient bilinguals. Listeners in these studies complete identification and discrimination tasks, and then identify whether a stimulus is the typical pronunciation of the word or if it is mispronounced. For the isolated segment tasks contrasting neighboring vowels in Catalan (e.g. /ɛ/~/e/), there are conflicting results: highly-proficient Spanish-dominant Spanish-Catalan bilinguals in Barcelona were unable to reliably distinguish the

Catalan mid-vowels is isolation (Sebastián-Gallés & Soto-Faraco, 1999), while their peers in Majorca were successful (Amengual, 2015). However, Spanish-dominant bilinguals in both locales responded similarly poorly in the mispronunciation tasks, in which they heard a word's actual mid-vowel replaced with the neighboring vowel (e.g., /ɛ/ replaced with /e/, as in /ərɛl/ 'root' pronounced as */ərel/). Sebastián-Gallés and colleagues (1999, 2005) attribute the lack of detail in Spanish-dominant bilinguals' representations of Catalan to their exposure to Spanish in the first years of life, before acquiring Catalan. However, Amengual's results indicate that early Spanish exposure itself is not the cause of early bilinguals' decreased discrimination abilities in the mispronunciation task, since listeners in Majorca could reliably perceive differences when the segments were presented in isolation. This suggests that, in both regions, the Spanish-dominant bilinguals' lexical representations of Catalan contain less phonetic detail for Catalan-specific contrasts, despite the ability of some listeners to discriminate the segments in other tasks. Amengual suggests that the greater range of acceptable Catalan pronunciations is due at least in part to more variable Spanish-accented input in Catalan.

Flege's work on cross-language speech perception and this work on Spanish-Catalan bilinguals leads to several important challenges for future work on the nature of bilinguals' linguistic representations. First, given the differences in results between the segment identification tasks and mispronunciation tasks, it is important to use tasks that require listeners to respond to input that forces linguistic and not just auditory processing. Second, early acquisition of, extensive exposure to, and high proficiency in both languages may not always be sufficient for bilinguals to develop native-like phonetic representations of the less-dominant language, if the results from Spanish-Catalan bilinguals can be generalized. However, the findings of Flege and colleagues (1987, 1994, 1999, 2004, 2007) and Mack (1999) indicate that the Spanish-Catalan results cannot be generalized to Spanish-English, Italian-English, and French-English bilinguals. Future work must continue to probe differences in the learning situations of these bilinguals that would contribute to divergent results in segment identification and classification tasks. Finally, the scope of these cross-language perception tasks has typically been restricted to testing contrasts in only one of the bilinguals' languages. In order to more completely understand the mutual interaction between the phonological categories of both of bilingual's two languages, it is necessary to include stimuli particular to both languages.

2.2.2 Children

After infants' sensitivities narrow to ambient language categories during the first year of life (see Section 2.1.1), language-specific contrasts are tested in children in word-learning tasks. By age 2, children can use phonetic detail to learn similar words (e.g. *bin* and *din*) when they are presented in referential expressions (Fennell & Waxman, 2010), and they can recognize mispronunciations as acceptable labels of familiar objects (Mani & Plunkett, 2007). Two-year-olds also adapt to foreign-accented productions after exposure to the accent (Schmale et al., 2012). It is also around this age that bilingual children acquiring both languages simultaneously begin to show systematic differentiation of their languages, and use the appropriate language with each parent more often than would be predicted by the child's own dominance (Genesee et al., 1995; Nicoladis & Genesee, 1996; Nicoladis, 1998; Paradis & Nicoladis, 2007).[3] Toddlers thus have sufficiently detailed

---

[3] For example, if the child is judged to be 80% dominant in French and 20% dominant in English, the child might be expected to use French 80% of the time and English 20% of the time, across all interactions. If a child recognizes that a parent's preferred language in English, and if the child is able to discriminate French and English, the child is expected to use English in significantly more than 20% of the interaction with this parent.

phonological representations that acoustic variation can be integrated into long-term representations, especially if there is a social and/or pragmatic reason to attend to the variation, as in the case of bilingualism.

By the end of preschool (age 5), children can associate acoustic information (e.g. voice) with speaker-specific traits and preferences (Creel, 2012b) and so are beginning to incorporate talker- and context-specific information in their representations. Monolingual children may be hypersensitive to phonetic deviation from canonical pronunciations and may be unable to map the phonetic information of a segment with the phonemic level of representation. This inability to generalize from phonetic information to phonological category prevents the association of some accented productions with the canonical lexical forms the children are most familiar with (Nathan et al., 1998). Children at this age also respond more slowly to mispronunciations but still accept them as variants of target words (Creel, 2012a), and the better performance of children in Creel (2012a) compared to Nathan et al. (1998) may be due to the fact that in the latter study children were exposed to naturalistic speech, a Glaswegian accent, with potentially multiple deviant acoustic cues, while in the Creel study the mispronunciations were created by producing a single change in vowel. Further evidence for children's ability to identify sounds unlike their own phonological categories comes from Kinzler et al., (2009, 2011), in whose studies children have shown a preference for native-accented over foreign-accented speakers. As for the bilingual preschoolers, language differentiation becomes increasingly more reliable, although their developing language dominance also influences how often children use each language with an interlocutor (e.g., Paradis & Nicoladis, 2007).

However, there is evidence that not all deviations from children's typical or familiar categories are equally perceptible. In early elementary school, children are not sensitive to the differences between local and regional dialects, even though they can reliably differentiate their local dialect from foreign-accented speech (Floccia et al., 2009; Girard et al., 2008). This finding indicates there must be acoustic properties of foreign-accented speech that make this variety more salient to children than regional accent differences. This possibility was further explored in Wagner et al. (2013), in which children heard the local accent, a regional variety, and Indian English, a native variety that differs more than other regional accents but is more consistent than foreign-accented speech. Children categorized the accents both explicitly in a categorization task, as in Floccia et al. (2009) and Girard et al. (2008), and implicitly, in a task that required children to associate accents with different properties. In the explicit tasks, children again were unable to distinguish local and regional varieties but could separate their local dialect from Indian English. However, in the implicit task, children could differentiate all three accents. Wagner and colleagues propose that children perceive dialects on a continuum between their local variety at one extreme and foreign-accented speech at the other. Regional varieties, which at least in English differ principally in vowels, are closer to the local dialect than Indian English, which differs more significantly in both consonants and vowels, and falls between regional accents and foreign accents on the continuum. This may also be why younger children in Nathan et al. (1998) struggled with the Scottish accent, if there were substantial consonantal differences in the accented productions (e.g., the Scottish trill /r/), while children in Creel (2012a) could better assimilate vowel mispronunciations into existing lexical representations. There may be other cues distinguishing kinds of accents, besides the consonant/vowel contrast, such as the new and similar phone distinction described by Flege (1995) in the SLM. Children may be more sensitive to new phones in Indian English and foreign-accented speech, but may be less able to identify deviations in similar phones across regional dialects. Such an account could motivate children's differential access to phonological information in Wagner et al. (2013) and the difficulty they had discriminating local and regional dialects in Floccia et al. (2009) and Girard et al. (2008).

Unfortunately experimental work considering cross-language speech perception in children after infancy is limited to monolinguals, as most work with bilingual children's language development has been observational or has not focused on phonological development. A partial exception is Genesee et al. (1995), who tested bilingual toddlers' use of each language with a new interlocutor, but the gap remains between the more acoustically-driven work with monolinguals and the more holistic language-differentiation work with bilinguals. In order to understand how phonological categories develop and interact in the representations of both monolinguals and bilinguals, experimental methodologies need to be developed to test cross-language speech perception in young children.

2.3 The current project

Previous work has demonstrated that language-specific phonological categories are established early in life and shape our perception of other-language sounds, but even so the internal structure of these categories and their boundaries are subject to the influence of ambient language(s) and ongoing input. Monolingual and bilingual listeners' abilities to identify and discriminate language-specific segments, particularly those from the bilinguals' L2, have been well-studied, and evidence indicates that early Spanish-English bilinguals often show native-like sensitivities to sounds in their L2, English. However, the results from research with early Spanish-Catalan bilinguals are more dubious and indicate differences in listener sensitivities according to whether tests involved isolated segments or real words. The behavior of late learners is more variable on the whole than early bilinguals, but in some tasks late bilinguals also successfully discriminate L2 sounds. It is less well understood how these sensitivities develop from infancy to adulthood, since few discrimination tasks have been developed for children.

Listeners in these cross-language speech perception tasks typically hear segments in isolation, but this approach may not adequately gauge how listeners use these cues in their lexical representations. In fact these discrimination tasks may not require linguistic processing at all. It is important that investigations into the nature of bilinguals' representations of their languages use tasks in which listeners process segments heard in phonological context to avoid discrimination that relies on auditory processing alone. Previous work contrasting isolated segments with those same segments in word-length stimuli indicates listeners may show different sensitivities when listening in context and when hearing a segment in isolation, so listeners may access more acoustic detail when contrasting segments than when retrieving lexical items. To further probe what listeners understand about the phonology of their language(s), a wider range of stimuli should be considered, beyond the VOT and vowel continua that have been the focus of previous work (although earlier studies started with these contrasts for good reason). There are theoretical reasons to suppose that listeners do not treat all language-specific segments equally and that access to different classes of segments may vary. Some segments and contrasts are expected to be easier for non-native listeners to map to new, separate categories, while categories perceived to be more similar to native sounds should be more difficult to discriminate. It remains to be seen how the association of language and language-specific cues varies as a function of cue type and listener experience. An investigation into these factors will extend previous work on isolated segments by testing attention to phonetic detail in the context of language, will inform models of speech perception (especially those modeling L2 category learning), and will contribute to our understanding of the plasticity of speech perception.

This dissertation investigates the sensitivity of listeners from different language backgrounds to two classes of language-specific segmental cues. The current project includes both English-specific cues and Spanish-specific cues and tests children and adults with different ages of acquisition to more completely understand how language experience influences the perception of segmental cues in context. The three studies presented here use a novel task in which listeners are asked to associate nonce words containing these language-specific cues with the appropriate language. Monolingual and bilingual listeners categorized stimuli containing a Spanish- or English-specific segmental cue as sounding more like Spanish or English. By manipulating a single cue in a stimulus word, and holding constant the remaining segments, conclusions can be drawn regarding whether listeners are aware of the link between the target segment and its presence in the lexical representations in only one of their languages. There are three benefits of using nonce words in this study. First, creating nonce words allows for control over the non-target segments and phonotactic properties of the stimuli. Second, the use of nonce words prevents the influence of word frequency on listeners' language decisions. Third, making decisions about nonce words forces listeners to access generalizations about the phonological inventories of their languages by abstracting across their lexical representations of each language.

The language-specific segments tested here represent categories that are unique to English or Spanish and segments that vary in how they are implemented phonetically between the Spanish variant and the English variant; these categories were classified as either "phonemic" or "phonetic." Phonemic cues, like Flege's new phones and Best's separate categories, were expected to be strong indicators of language and were defined as segments for which there is no analogous sound in the other language. Phonetic cues were modeled on what Flege referred to as similar phones and would be predicted by Best to be mapped to a native-language category; these were defined as cues for which the same category exists in both languages, and so these are expected to be more difficult for listeners to use.

It is expected that the bilingual groups will show greater sensitivity to language-specific cues than the monolinguals, since the bilinguals' considerable exposure to both English and Spanish productions should permit more reliable associations between language and the sounds present in the lexical representations of each language. Early bilinguals have had earlier and longer exposure to both languages than the late bilinguals, so these early learners might be more sensitive to more difficult cues like the phonetic segments, especially if native-like differentiation of these sounds is related to age of acquisition. On the other hand, many early Spanish-English bilinguals from Texas frequently codeswitch between their languages and have grown up doing so in their bilingual communities, so the early bilinguals may be less practiced at separating English from Spanish than late bilinguals who learned English after puberty and typically in foreign language classrooms.

Experiment 1 tests the categorization accuracy and response times of adults to English- and Spanish-specific phonemic and phonetic cues. Listeners included English monolinguals, early Spanish-English bilinguals who have spoken Spanish at home since birth and who learned English beginning in kindergarten, and late Spanish-English bilinguals who are from Spanish-speaking countries and did not learn English until moving to the U.S. after age 13. Listeners from all backgrounds are expected to categorize the phonemic cues more accurately than the phonetic cues, and the bilingual groups are expected to make more accurate categorization decisions than the monolinguals. It is unclear whether earlier acquisition of both languages results in the early bilinguals showing greater sensitivity to language-segment associations than the late bilinguals. Experiment 2 probes the speed-accuracy trade-off possibly responsible for differences between the early and late bilinguals found in Experiment 1. Adult listeners from the same three backgrounds completed a timed version of Experiment 1 in which responses to the same nonce-word stimuli were speeded or delayed. If early and late bilingual listeners have fundamentally similar associations between languages and segments, on account of speaking the same languages, the delayed condition may allow both groups to respond more accurately than in the speeded conditions. If language representations or the details stored in them differ across the bilingual groups, one group may be more accurate than the other in all conditions. Finally, Experiment 3 further probes how language experience affects cue sensitivity by investigating how listeners at an earlier stage of linguistic development respond to language-specific cues. Experiment 3 adapts the task in Experiment 1 for six-year-old children, and listeners included English monolinguals, children exposed to only English at home who are enrolled in a Spanish-English dual language program at school, and children exposed to Spanish at home who are enrolled in a Spanish-English dual language program at school. Children are expected to show less sensitivity to language-specific segments than adults overall, and as with the adult bilinguals, the children who are learning Spanish and English are expected to categorize nonce words more accurately than monolingual children.

This project pursues answers to the following research questions and tests the following hypotheses:

Experiment 1

1. Are listeners able to use phonemic and phonetic segments to distinguish Spanish and English, and does language experience affect sensitivity to the cues?

    a. Listeners are expected to be more sensitive to phonemic cues, modeled on the new phones in Flege's SLM, than to phonetic cues, which are like similar phones in the SLM. Phonemic cues should be easier for listeners of all language backgrounds to perceive as distinct categories, and so are hypothesized to be easier to associate with the appropriate language.

    b. Bilingual listeners are expected to categorize the nonce words more accurately than monolingual listeners, since bilinguals have linguistic representations of both Spanish and English and can contrast these representations. Previous work suggests early bilinguals will be as sensitive to English segments as monolingual listeners, so the early bilingual advantage over monolinguals might be limited to the Spanish-specific segments. Early bilinguals are expected to show greater sensitivity to English-specific cues than the late learners of English.

Experiment 2

2. Does listener accuracy change as the result of having more or less time to access language-specific phonological information?

    a. If early and late bilinguals have access to fundamentally equivalent information about each language's phonological categories, the accuracy of both groups will improve and become more similar when more time is allowed to make the categorization decision. When responses are forced to be faster, both groups will respond less accurately and more similarly to each other.

b. If one group of bilinguals has encoded or has greater access to phonetic detail in their representations of English and Spanish, neither bilingual group is expected to perform differently than they did in Experiment 1, no matter how much time is allowed for a response.

Experiment 3

3. Are six-year-olds able to use phonemic and phonetic segments to distinguish Spanish and English, and does language experience affect sensitivity to the cues?

    a. Children are expected to be less accurate than adults in the use of language-specific segments to make categorization decisions since their linguistic representations are still developing. However, like adults, children are expected to be more sensitive to phonemic cues than to phonetic cues.

    b. Children learning English and Spanish are expected to be more accurate than monolingual children due to having developing representations of both languages to compare.

4. How does language experience from childhood to adulthood change listeners' sensitivity to language-specific segments?

    a. The monolingual children are expected to have comparable accuracy to the monolingual adults, who they grow into, since both groups continue to have the representations of one language on which to base their judgments.

    b. The children who speak Spanish at home are expected to have comparable accuracy to the early bilingual adults, since these children will have the same linguistic profile as the early bilingual adults when they grow up.

This dissertation makes three novel contributions to cross-language speech perception and the nature of phonological categories in bilinguals. First, the present studies employ a novel task that requires listeners to attend to individual segmental differences in the context of a word. This methodology allows for the salience of different language-specific sounds to be directly compared, while also presenting the segments in a context more naturalistic than isolation. Second, this work evaluates listener sensitivity to a range of segmental cues. Most stimuli in earlier studies have been very limited in scope, including only a particular vowel contrast or a single subsegmental cue like VOT. By including a range of both phonemic and phonetic segments, this project sheds light on the relative importance of different sounds in listeners' language decisions. Finally, each of the three experiments expressly compares listeners of different language backgrounds. This study builds on previous work by more completely assessing language-specific cues from both Spanish and English, including English monolinguals and early and late Spanish-English bilinguals, and comparing the sensitivities of children and adults.

The remainder of the dissertation is organized as follows. Chapter 3 describes the segmental cues and nonce-word stimuli developed for the categorization task. Chapter 4 reports on Experiment 1, in which monolingual and early and late bilingual adults categorized the language-specific nonce words. Chapter 5 presents findings from Experiment 2, a categorization task that manipulated listener response time to address an unexpected difference between early and late bilinguals in Experiment 1. Chapter 6 describes Experiment 3, in which six-year-olds completed an adapted version of Experiment 1 to better understand developmental issues in cue sensitivity. Finally, Chapter 7 provides an overview of findings of the three experiments and discusses the conclusions, implications, and limitations of this project.

3. Materials

Chapters 4, 5, and 6 describe three experiments that tested listener sensitivity to language-specific segment cues in a nonce-word categorization task. The characteristics of the participants and the procedure of each task vary across the three studies, but the nonce-word stimuli used in all three experiments are the same. This chapter describes the properties of the stimuli used in the project.

3.1 Language-specific target segments

To test listeners' sensitivity to different classes of sounds, target segments were chosen to represent segmental categories unique to one language ("phonemic" cues) and shared categories that are realized at different points along a continuum in each language ("phonetic" cues). Three language-specific phonemic cues were chosen for the categorization task: the English-specific segments /θ/ and /ɹ/, and the Spanish-specific trill /r/. The selection of phonemic cues was limited to those sounds that form categories not present in the other language. For example, the English voiced alveolar approximant /ɹ/ and the Spanish voiced alveolar trill /r/ are not different extremes of a continuum between /ɹ/ and /r/, in the way that English and Spanish voiced and voiceless stops vary along a single dimension (VOT). That is, there is not a single dimension or acoustic correlate that distinguishes /ɹ/ and /r/ that could be increased or decreased to create one from another, since the two sounds are produced with fundamentally different manners of articulation (/ɹ/ as an approximant and /r/ as a trill).[4]

[4] See also the discussion in Flege (1987) about the contrast between the French /y/ and English /u/ representing that of a "new" phone for L1 English learners of French. While [y] exists allophonically in English and the French /y/ may initially be mapped to English /u/, Flege emphasizes that learners will realize that /y/ is distinct both from other French vowels and from English /u/. That is, the L2 category is independent from any L1 category, even if initially perceived as similar. The same argument is made here

Vowels were excluded as phonemic cues since English-specific vowels (e.g. /ɪ/) can be differentiated from the nearest shared vowels (e.g. /i/) by both spectral cues and duration differences. Non-native listeners may rely on vowel duration cues more than native listeners (Escudero, 2006; Kondaurova & Francis, 2008), and the desire here was to ensure as much as possible that both monolingual and bilingual listeners were attending to the same acoustic property of the target segment. For listeners who rely on duration to differentiate vowels, duration creates a continuum between a shorter vowel and a longer vowel, making such a contrast a phonetic cue as defined in this study. Since all five Spanish vowels exist in English, there were no Spanish-specific vowels to consider for phonemic cues.

One additional English-specific cue was identified for inclusion as a phonemic cue, /θ/. Although /θ/ is a phoneme in Peninsular Spanish (it is produced as /s/ in Latin America), it was included as an English-specific phoneme since exposure to Peninsular Spanish among our listeners was expected to be very limited, and native speakers of Peninsular Spanish were excluded from the study. Early Spanish-English bilingual listeners living in Central Texas, where this study was conducted, may have some exposure to Peninsular Spanish, for example through movies, but are most familiar with Latin American dialects of Spanish. The late bilingual participants likely have more exposure to Peninsular Spanish than early bilinguals, but it is not expected that this exposure would be

for the newness of Spanish /r/ for L1 English listeners: despite the possibility that early learners associate Spanish /r/ with English /ɹ/, the Spanish sound is also perceived as very distinct from English /ɹ/ and it further does not exist even allophonically in American English. This is in contrast to the statements Best (1991) makes about these sounds being language-specific instantiations of a category /r/. However, given Flege's description and examples of "similar" phones varying alone continua of VOT or formants, it is clear that the differences between Spanish /r/ and English /ɹ/ are representative of new phones and not of a single, continuous category.

more influential on L1 representations than native dialect phonology. In fact, many monolingual English listeners probably have exposure to the trill /r/ in Scottish English, also through media, but it would be surprising if their language-segment associations reflect occasional exposure to the trill /r/ in English.

In addition to the phonemic cues, phonetic cues were also tested, and these vary along a continuum. These sound categories exist in both languages but their articulation in each language is characterized by subsegmental differences in place of articulation. Two language-specific phonetic segments were chosen for the task, the lateral approximant /l/ and the high back vowel /u/. The lateral approximant is produced as a 'light' [l] at the alveolar ridge in Spanish, while in American English the segment is realized as the 'darker' [ɫ], with an additional closure near the velum, particularly in closed syllables (Recasens, 2004, 2012). The back vowel /u/ differs along F2 in English and Spanish: it is fronted to [ʉ] for many speakers of American English and is produced further back, as [u], in Spanish (Bradlow, 1995; Clopper et al., 2005; Mendez, 1982).

3.2 NONCE WORDS

Nonce words were created to test the contributions of specific sounds to listeners' conceptualizations of Spanish and English while avoiding the influence of lexical knowledge on categorization. All nonce words were disyllabic and were composed of sounds common to both languages, except for a single target segment. This single language-specific segment served as the cue to which language the word must belong. The remaining segments in the nonce words exist in both English and Spanish (at least phonemically, as in the case of the English unstressed schwa allophone of /a/) and were expected to differ minimally between the two languages. By using segments common to English and Spanish, listeners were obligated to use the target segment to make the language categorization decision. The segments identified as common to English and Spanish were the vowels /i, a/, the fricatives /m, f, s, h[5]/, and the affricate /tʃ/. These consonants do not differ between the languages in point of articulation or in voicing. The vowels /i, a/ were chosen from among the five common vowels /i, e, a, o, u/ because they are not mid-vowels (/e, o/), which are diphthongized in American English, and because /u/ is included here as a target segment. Although /a/ is more variable than /i/ across the languages (Bradlow, 1995), it was included to increase the number of possible stimuli.

Table 3.1: Nonce words with language-specific phonemes /θ,ɹ,r/.

English phoneme /θ/	English phoneme /ɹ/	Spanish phoneme /r/
/tʃiθə/	/tʃaɹə/	/tʃira/
/fiθə/	/fiɹə/	/fara/
/hiθə/	/hiɹə/	/fira/
/maθə/	/maɹə/	/mara/
/saθə/	/ɹatʃə/[6]	/mira/
/siθə/	/ɹitʃə/	/ratʃa/[6]
/θitʃə/	/ɹimə/	/ritʃa/
/θisə/	/siɹə/	/sira/

For each target segment, eight nonce CVCV or CV/l/CV words were constructed from the set of segments overlapping in English and Spanish. The CV/l/CV structure was

[5] The phoneme identified here as /h/ is alternately realized as /x/ in some dialects of Spanish (Hualde, 2005). The speaker chosen to record the stimuli uses /h/ in his dialect of Spanish (Colombian).

[6] *Racha* is in fact a Spanish word meaning 'gust of wind.' The analyses in this and subsequent chapters exclude responses to this item and to the English nonce word /ɹatʃə/, since bilingual listeners may have interpreted this English stimulus as the Spanish word *racha* produced with an English accent and not as a uniquely English word.

included to provide two phonological contexts for /l/ stimuli that were both permissible in Spanish and in which /l/ was most likely to be velarized to [ɫ] in American English (Recasens, 2012). Each nonce word was a possible, but non-existent, word in both languages, and all words ended with /a/, which was reduced to [ə] in the English stimuli. See Table 3.1 for the set of stimuli containing language-specific phonemes and Table 3.2 for the set of stimuli containing language-specific phonetic segments.

Table 3.2: Nonce words with language-specific phonetic variations of /l,u/.

/l/		/u/	
English	Spanish	English	Spanish
[tʃaɫsə]	[tʃaltʃa]	[tʃʉtʃə]	[tʃuma]
[faɫmə]	[filfa]	[fʉtʃə]	[fufa]
[hiɫfə]	[lafa]	[fʉfə]]	[fusa]
[ɫitʃə]	[litʃa]	[fʉsə]	[mufa]
[ɫifə]	[lifa]	[hʉtʃə]	[muma]
[maɫfə]	[malfa]	[hʉsə]	[sutʃa]
[saɫfə]	[silma]	[mʉmə]	[hutʃa]
[siɫtʃə]	[halfa]	[sʉfə]	[husa]

The inclusion of disyllabic words meant that the second English vowel was reduced to schwa, resulting in an additional vowel-quality cue beyond the language-specific target segment. However, this strategy was preferred to the development of monosyllabic words; Spanish has relatively few monosyllabic words compared to English (cf. Costa & Caramazza, 1999) so monosyllables could be biased towards English responses. The final open syllable was also used consistently since the set of possible word-final consonants in

Spanish is very small: /ð, s, n, l, ɾ/. Some of these are subject to lenition (/ð/) or aspiration (/s/), or are already included as a language-specific target segment (/l/). Word-final Spanish /ɾ/ is also in free variation with /r/, a target segment. The inclusion of a second syllable and vowel reduction was therefore preferred. Vowel reduction and its potential influence on listeners' language decisions are addressed in the discussion (see Section 4.4).

3.3 STIMULI RECORDINGS AND SPEAKER

A single speaker recorded both English and Spanish stimuli, and this was crucial to the experimental task. Having all stimuli in one voice was preferred over recording two monolinguals to avoid voice being a cue to language, and using natural productions of the stimuli ensured there were no acoustic artefacts from splicing or otherwise manipulating segments within a word frame. Using natural productions from a single talker also permitted the selection of any segment as target segment, regardless of difficulties isolating the single segment (e.g. with the English /ɹ/) to be cut and copied into a common word frame.

Since it was also important for the stimuli to lack any additional language-specific cues, or accent, beyond the controlled target segment, care was taken to recruit a nearly balanced Spanish-English bilingual who produced both languages as natively as possible. The chosen talker was a 37-year-old Spanish-English bilingual who was born and raised in Colombia until the age of 7 at which point he moved to the state of New York with his family. He continued to speak Spanish at home in New York, and as an adult he moved to Texas for graduate school, during part of which he lived in Guatemala and Spain to conduct research. While most of his current daily interactions were in English, he also used Spanish on a daily basis with his family and frequently for translating and interpreting professionally at work.

The English and Spanish nonce words were recorded in separate sessions to further ensure minimal cross-linguistic transfer. The nonce words were recorded on a MOTU UltraLite-mk3 Hybrid recorder at a sampling frequency of 44.1 kHz (16 bit) in a sound-attenuated booth. The talker repeated each nonce word three times so that the clearest repetition could be chosen. The words were written in English and Spanish orthography (e.g. English *leefuh* for [ɫifə] and Spanish *chirra* for /tʃira/) and not in the International Phonetic Alphabet (IPA), so for some items the talker was coached to understand the intended pronunciation. The pitch contours of the selected repetitions varied, particularly when the talker was unsure of the pronunciation, so the pitch contours of all stimuli were manipulated to match a naturally-produced token with a falling contour. Using Praat (Boersma & Weenink, 2012), the first pitch point of each stimulus word was moved to 170Hz and the last pitch point was moved to 124Hz, to match the values of the model token, and the intervening pitch points were deleted from all tokens, including from the model. When the original pitch tier was replaced with the manipulated pitch tier, Praat interpolated pitch points along the line from 170Hz to 124Hz at each pulse.

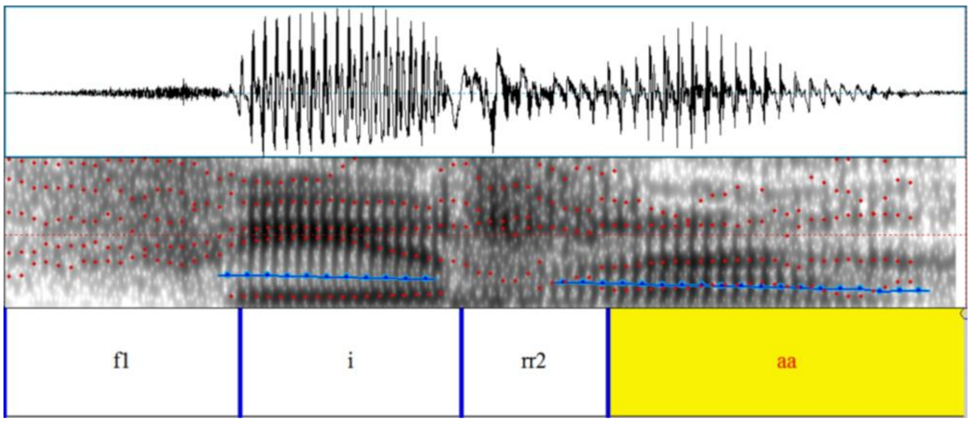
Acoustic properties of the stimuli

Segmental properties of each stimulus were also measured using Praat to ensure that the Spanish and English productions differed in the expected dimensions. See Figure 3.1 for sample waveforms and spectrograms from Spanish and English. The text grid below the spectrogram delineates each segment of the word, and formants are tracked in red and F0 in blue. The duration and first three formants of the filler vowels in both syllables of the nonce words were measured, and the same measurements were collected for the target /l, u/ segments. Recall that the vowels /i,a/ were used in the first vowel position of the disyllabic nonce words to create a sufficient number of non-word stimuli, and the second vowel (V_2) of each nonce word was realized as the full-vowel [a] in Spanish words and as the reduced [ə] in English words. See Table 3.3 for the mean and standard deviation of the duration and formant values for each segment. The English and Spanish productions of the non-target vowels are reported in (A), and the measurements of the language-specific variants of the target segments /l, u/ are in panel (B). Formant values are the mean of the measurements taken at the midpoint of each vowel. Standard deviations are included in parentheses.

In order to test whether the English and Spanish variants were distinct from each other, the concordance statistic (c-statistic) of a logistic regression model was analyzed. The c-statistic is the proportion of outcomes that are correctly predicted by the fitted model. Constructing such a model for the c-statistic was preferable to testing for differences between duration and each formant separately since listeners hear the multiple acoustic cues at once; that is, listeners may attend to differences in all three dimensions (F1, F2, and duration), so all three should be considered together when determining if the sounds were distinct in English and Spanish.

Figure 3.1: Sample waveforms and spectrograms of Spanish (A) and English (B) nonce words.

(A) Spanish nonce word *firra* /fira/



(B) English nonce word *chalsuh* [tʃałsə]

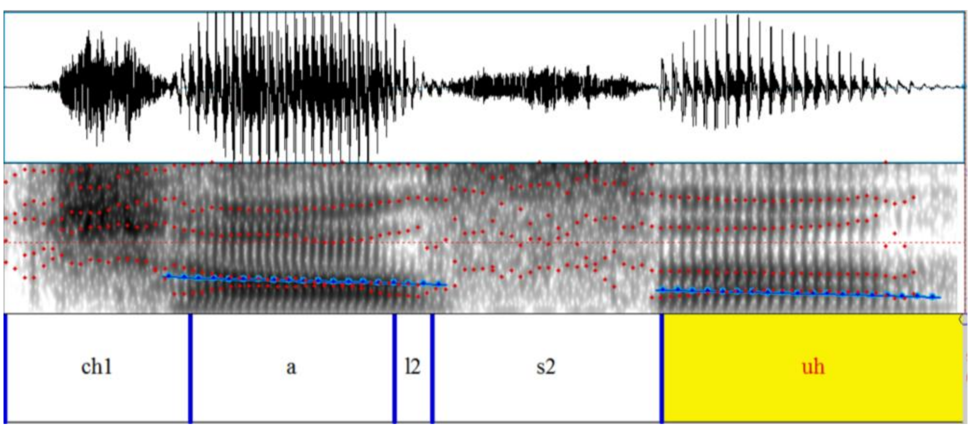


Table 3.3: Acoustic properties of filler vowels (A) and target segments (B).

(A) Filler vowels

	Duration (ms)		F1 (Hz)		F2 (Hz)	
	English	Spanish	English	Spanish	English	Spanish
/i/	87.0 (22.6)	95.6 (20.3)	369.7 (47.4)	361.0 (31.9)	2245.3 (243.7)	2196.3 (107.9)
/a/	116.9 (19.0)	99.1 (14.4)	878.8 (67.4)	835.7 (15.1)	1189.4 (74.6)	1524.6 (55.1)
V_2	174.4 (29.0)	141.5 (31.4)	693.7 (67.6)	769.8 (130.8)	1367.4 (143.3)	1484.5 (97.7)

(B) Target segments

	Duration (ms)		F1 (Hz)		F2 (Hz)		F3 (Hz)	
	English	Spanish	English	Spanish	English	Spanish	English	Spanish
/l/	63.8 (22.9)	77.7 (17.9)	581.6 (134.7)	383.4 (88.3)	1141.4 (260.3)	1917.4 (331.8)	2999.2 (253.4)	2937.6 (375.9)
/u/	81.7 (11.9)	82.7 (18.3)	415.8 (22.2)	484.5 (170.9)	1560.9 (178.5)	1174.0 (372.5)		

For each vowel, a logistic regression model was constructed in R (RStudio 0.99.489) using the *rms* package (v4.2-1) with language (English, Spanish) as the dependent variable and the duration and midpoint measures of F1 and F2 as fixed effects.

The model for English and Spanish /l/ additionally included the midpoint measure of F3 as a fixed effect. All measurements were centered and scaled.

For the two target segments that were measured, /l/ and /u/, it was expected that the formants and the duration of the segment would be sufficient to distinguish the English and Spanish variants. The model with these three main effects as well as the midpoint of F3 made perfect discrimination between Spanish [l] and English [ɫ] (C=1.000). For Spanish [u] and English [ʉ], the model was also successful in discriminating one variant from the other (C=1.000).

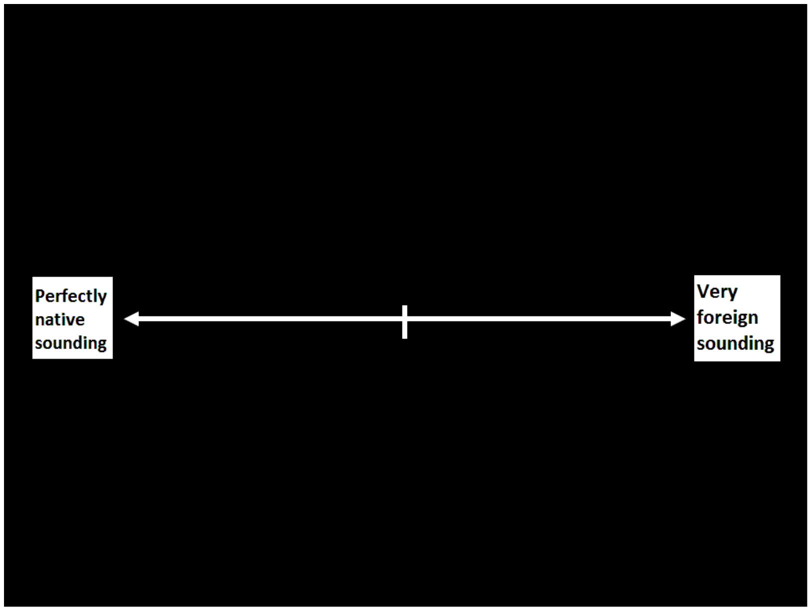
The other three segments were filler vowels: the two vowels /i, a/, which were used in the first syllables of the nonce words, and the final vowel of the stimuli. The model for /i/ produced a c-statistic of 0.681, which represents a moderately good fit to the differences between English and Spanish /i/, but which falls short of the clear distinction between the phonetic variants described above. For /a/ in the position of nucleus of the first syllable, the model was highly successful for discrimination (C=1.000). Finally, the model for the second (unstressed) vowel in the nonce words fit well (C=0.853). The acoustic distance between English and Spanish vowels in stressed and unstressed positions was expected (cf. Bradlow, 1995), and see Section 3.4 for a discussion of how the results of Experiment 1 should be understood in light of these additional acoustic differences.

3.4 Accentedness rating study

To ensure that the talker's productions were as native-like as possible in both languages, an accentedness rating study was completed. Native English and native Spanish listeners rated the nativeness of the productions of eight talkers, including the stimuli talker. All talkers recorded Æsop's *The North Wind and the Sun* in Spanish and English, and the final set of eight talkers included one male and one female from each of the following four groups: monolingual English talkers, L1 English talkers who learned Spanish late and had completed college and graduate coursework in Spanish, L1 Spanish talkers from Latin America who learned English late and had moved to the U.S. to attend college, and early Spanish-English bilinguals (including the stimuli talker). The recordings from these eight talkers were divided into seven phrases, yielding 56 sound files of the talkers' English and 56 sound files of their Spanish.

Ten monolingual English listeners and 10 L1 Spanish listeners from Latin America who learned English after age 14 were recruited for the accentedness rating experiment. None participated in the main study. Listeners heard productions in their native language and decided how native-sounding or foreign-sounding each production was by using the mouse to click on a horizontal line. The line appeared on the screen after the audio presentation of each sentence and represented a continuum between "Perfectly native sounding" (labeled as such at the left extreme) and "Very foreign sounding" (so labeled at the right extreme). See Figure 3.2. The Spanish translations "Suena totalmente nativo" and "No suena nada nativo" were used in the Spanish version with native Spanish listeners and the talkers' Spanish productions. The accentedness rating was recorded as the x-intercept of the mouse at the click. The 56 sentences were randomized for each listener.

Figure 3.2. Screen image presented to English listeners in the accentedness rating study.



Accentedness ratings were converted to *z*-scores to account for listeners using the continuum differently; see Table 3.4 for the mean accentedness rating in English and Spanish for each talker. Standard deviations are included in parentheses, and the scores for the stimuli talker are in bold. The *z*-transformed accentedness ratings for English and Spanish productions were submitted to separate mixed-effects linear regression models using the *lme4* (v1.1-7) and *lmerTest* (v2.0-20) packages in R (RStudio 0.99.489). Listener was included as a random effect on the intercept. Including talker as a fixed effect significantly improved the fit of a model with the random intercept alone, for both the English model (χ^2=1317.3, df=7, p<0.001) and the Spanish model (χ^2=948.25, df=7, p<0.001). The stimuli talker (early bilingual male) was designated as the referent class for the talker variable in both models.

Table 3.4: Mean accentedness rating, as a z-score, for English and Spanish productions.

	Mono male	Mono female	L1Eng male	L1Eng female	**Early male**	Early female	L1Span male	L1Span female
English	-0.58 (0.31)	-0.76 (0.21)	-0.71 (0.17)	-0.47 (0.38)	**-0.63 (0.23)**	-0.02 (0.47)	1.49 (0.36)	1.69 (0.19)
Spanish	1.40 (0.23)	1.37 (0.18)	0.42 (0.72)	-0.21 (0.54)	**-0.87 (0.29)**	-0.42 (0.52)	-0.89 (0.34)	-0.80 (0.33)

The model summary for the English productions is presented in Table 3.5. Figure 3.3 displays the mean accentedness rating of each talker's English productions and standard error bars illustrating 97.5% confidence intervals. The different listener groups are represented by different colors, and the male speakers' ratings are in the striped bars; the green striped bar, in the black box, thus represents the stimuli talker of interest here. The left extreme of the x-axis represents "Perfectly native sounding" and the right extreme represents "Very foreign sounding." The intercept for the stimuli talker's English productions was significantly less than zero ($p<0.001$) and was thus closer to the "Perfectly native sounding" (left) extreme than to the center. The stimuli talker's English was not rated as significantly different from the monolingual English male ($p=0.29$) or the L1 English male ($p=0.12$), and he was rated as significantly more native sounding than all other talkers (at least $p<0.01$) except for the monolingual English female ($p<0.05$).[7]

[7] The monolingual English female was also rated as significantly more native sounding than the monolingual English male ($p<0.001$) and the L1 English female ($p<0.001$), who were also raised as monolingual English speakers. The speed with which the monolingual English female read the story may have influenced how accented she was rated (cf. Munro & Derwing, 2001), but importantly, the stimuli

Table 3.5: Summary of mixed-effects linear regression model fitting accentedness ratings of English productions.

English productions				
Predictor	*Estimate*	*Standard Error*	*t value*	*p value*
Intercept (Stimuli talker)	-0.632	0.037	-17.186	<0.001
Mono male	0.0554	0.052	1.065	NS
Mono female	-0.131	0.052	-2.516	<0.05
L1 Eng male	-0.082	0.052	-1.575	NS
L1 Eng female	0.163	0.052	3.127	<0.01
Early female	0.613	0.052	11.785	<0.001
L1 Span male	2.123	0.052	40.797	<0.001
L1 Span female	2.318	0.052	44.537	<0.01

Random effects	*Variance*
Listener	<0.001
Residual	0.095

talker's accent in English was not rated different from two male talkers who grew up as monolingual English speakers.

Figure 3.3: Mean accentedness ratings for eight talkers' productions of English.

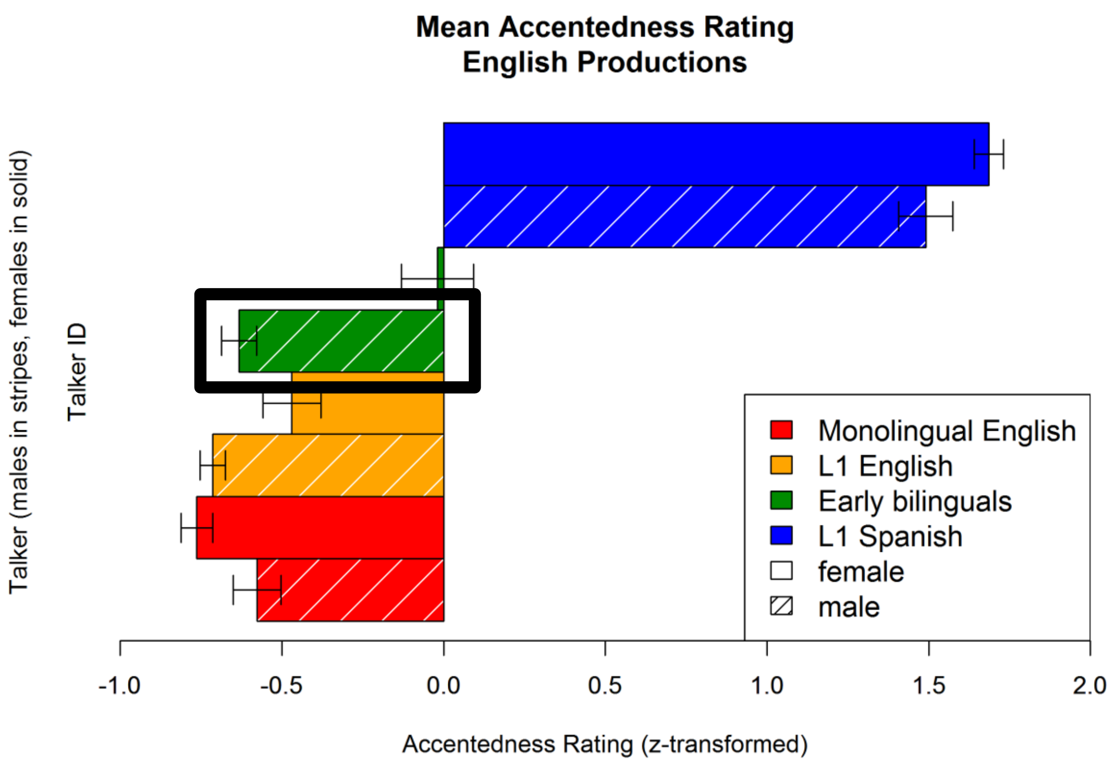


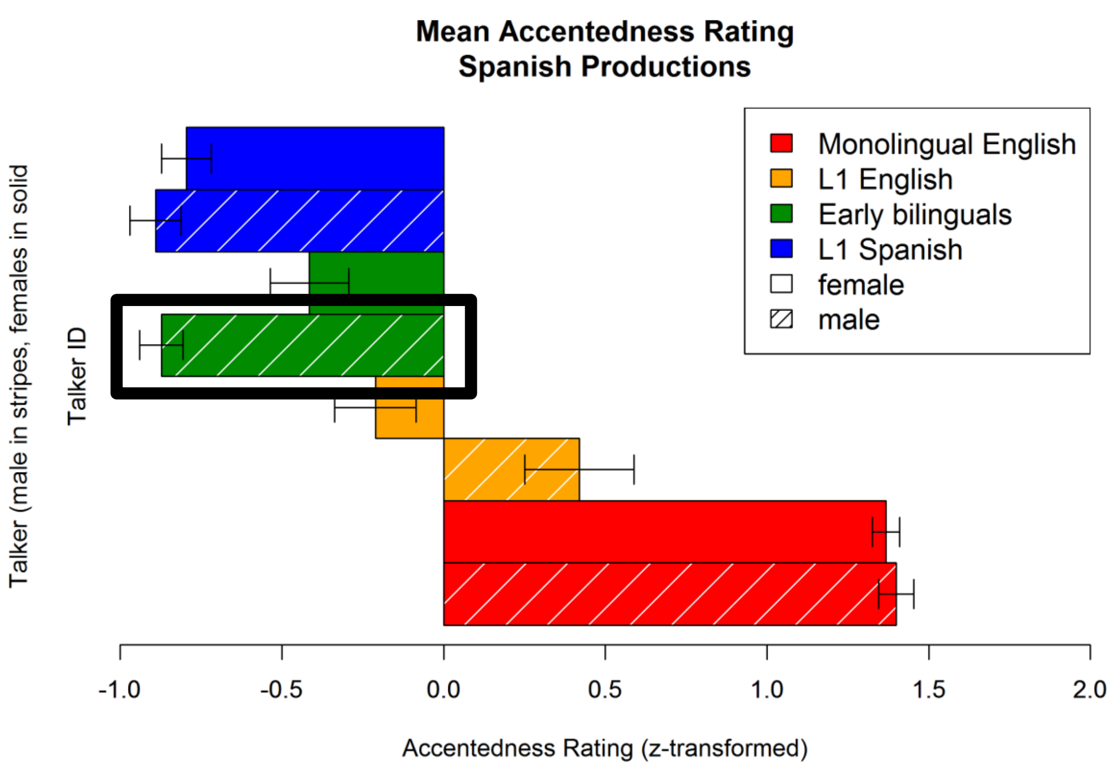
Table 3.6 presents the model summary for the Spanish productions, and the mean accentedness ratings for Spanish are displayed in Figure 3.4. As in Figure 3.3, the stimuli talker is represented by the green striped bar, in the black box. Values on the left of the x-axis indicate speech that was rated as "Perfectly native sounding" and values on the right represents ratings of "Very foreign sounding." The stimuli talker's Spanish rating was significantly less than zero ($p<0.001$), indicating that he was rated closer to "Perfectly native sounding" (the left extreme) than to the center for his Spanish productions as well. His Spanish was also rated as significantly more native sounding than all the other talkers ($p<0.001$), except for the L1 Spanish male and female, with whom there was no significant difference in rating (for L1 Spanish male, $p=0.80$; for L1 Spanish female, $p=0.29$).

Table 3.6: Summary of mixed-effects linear regression model fitting accentedness ratings of Spanish productions.

Spanish productions				
Predictor	*Estimate*	*Standard Error*	*t value*	*p value*
Intercept (Stimuli talker)	-0.873	0.051	-17.062	<0.001
Mono male	2.272	0.072	31.404	<0.001
Mono female	2.241	0.072	30.970	<0.001
L1 Eng male	1.292	0.072	17.861	<0.001
L1 Eng female	0.661	0.072	9.144	<0.001
Early female	0.458	0.072	6.323	<0.001
L1 Span male	-0.018	0.072	-0.255	NS
L1 Span female	0.077	0.072	1.070	NS

Random effects	*Variance*
Listener	<0.001
Residual	0.183

Figure 3.4: Mean accentedness ratings for eight talkers' productions of Spanish.



Note that the stimuli talker is the only talker whose productions were described as native-like (extending significantly below 0, to the left) for both English and Spanish productions. The results of this ratings study indicate that the stimuli talker's English is perceived to be native-like, and it is unlikely that his productions of the English nonce-word stimuli were foreign-accented. Likewise, the stimuli talker's Spanish was rated as native as the speech of the L1 Spanish talkers who lived in Spanish-speaking countries until age 18 and learned English after puberty. The stimuli talker's Spanish is thus perceived to be native-accented, and his productions of the Spanish nonce-words are unlikely to have been produced with an English accent.

4. Experiment 1: Adults' perception of language-specific acoustic cues

4.1 Introduction

This chapter presents the results of Experiment 1, which tested adults' sensitivity to English- and Spanish-specific segmental cues and compared the categorization patterns of monolingual English listeners and early and late Spanish-English bilinguals. Most evidence indicates that monolingual and early bilingual listeners can differentiate language-specific segments in isolation (although see Sebastián-Gallés & Soto-Faraco, 1999, and Sebastián-Gallés et al., 2005), but fewer studies have tested how listeners use such cues in word-length stimuli. Conflicting results have emerged from the literature with word-length stimuli as well: in Flege & Munro (1994), monolingual listeners were able to judge the language in which *taco* had been produced, while in Sebastián-Gallés et al. (2005) and Amengual (2015), early bilinguals could not reliably identify mispronunciations varying in a single segment. It remains unclear whether listeners attend to all language-specific acoustic cues equally, or which segmental cues are most strongly associated with language. These findings also call into question the sensitivity of bilingual listeners to sounds from both of their languages.

The present study tests listener sensitivity to two classes of language-specific segments: phonemic cues and phonetic cues. These categories are parallel to the new vs. similar distinction made in second language acquisition literature (Flege, 1987, 1995). Evidence suggests that sound categories that are "new" to an L2 and have no counterpart in the L1, like the phonemic cues proposed here, are easier to produce authentically than "similar" L2 phones. Similar phones differ along some particular parameter from the L1 variant, e.g. VOT or F1, and are like the phonetic cues used here. Flege's Speech Learning Model (SLM) predicts that L2 phones that are similar to L1 phones will be more difficult to perceive and produce like native speakers do. Therefore listeners in this task are expected to categorize phonemic segments more accurately than phonetic segments.

Experiment 1 also compares listeners from three different language backgrounds: English monolinguals and early and late Spanish-English bilinguals. While there is evidence indicating that early bilinguals are as sensitive to L2 contrasts as native speakers (Flege et al., 1999; Flege & MacKay, 2004; Mack, 1999), listeners in these studies are typically only tested in contrasts from their L2. In this study bilinguals judge segments from both languages to more completely understand the role of language experience in cue salience. Spanish-English bilingual listeners who have extensive exposure to the phonologies of both languages are expected to categorize language-specific segments more accurately and faster than monolingual listeners, given their competence in Spanish and previous work showing successful L2 discrimination for these groups.

Experiment 1 seeks to answer the following research question and tests the following predictions:

1. Are listeners able to use phonemic and phonetic segments to distinguish Spanish and English, and does language experience affect sensitivity to the cues?
    a. Listeners are expected to be more sensitive to phonemic cues, modeled on the new phones in Flege's SLM, than to phonetic cues, which are like similar phones in the SLM. Phonemic cues should be easier for listeners of all language backgrounds to perceive as distinct categories, and so are hypothesized to be easier to associate with the appropriate language.
    b. Bilingual listeners are expected to categorize the nonce words more accurately than monolingual listeners, since bilinguals have linguistic representations of both Spanish and English and can contrast these representations. Previous work

suggests early bilinguals will be as sensitive to English segments as monolingual listeners, so the early bilingual advantage over monolinguals might be limited to the Spanish-specific segments. Early bilinguals are expected to show greater sensitivity to English-specific cues than the late learners of English.

4.2 METHODOLOGY

4.2.1 Participants

Listeners were recruited in one of two ways. The majority of participants (n=53) were contacted through the Department of Linguistics subject pool and received course credit for their participation. To supplement the subject pool participants, the remaining participants (n=27) were Spanish-English bilinguals, both early and late, who responded to an advertisement in the University of Texas Events Calendar soliciting volunteers to participate in linguistics studies. These participants were paid $10/hour for their time.

Listeners completed a language history questionnaire (Chan, 2014) that included questions about participants' biographical information, the places they had lived and for how long, their language exposure and proficiency, and their language(s) of education. Based on their responses to the questionnaire, participants were divided into three groups: monolingual English speakers with minimal or no exposure to Spanish (Monolingual), Spanish-English bilinguals from the U.S. who acquired both languages in early childhood (Early Bilinguals), and Spanish-English bilinguals from Spanish-speaking countries who acquired English as adults (Late Bilinguals). Participants who did not fit into one of these groups were not included in the final sample (n=24). See Table 4.1 for a summary of participant characteristics.

Forty participants (21 females) were included in the Monolingual group. All members of this group were from the U.S., had heard English from birth, did not hear another language at home, and were not proficient in any other language. Participants ranged in age between 18 and 29, and the mean age of the group was 20. Of the 40 Monolingual listeners, 24 had studied Spanish in middle and/or high school. One additional participant had some Spanish classes in elementary school, and one further participant reported learning some Spanish as a toddler outside the home. All 26 listeners with some exposure to Spanish reported very low proficiency in the language.

The Early Bilingual group included 18 participants (15 females) who ranged in age from 18 to 29, with a mean age of 20 years. Eleven of the listeners in the Early Bilingual group were born and raised in the United States, and the remaining seven participants were born in Mexico (n=6) or Colombia (n=1) and moved to the U.S. before they began elementary school. All listeners in the Early Bilinguals group had learned Spanish at home since birth. Seven participants also learned English at home since birth (four of the U.S.-born participants, three of the foreign-born participants). The remaining 11 participants began learning English when they started elementary school.

Twenty-two listeners (11 females) were categorized as Late Bilinguals since they were born and raised in a Spanish-speaking country and moved to the U.S. after age 14. Listeners in this group ranged in age between 18 and 43, with a mean age of 28 years. Only Late Bilinguals from Latin America participated; listeners from Spain were excluded since /θ/ is phonemic in Peninsular Spanish and the present study included /θ/ as an English-specific phoneme. Listeners were from Mexico (n=11), Argentina (n=2), Peru (n=2), Ecuador (n=2), Bolivia (n=1), Venezuela (n=1), Colombia (n=1), the Dominican Republic (n=1), or some combination of these countries (n=1). Late Bilinguals ranged in the age at which they moved to the U.S. between 14 and 28, with mean age of arrival of 20. All listeners had learned only Spanish at home since birth. Although all had studied English at least informally in school before they moved to the U.S., Spanish was the only language of instruction in both primary and secondary school for all Late Bilingual participants.

Table 4.1: Demographic information and language background for participants in Experiment 1.

	Monolinguals	Early Bilinguals	Late Bilinguals
N	40	18	22
mean age	20	20	28
age range	18-29	18-29	18-43
females	21	15	11
mean age (in years) when learned English	0	3.7	10
mean age (in years) when learned Spanish	12.5	0	0
mean age (in years) when moved to U.S.	NA	1.3	20.1

4.2.2 Procedure

Participants completed the nonce-word categorization experiment in the UT Phonetics Lab in the Department of Linguistics at the University of Texas at Austin. The experimenter obtained written informed consent from the participant before beginning the study, in accordance with the recommendations of the Institutional Review Board at UT

Austin. Listeners answered an online language history questionnaire and were tested for normal hearing, and this was followed by the categorization experiment.

Listeners performed the language categorization task in a sound-attenuated booth on a PC running E-Prime 2.0 (Psychology Software Tools, 2010). Listeners wore Sennheiser HD 280 Pro headphones and were oriented to the serial response button box (Psychology Software Tools, 2003). Participants were instructed to place their index and middle fingers on the two leftmost buttons, which were labeled with "ENG" and "SPAN," the order of which was counterbalanced across participants. The language that corresponded to each button was also presented on the computer screen, e.g. "ENGLISH" appeared on the left side of the screen for the group of participants who used the left button to indicate English words. Listeners began with a practice block in which they first read instructions presented on-screen and then heard 20 real words (10 English, 10 Spanish). Listeners decided if each word sounded more like English or more like Spanish.

After the practice block, the test portion began. At test, listeners were told they would hear "snippets of speech that were taken out of longer recordings while the speaker was talking in either English or Spanish," and they were asked to decide if what they heard sounded more like it came from the English recording or the Spanish recording. This wording and context was provided after piloting indicated that some listeners had the impression that they were hearing accented productions instead of words from two languages. That is, pilot participants reported confusion in deciding whether a stimulus was (for example) an English word or an English-accented Spanish word, and so whether they should press the English or Spanish button. To avoid this confusion between accent and language, the categorization task was rephrased to ask about the language being used to produce the word. Listeners categorized the 56 nonce words eight times, and stimuli were randomized within each of the eight blocks, for a total of 448 trials. There was a one second pause between a listener's response and the onset of the audio for the next stimulus. Reaction time (RT) was calculated from the onset of the audio file, and categorization decision and RT were recorded for each trial.

4.3 Results

Categorization decision (Spanish or English) and reaction time (RT) were recorded for each trial. Decisions were coded as accurate if words containing the English-specific phoneme /ɹ/ or /θ/ or the English variants [ɫ] or [ʉ] were classified as English and if words with the Spanish-specific phoneme /r/ or the Spanish variants [l] or [u] were classified as Spanish. Trials with the Spanish stimulus *racha* /ratʃa/ and the English stimulus *rachuh* /ɹatʃə/ were excluded from the analyses (cf. Section 3.2 and footnote 5). RTs were calculated by subtracting the length of the stimulus .wav file from the time calculated by E-Prime between trial onset and button press. This ensured that the RTs analyzed here reflected the length of time for the listener to make a categorization decision, after hearing the end of the stimulus word. Trials with negative RTs (n=665; 1.9%) thus reflected responses that were made before the end of the sound file had played and were discarded as spurious responses. Trials more than three standard deviations above or below a participant's mean were excluded as outliers (n=228; 0.7%). The spurious responses and outliers accounted for 2.6% of all trials (n=893), after Spanish *racha* and the English *rachuh* were removed. RTs were log-transformed from milliseconds to normalize the distribution of responses for the regression analyses. Less than 0.5% of the remaining responses exceeded 5000ms and the distance of these from the mean was reduced in the log transformation. The following analyses include the remaining 33,667 trials (Monolinguals: n=16,800; Early Bilinguals: n=7,441; Late Bilinguals: n=9,426). Accuracy (correct, incorrect) and log-transformed RT were submitted to separate regression analyses, which were analyzed using Bayesian inference with the *glmer2stan* package (v0.995) in R (v3.2.2) to interface with STAN via RStan (v2.8.2).

Table 4.2: Mean accuracy of each listener group for each stimulus type in Experiment 1.

		Monolinguals	Early Bilinguals	Late Bilinguals
Phonemic Cues	Spanish /r/	95.6 (20.5)	96.9 (17.3)	97.6 (15.4)
	English /ɹ/	87.5 (33.1)	93.0 (25.6)	96.1 (19.3)
	English /θ/	60.0 (49.0)	66.5 (47.2)	77.3 (41.9)
Phonetic Cues	Spanish [l]	93.0 (25.4)	95.0 (21.9)	91.8 (27.5)
	English [ɫ]	74.8 (43.4)	79.9 (40.1)	83.5 (37.1)
	Spanish [u]	88.9 (31.4)	85.7 (35.0)	89.5 (30.6)
	English [ʉ]	66.2 (47.3)	66.5 (47.2)	68.5 (46.5)

4.3.1 Accuracy analysis

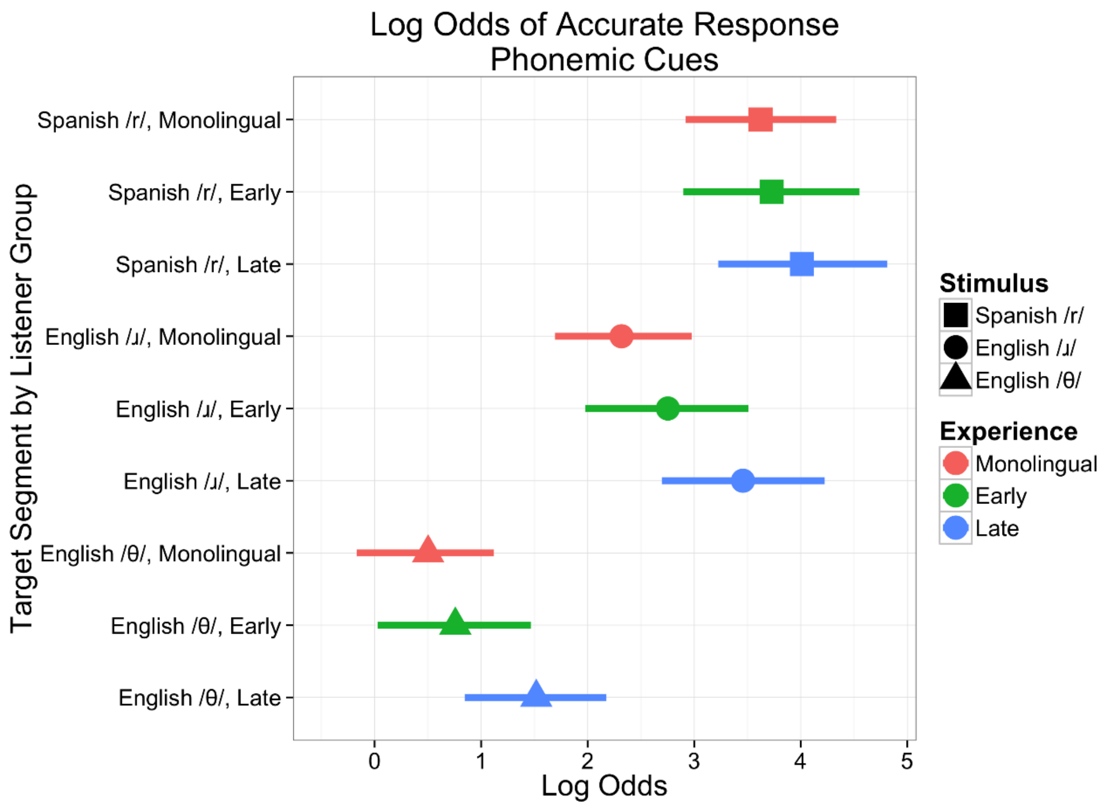
Mean accuracy rates and standard deviations for each stimulus type for each listener group are reported in Table 4.2. The accuracy results were analyzed using Bayesian mixed effects logistic regression models. The models were fit via a Markov Chain Monte Carlo procedure using STAN (Gelman et al., 2015). Model comparison was performed using the Deviance Information Criterion (DIC; Spiegelhalter et al., 2002). For all models, participant and stimulus word were included as random effects on the intercept. First a model was constructed with these random effects and a fixed effect of stimulus class (phonemic, phonetic), and this was compared to a model with the random effects and a fixed effect of target segment (seven levels: /r, ɹ, θ/ and [l, u, ɫ, ʉ]). The model with the seven levels of target segment provided a better fit to the data than the model with the stimuli divided only by phonemic or phonetic cue, indicating that variation in accuracy was better explained at the level of segment than by considering the stimuli as representative of the phonemic and phonetic categories. The next model tested included target segment and listener language group (three levels: Monolingual, Early Bilingual, Late Bilingual) as fixed effects, and this improved the fit over the model with target segment as the sole fixed effect. A model with an interaction between these two fixed effects provided an improved fit over a model without the interaction term. See Table 4.3 for the model summary. The reference group, reflected in the model intercept, represents the accuracy of Monolinguals categorizing stimuli with the Spanish phonemic cue /r/. The fitted log odds of accuracy for each target segment and listener language group are plotted in Figure 4.1, with the phonemic cues in panel (A) and the phonetic cues in panel (B). The error bars represent the 95% Bayesian credible intervals.

Table 4.3: Summary of mixed effects logistic regression model fitting accuracy results in Experiment 1.

Predictor	*Mean*	*Posterior SD*	*95% CI*	*p value*
Intercept (Monolingual, Spanish /r/)	3.623	0.358	(2.918, 4.332)	<0.0001
English /ɹ/	-1.305	0.437	(-2.151, -0.444)	<0.0001
English /θ/	-3.121	0.429	(-3.958, -2.268)	<0.0001
Spanish [l]	-0.481	0.447	(-1.375, 0.414)	<0.10
English [ɫ]	-2.118	0.420	(-2.955, -1.279)	<0.0001
Spanish [u]	-1.133	0.417	(-1.949, -0.293)	<0.001
English [ʉ]	-2.837	0.415	(-3.671, -2.028)	<0.0001
Early Bilinguals	0.104	0.348	(-0.575, 0.785)	NS
Late Bilinguals	0.389	0.313	(-0.215, 0.983)	NS
English /ɹ/ * Early	0.331	0.269	(-0.205, 0.843)	<0.05
English /θ/ * Early	0.152	0.241	(-0.343, 0.610)	<0.0001
Spanish [l] * Early	0.016	0.277	(-0.545, 0.532)	NS
English [ɫ] * Early	0.177	0.249	(-0.322, 0.653)	<0.0001
Spanish [u] * Early	-0.670	0.251	(-1.179, -0.196)	<0.0001
English [ʉ] * Early	-0.200	0.240	(-0.687, 0.242)	<0.0001
English /ɹ/ * Late	0.751	0.270	(0.222, 1.274)	NS
English /θ/ * Late	0.267	0.230	(0.172, 1.082)	<0.0001
Spanish [l] * Late	-0.802	0.248	(-1.303, -0.326)	<0.05
English [ɫ] * Late	0.286	0.238	(-0.190, 0.761)	<0.0001
Spanish [u] * Late	-0.498	0.242	(-0.977, -0.021)	<0.001
English [ʉ] * Late	-0.285	0.228	(-0.739, 0.161)	<0.0001

Random effects	*Variance*
Listener	0.894
Stimulus Word	0.775

Figure 4.1: Log odds of accuracy for phonemic cues in Experiment 1.

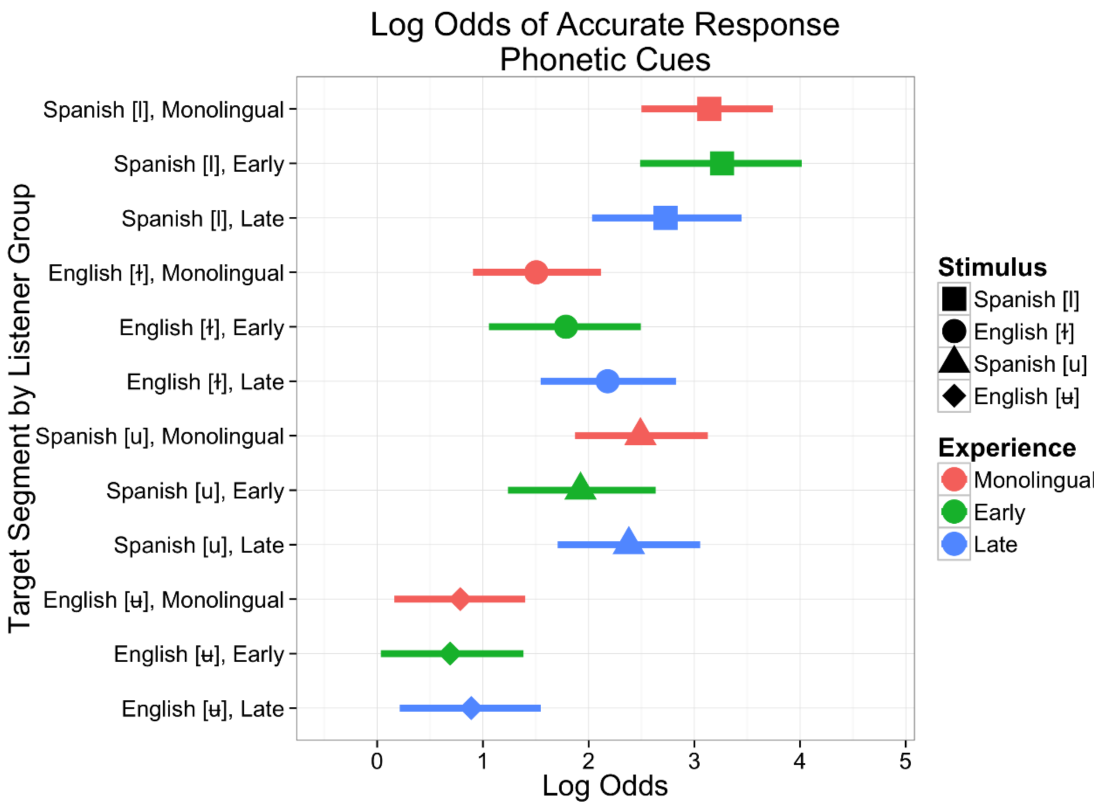


Phonemic cues

Overall, listeners categorized the Spanish segment /r/ the most accurately, followed by the English /ɹ/, and finally the English /θ/. The accuracy of Spanish /r/ was significantly greater than of English /ɹ/ for Monolinguals (β=1.305, posterior SD=0.103, p<0.0001) and Early Bilinguals (β=0.974, posterior SD=0.481, p<0.01), and marginally significant for Late Bilinguals (β=0.554, posterior SD=0.488, p<0.10). All listener groups were significantly more accurate categorizing stimuli with Spanish /r/ than English /θ/ (Monolinguals: β=3.121, posterior SD=0.429, p<0.0001; Early: β=2.97, posterior SD=0.467, p<0.0001; Late: β=2.49, posterior SD=0.468, p<0.0001). The difference in accuracy between English /ɹ/ and English /θ/ also reached significance for all three listener groups (Monolinguals: $\beta=1.815$, posterior SD=0.422, $p<0.0001$; Early: $\beta=1.995$, posterior SD=0.440, $p<0.0001$; Late: $\beta=1.940$, posterior SD=0.445, $p<0.0001$).

There were significant interactions between phonemic segment and listener language background for both English segments, but not for Spanish /r/. For English /ɹ/, Late Bilinguals performed significantly more accurately than Monolinguals ($\beta=1.140$, posterior SD=0.280, $p<0.001$) and Early Bilinguals ($\beta=0.705$, posterior SD=0.353, $p<0.05$). The difference in accuracy between Monolinguals and Early Bilinguals for English /ɹ/ did not reach significance ($\beta=0.435$, posterior SD=0.306, $p=0.10$). With respect to stimuli containing the English /θ/, Late Bilinguals also outperformed Monolinguals ($\beta=1.015$, posterior SD=0.244, $p<0.001$) and Early Bilinguals ($\beta=0.760$, posterior SD=0.306, $p<0.05$), and there was again no difference between the latter two groups ($\beta=0.255$, posterior SD=0.279, $p=0.22$). There were no differences in accuracy among the groups for stimuli with the Spanish /r/ (Monolingual vs. Early: $\beta=0.104$, posterior SD=0.348, $p=0.39$; Monolingual vs. Late: $\beta=0.389$, posterior SD=0.313, $p=0.13$; Early vs. Late: $\beta=0.285$, posterior SD=0.398, $p=0.24$). In sum, all groups very accurately categorized Spanish /r/ and struggled with English /θ/, and Late Bilinguals outperformed their peers in categorizing the English phonemes /ɹ/ and /θ/.

Figure 4.2: Log odds of accuracy for phonetic cues in Experiment 1.



Phonetic cues

Accuracy for both Spanish phonetic cues was significantly greater than for their English counterparts for all listener groups. Spanish [l] was categorized more accurately than English [ɫ] (Monolinguals: β=1.637, posterior SD=0.405, p<0.0001; Early Bilinguals: β=1.476, posterior SD=0.426, p<0.0001; Late Bilinguals: β=0.548, posterior SD=0.415, p<0.05), and Spanish [u] was categorized more accurately than English [ʉ] (Monolinguals: β=1.704, posterior SD=0.404, p<0.0001; Early: β=1.233, posterior SD=0.411, p<0.0001; Late: β=1.490, posterior SD=0.409, p<0.0001). Responses to Spanish [l] were more accurate than to Spanish [u] for Monolinguals (β=0.652, posterior SD=0.414, p<0.05) and Early Bilinguals (β=1.338, posterior SD=0.439, p<0.0001), but there was no difference for the Late Bilinguals (β=0.348, posterior SD=0.422, *p*=0.16). All listener groups categorized the English variant [ɫ] more accurately than the English [ʉ] (Monolinguals: β=0.719, posterior SD=0.396, *p*<0.05; Early: β=1.096, posterior SD=0.406, *p*<0.001; Late: β=1.290, posterior SD=0.403, *p*<0.0001).

There were significant, or near significant, interactions between target segment and listener language background for both Spanish phonetic variants and for English [ɫ]. For Spanish [l], Early Bilinguals trended towards better accuracy than Late Bilinguals (β=0.533, posterior SD=0.337, *p*=0.06), and there was no difference between the Monolinguals and either bilingual group (vs. Early: β=0.120, posterior SD=0.315, *p*=0.36; vs. Late: β=0.413, posterior SD=0.266, *p*=0.12). Monolinguals trended towards more accurate responses than Early Bilinguals for Spanish [u] (β=0.567, posterior SD=0.290, *p*<0.10), and there was no difference in accuracy between Late Bilinguals and the other groups (vs. Monolinguals: β=0.110, posterior SD=0.259, *p*=0.38; vs. Early: β=0.457; posterior SD=0.320, *p*=0.10). Late Bilinguals categorized stimuli with English [ɫ] more accurately than Monolingual listeners (β=0.675, posterior SD=0.251, *p*<0.05), and the Early Bilinguals did not significantly differ from either other group (vs. Monolinguals: β=0.281, posterior SD=0.285, *p*=0.18; vs. Late: β=0.394, posterior SD=0.317). There were no differences in accuracy among the groups for English [ʉ] (Monolinguals vs. Early: β=0.096, posterior SD=0.279, *p*=0.39; Monolinguals vs. Late: β=0.104, posterior SD=0.242, *p*=0.38; Early vs. Late: β=0.200, posterior SD=0.304, *p*=0.29). To summarize, the three groups were more accurate with Spanish stimuli and with /l/ variants, except for the Late Bilinguals who categorized Spanish [l] as well as Spanish [u], and the only significant interaction between groups was for English [ɫ], for which the Late Bilinguals were significantly more accurate than the Monolinguals.

The results of the accuracy analyses for the phonemic and phonetic cues are summarized in Tables 4.4 and 4.5. Table 4.4 summarizes how Spanish and English phonemic (A) and phonetic (B) stimuli were categorized by each listener group, and Table 4.5 summarizes how the listener groups compared across the phonemic (A) and phonetic (B) segments. The “=” is used to illustrate differences that were not significant, and the “>” and “<” indicate significant differences. The “»” and “«” represent differences that approached significance.

Table 4.4: Summary of accuracy results for target segment comparisons in Experiment 1.

(A) Phonemic Cues

<table>
<tr><th></th><th>Target Segments</th></tr>
<tr><td>Monolinguals</td><td rowspan="2">Spanish /r/ > English /ɹ/ > English /θ/</td></tr>
<tr><td>Early Bilinguals</td></tr>
<tr><td>Late Bilinguals</td><td>Spanish /r/ = English /ɹ/ > English /θ/</td></tr>
</table>

(B) Phonetic Cues

<table>
<tr><th></th><th>Cross-language</th><th>Cross-segment</th></tr>
<tr><td>Monolinguals</td><td rowspan="3">Spanish [l] > English [ɫ]
Spanish [u] > English [ʉ]</td><td rowspan="2">Spanish [l] > Spanish [u]
English [ɫ] > English [ʉ]</td></tr>
<tr><td>Early Bilinguals</td></tr>
<tr><td>Late Bilinguals</td><td>Spanish [l] = Spanish [u]
English [ɫ] > English [ʉ]</td></tr>
</table>

Table 4.5: Summary of accuracy results for listener group comparisons in Experiment 1.

(A) Phonemic Cues

<table>
<tr><th></th><th>Listener Groups</th></tr>
<tr><td>Spanish /r/</td><td>Monolinguals = Early = Late</td></tr>
<tr><td>English /ɹ/</td><td rowspan="2">Monolinguals = Early < Late</td></tr>
<tr><td>English /θ/</td></tr>
</table>

(B) Phonetic Cues

	Listener Groups
Spanish [l]	Early » Late Monolinguals = Early, Monolinguals = Late
English [ɫ]	Early < Late Monolinguals = Early, Monolinguals = Late
Spanish [u]	Monolinguals » Early Monolinguals = Late, Early = Late
Spanish [ʉ]	Monolinguals = Early = Late

4.3.2 Reaction time analysis

Mean RTs and standard deviations to each stimulus type for each listener group are reported in Table 4.6. Log-transformed RTs were analyzed using a Bayesian mixed effects linear regression model with listener language group (three levels: Monolingual, Early Bilingual, Late Bilingual), target segment (seven levels: /r, ɹ, θ/ and [l, u, ɫ, ʉ]), and accuracy (correct, incorrect) as fixed effects and participant and stimulus word as random intercepts. These models were also fit via a Markov Chain Monte Carlo procedure using STAN, as described above. A model with an interaction between language group and target segment provided an improved fit over a model without this interaction term. The fit of the model was improved further by adding accuracy as a main effect, and again by adding a three-way interaction with accuracy. See Table 4.7 for the summary of this model. The reference group, reflected in the model intercept, represents the log RT of inaccurate responses by Monolinguals categorizing stimuli with the Spanish phonemic cue /r/. The fitted log RT for each target segment and listener language group are plotted in Figures 4.3 and 4.4 for phonemic cues and in Figures 4.5 and 4.6 for phonetic cues. For clarity of presentation given the number of conditions (3 listener groups x 7 segments x 2 accuracy levels), Figures 4.3 and 4.5 present only the accurate responses and those conditions with a significant interaction with accuracy are included in Figures 4.4 and 4.6. The error bars represent the 95% Bayesian credible intervals. The differences among accurate trials across segments and listener groups are reported in the results below, and interactions with accuracy are described afterwards.

Table 4.6: Mean RT (in milliseconds) in correct and incorrect trials for each listener group and stimulus type in Experiment 1.

(A) Correct trials

		Monolinguals	Early Bilinguals	Late Bilinguals
Phonemic Cues	Spanish /r/	538.0 (591.8)	530.4 (545.1)	639.8 (675.8)
	English /ɹ/	458.4 (497.8)	519.7 (601.3)	570.7 (509.5)
	English /θ/	647.3 (682.4)	763.0 (837.7)	761.8 (744.6)
Phonetic Cues	Spanish [l]	589.3 (663.0)	569.8 (650.5)	737.2 (744.8)
	English [ɫ]	537.0 (698.5)	629.1 (785.2)	694.5 (820.5)
	Spanish [u]	601.7 (607.9)	721.0 (766.7)	817.8 (837.1)
	English [ʉ]	653.6 (785.0)	818.4 (876.7)	863.7 (744.7)

(B) Incorrect trials

		Monolinguals	Early Bilinguals	Late Bilinguals
Phonemic Cues	Spanish /r/	967.5 (1371.5)	799.0 (900.0)	849.0 (837.8)
	English /ɹ/	1008.8 (1103.2)	902.9 (1432.9)	1219.6 (1160.8)
	English /θ/	832.2 (831.2)	825.8 (1043.2)	1103.9 (1209.1)
Phonetic Cues	Spanish [l]	1087.5 (1497.2)	1119.0 (1310.8)	1185.3 (1102.1)
	English [ɫ]	1009.2 (1170.7)	785.2 (959.5)	982.3 (1140.5)
	Spanish [u]	797.6 (930.6)	901.4 (882.1)	1146.8 (861.8)
	English [ʉ]	830.1 (827.9)	751.7 (910.4)	948.2 (973.8)

Table 4.7: Summary of mixed effects linear regression model fitting reaction time in Experiment 1.

Predictor	*Mean*	*Posterior SD*	*95% CI*	*p value*
Intercept (Monolingual, Spanish /r/, incorrect)	6.226	0.105	(6.021, 6.438)	<0.0001
English /ɹ/	0.009	0.103	(-0.199, 0.207)	NS
English /θ/	0.051	0.095	(-0.138, 0.239)	NS
Spanish [l]	0.155	0.109	(-0.068, 0.363)	NS
English [ɫ]	0.040	0.097	(-0.160, 0.227)	NS
Spanish [u]	0.036	0.102	(-0.168, 0.232)	NS
English [ʉ]	0.023	0.097	(-0.164, 0.212)	NS
Early Bilinguals	0.123	0.181	(-0.234, 0.466)	NS
Late Bilinguals	0.149	0.180	(-0.220, 0.502)	<0.10
Response correct	-0.249	0.078	(-0.403, -0.101)	<0.001
English /ɹ/ * Early	-0.034	0.180	(-0.374, 0.321)	NS
English /θ/ * Early	-0.151	0.159	(-0.450, 0.163)	NS
Spanish [l] * Early	0.056	0.186	(-0.305, 0.425)	<0.01
English [ɫ] * Early	-0.311	0.162	(-0.623, 0.004)	<0.10
Spanish [u] * Early	0.022	0.166	(-0.304, 0.353)	<0.05
English [ʉ] * Early	-0.102	0.158	(-0.409, 0.210)	NS
English /ɹ/ * Late	0.233	0.191	(-0.135, 0.611)	<0.0001
English /θ/ * Late	0.216	0.161	(-0.095, 0.548)	<0.0001
Spanish [l] * Late	0.108	0.179	(-0.233, 0.474)	<0.0001
English [ɫ] * Late	-0.114	0.165	(-0.429, 0.219)	NS
Spanish [u] * Late	0.268	0.172	(-0.069, 0.610)	<0.0001
English [ʉ] * Late	0.110	0.160	(-0.202, 0.430)	<0.01
English /ɹ/ * Correct	-0.177	0.091	(-0.355, 0.003)	<0.0001
English /θ/ * Correct	0.183	0.084	(0.022, 0.350)	NS
Spanish [l] * Correct	-0.088	0.097	(-0.277, 0.102)	<0.01
English [ɫ] * Correct	-0.104	0.085	(-0.264, 0.067)	<0.0001
Spanish [u] * Correct	0.085	0.090	(-0.092, 0.263)	<0.05
English [ʉ] * Correct	0.173	0.084	(0.012, 0.339)	NS
Early * Correct	-0.114	0.181	(-0.234, 0.466)	NS
Late * Correct	0.098	0.180	(-0.220, 0.502)	<0.05
English /ɹ/ * Early * Correct	0.068	0.184	(-0.305, 0.420)	NS
English /θ/ * Early * Correct	0.196	0.163	(-0.130, 0.507)	<0.01
Spanish [l] * Early * Correct	-0.089	0.189	(-0.465, 0.276)	NS
English [ɫ] * Early * Correct	0.384	0.167	(0.054, 0.703)	NS
Spanish [u] * Early * Correct	0.116	0.172	(-0.228, 0.446)	<0.01

Table 4.7, cont.

English [ʉ] * Early * Correct	0.234	0.164	(-0.088, 0.551)	<0.01
English /ɹ/ * Late * Correct	-0.197	0.195	(-0.578, 0.179)	NS
English /θ/ * Late * Correct	-0.313	0.166	(-0.646, 0.005)	<0.0001
Spanish [l] * Late * Correct	-0.061	0.182	(-0.430, 0.291)	<0.001
English [ɫ] * Late * Correct	0.145	0.170	(-0.200, 0.475)	<0.05
Spanish [u] * Late * Correct	-0.179	0.176	(-0.531, 0.169)	<0.0001
English [ʉ] * Late * Correct	-0.052	0.166	(-0.383, 0.265)	<0.0001

Random effects	*Variance*
Listener	0.365
Stimulus Word	0.102

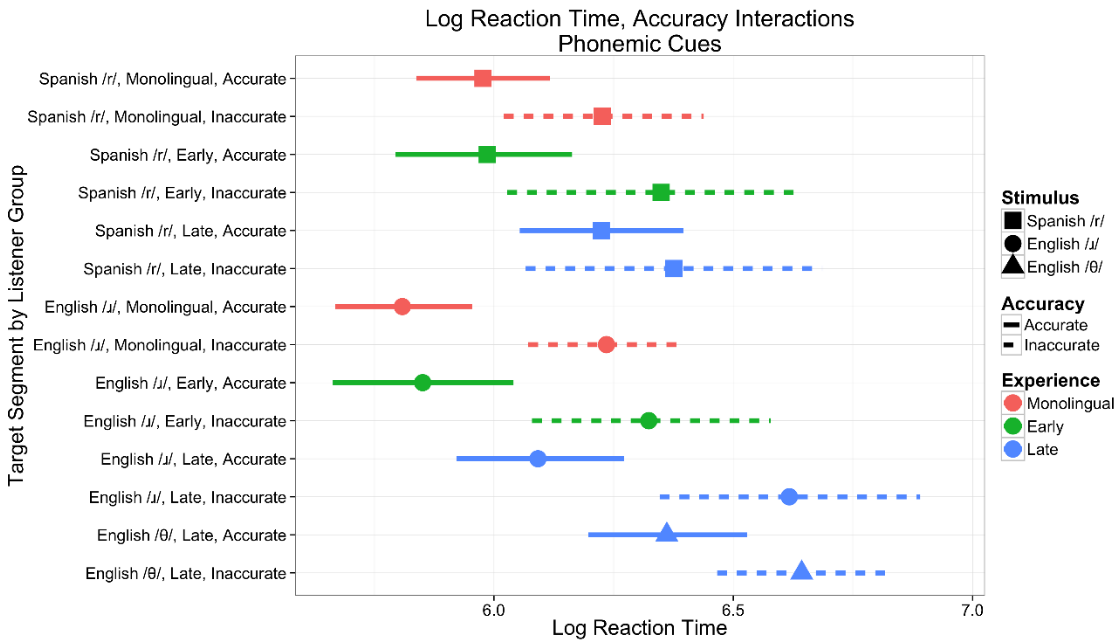
Phonemic cues

The difference in log RT between Spanish /r/ and English /ɹ/ was significant only for Monolinguals ($\beta=0.168$, posterior SD=0.061, $p<0.05$), with the English /ɹ/ responded to faster than the Spanish /r/. A difference was not found for the other groups (Early: $\beta=0.135$, posterior SD=0.064, $p=0.14$; Late: $\beta=0.132$, posterior SD=0.063, $p=0.14$). Monolinguals and Early Bilinguals responded significantly faster to Spanish /r/ than to English /θ/ (Monolinguals: $\beta=0.234$, posterior SD=0.058, $p<0.001$; Early: $\beta=0.280$, posterior SD=0.063, $p<0.05$), but there was no difference in RT for the Late Bilinguals ($\beta=0.137$, posterior SD=0.059, $p=0.13$). All three groups responded more quickly to English /ɹ/ than English /θ/ (Monolinguals: $\beta=0.403$, posterior SD=0.058, $p<0.0001$; Early: $\beta=0.415$, posterior SD=0.064, $p<0.001$; Late: $\beta=0.269$, posterior SD=0.061, $p<0.05$).

Figure 4.3: Log reaction time for phonemic cues in accurate trials in Experiment 1.

Monolinguals responded faster to Spanish /r/ than either Early (β=0.258, posterior SD=0.135, $p<0.001$) or Late Bilinguals (β=0.497, posterior SD=0.129, $p<0.0001$), and Early Bilinguals responded faster than Late Bilinguals (β=0.238, posterior SD=0.116, $p<0.05$). For English /ɹ/, Monolinguals responded faster than Early (β=0.292, posterior SD=0.132, $p<0.0001$) and Late Bilinguals (β=0.533, posterior SD=0.127, $p<0.0001$), and Early Bilinguals were faster than Late Bilinguals (β=0.241, posterior SD=0.116, $p<0.05$). For stimuli with English /θ/, Monolinguals were again significantly faster than Early (β=0.303, posterior SD=0.132, $p<0.0001$) and Late Bilinguals (β=0.399, posterior SD=0.128, $p<0.0001$), but for this segment there was no difference between the bilingual groups (β=0.095, posterior SD=0.116, $p=0.22$).

Figure 4.4: Log reaction time for phonemic cues in conditions with a significant interaction with accuracy in Experiment 1.



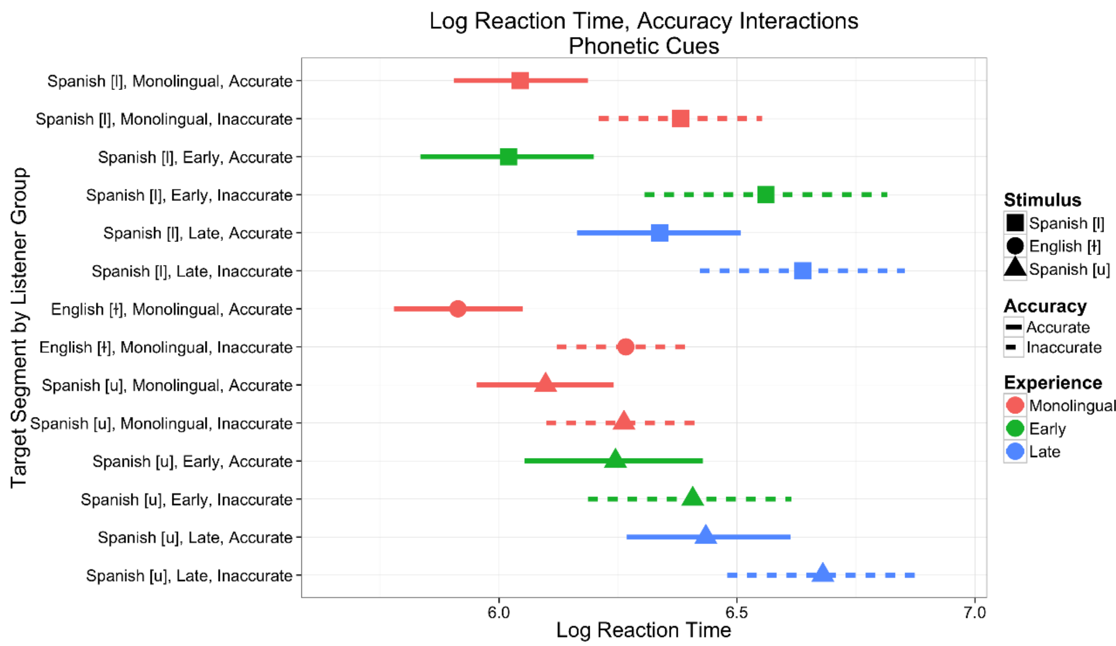
The interaction with accuracy showed significantly faster responses in correct trials than in incorrect trials for Monolinguals' responses to Spanish /r/ (β=0.249, posterior SD=0.078, *p*<0.001), and there were no accuracy-based difference for the bilingual groups (Early: β=0.114, posterior SD=0.154, *p*=0.18; Late: β=0.098, posterior SD=0.157, *p*=0.26). All three groups were also faster in correct than in incorrect trials in response to English /ɹ/ (Monolinguals: β=0.426, posterior SD=0.047, *p*<0.0001; Early: β=0.223, posterior SD=0.119, *p*<0.05; Late: β=0.276, posterior SD=0.132, *p*<0.05). For English /θ/, the interaction with accuracy showed that Early Bilinguals responded significantly faster on incorrect trials than on correct trials (β=0.265, posterior SD=0.91, *p*<0.01), with no difference by accuracy for the other groups (Monolinguals: β=0.066, posterior SD=0.030, *p*=0.18; Late: β=0.032, posterior SD=0.090, *p*=0.39). Overall, Monolinguals and Early Bilinguals responded fastest to English /ɹ/ stimuli, followed by Spanish /r/ and then /θ/, while Late Bilinguals' RTs were not different across the sounds. Monolinguals were fastest for all three segments, followed by the Early Bilinguals and then the Late Bilinguals, and in most accuracy interactions correct responses were faster than incorrect ones.

Phonetic cues

The difference in RT for Spanish [l] and English [ɫ] reached significance for Monolinguals (β=0.130, posterior SD=0.057, p<0.05), with their responses to the English variant faster than to the Spanish variant, but was not significant for the bilingual groups (Early: β=0.025, posterior SD=0.062, p=0.42; Late: β=0.146, posterior SD=0.060, p=0.12). No group showed a significant difference between the Spanish and English /u/ variants (Monolinguals: β=0.075, posterior SD=0.055, p=0.15; Early: β=0.068, posterior SD=0.061, p=0.29; Late: β=0.044, posterior SD=0.059, p=0.36). In comparing RTs on trials with Spanish [l] and [u], the Early Bilinguals responded significantly faster to [l] than to [u] (β=0.225, posterior SD=0.060, p<0.05); there was no difference in RT for the Monolinguals (β=0.054, posterior SD=0.055, p=0.22) and Late Bilinguals (β=0.097, posterior SD=0.058, p=0.21). For the English [ɫ] and [ʉ], all three groups responded significantly faster to trials with [ɫ] than to trials with [ʉ] (Monolinguals: β=0.259, posterior SD=0.055, p<0.0001; Early: β=0.318, posterior SD=0.062, p<0.01; Late: β=0.286, posterior SD=0.059, p<0.01).

Figure 4.5: Log reaction time for phonetic cues in accurate trials in Experiment 1.

Figure 4.6: Log reaction time for phonetic cues in conditions with a significant interaction with accuracy in Experiment 1.



There were also differences in RTs among the listener groups within each sound category. For Spanish [l], Monolinguals responded more quickly than the bilingual groups (Early: β=0.225, posterior SD=0.132, p<0.001; Late: β=0.543, posterior SD=0.127, p<0.0001), and the Early Bilinguals were faster than the Late Bilinguals (β=0.318, posterior SD=0.116, p<0.01). For the remaining segments, Monolinguals were faster than both bilingual groups, and Early Bilinguals trended towards responding significantly faster than Late Bilinguals. For English [ɫ], Monolinguals were faster than Early (β=0.330, posterior SD=0.132, p<0.0001) and Late (β=0.528, posterior SD=0.127, p<0.0001) Bilinguals, and Early Bilinguals trended towards responding faster than Late Bilinguals (β=0.198, posterior SD=0.116, p<0.10). For Spanish [u], Monolinguals were faster than Early (β=0.396, posterior SD=0.132, p<0.0001) and Late (β=0.586, posterior SD=0.127, p<0.0001) Bilinguals, and Early Bilinguals trended towards faster responses than Late

Bilinguals (β=0.190, posterior SD=0.117, $p<0.10$). For English [ʉ], Monolinguals responded significantly faster than Early Bilinguals (β=0.389, posterior SD=0.132, $p<0.0001$) and Late Bilinguals (β=0.555, posterior SD=0.128, $p<0.0001$), and the responses of Early Bilinguals were marginally faster than those of Late Bilinguals (β=0.166, posterior SD=0.118, $p<0.10$).

There was a three-way interaction with accuracy for Monolinguals and Early Bilinguals in the Spanish [l] condition, with correct responses having significantly faster RTs than incorrect responses (Monolinguals: β=0.337, posterior SD=0.055, $p<0.0001$; Early: β=0.291; posterior SD=0.124; $p<0.01$); there was no difference between accurate and inaccurate trials in this condition for Late Bilinguals (β=0.051, posterior SD=0.105, $p=0.33$). For the English [ɫ], Monolinguals responded significantly faster in correct than in incorrect trials (β=0.353, posterior SD=0.035, $p<0.0001$), and this difference in both bilingual groups approached significance, but was in the opposite direction, with incorrect RTs trending faster than correct (Early: β=0.166, posterior SD=0.096, $p=0.05$; Late: β=0.140, posterior SD=0.094, $p=0.07$). For the Spanish [u], Monolinguals were faster in correct than in incorrect trials (β=0.165, posterior SD=0.045, $p<0.05$), and there were no accuracy-based differences for the bilingual groups (Early: β=0.087, posterior SD=0.099, $p=0.21$; Late: β=0.003, posterior SD=0.100, $p=0.48$). For English [ʉ], there was no significant difference between accurate and inaccurate trials for Monolinguals (β=0.076, posterior SD=0.031, $p=0.13$), but Early (β=0.293, posterior SD=0.090, $p<0.01$) and Late (β=0.219, posterior SD=0.088, $p<0.01$) Bilinguals responded faster in incorrect trials than in correct trials. In sum, there were fewer significant differences between sounds for phonetic cues, although RTs for English [ɫ] tended to be shorter than for other sounds. Monolinguals were faster than the bilingual groups for all four segments, and the Early

Bilinguals were faster than the Late Bilinguals for words with Spanish [l]. For the English variants, there was also a tendency for the bilingual groups to respond faster on inaccurate trials than on accurate ones.

The results of the RT analyses for the phonemic and phonetic cues are summarized in Tables 4.8, 4.9, and 4.10. In Table 4.8, the RTs for each segment are compared for each listener group, and in Table 4.9 the responses of the listener groups are compared across the target segments. Table 4.10 summarizes the differences in RT in accurate and inaccurate trials. As in the accuracy summaries, the “=” designates differences that did not reach significance, the “>” and “<” indicate significant differences, and “»” and “«” represent differences that approached significance.

Table 4.8: Summary of RT results for target segment comparisons in Experiment 1.

(A) Phonemic Cues

	Target Segments
Monolinguals	English /ɹ/ < Spanish /r/ < English /θ/
Early Bilinguals	English /ɹ/ = Spanish /r/ < English /θ/
Late Bilinguals	English /ɹ/ < English /θ/ Spanish /r/ = English /ɹ/, Spanish /r/ English /θ/

(B) Phonetic Cues

<table>
<tr><th></th><th>Cross-language</th><th>Cross-segment</th></tr>
<tr><td>Monolinguals</td><td>Spanish [l] > English [ɫ]
Spanish [u] = English [ʉ]</td><td>Spanish [l] = Spanish [u]
English [ɫ] < English [ʉ]</td></tr>
<tr><td>Early Bilinguals</td><td rowspan="2">Spanish [l] = English [ɫ]
Spanish [u] = English [ʉ]</td><td>Spanish [l] < Spanish [u]
English [ɫ] < English [ʉ]</td></tr>
<tr><td>Late Bilinguals</td><td>Spanish [l] = Spanish [u]
English [ɫ] < English [ʉ]</td></tr>
</table>

Table 4.9: Summary of RT results for listener group comparisons in Experiment 1.

(A) Phonemic Cues

	Listener Groups
Spanish /r/	Monolinguals < Early < Late
English /ɹ/	
English /θ/	Monolinguals < Early, Monolinguals < Late Early = Late

(B) Phonetic Cues

	Listener Groups
Spanish [l]	Monolinguals < Early < Late
English [ɫ]	Early « Late Monolinguals < Early, Monolinguals < Late
Spanish [u]	
Spanish [ʉ]	

Table 4.10: Summary of RT results for accuracy interactions in Experiment 1.

(A) Phonemic Cues

		Trial Accuracy
Spanish /r/	Monolinguals	Correct < Incorrect
	Early Bilinguals	Correct = Incorrect
	Late Bilinguals	
English /ɹ/	Monolinguals	Correct < Incorrect
	Early Bilinguals	
	Late Bilinguals	
English /θ/	Monolinguals	Correct = Incorrect
	Early Bilinguals	Correct > Incorrect
	Late Bilinguals	Correct = Incorrect

Table 4.10, cont.

(B) Phonetic Cues

		Trial Accuracy
Spanish [l]	Monolinguals	Correct < Incorrect
	Early Bilinguals	
	Late Bilinguals	Correct = Incorrect
English [ɫ]	Monolinguals	Correct < Incorrect
	Early Bilinguals	Correct > Incorrect
	Late Bilinguals	
Spanish [u]	Monolinguals	Correct < Incorrect
	Early Bilinguals	Correct = Incorrect
	Late Bilinguals	
English [ʉ]	Monolinguals	Correct = Incorrect
	Early Bilinguals	Correct > Incorrect
	Late Bilinguals	

4.3.3 Variation across nonce words

The categorization patterns described above for each target segment include the decisions made for eight nonce words per target segment. While it may appear that listeners were more or less accurate in categorizing a given segment, it may be the case that language decisions were made differently depending on where the target segment occurred in the word, e.g., word-initial or word-internal (cf. the categorization of *taco* in Flege & Munro, 1994). There may also have effects of the filler vowels /i, a/ used in the first syllable, or some nonce words may have been outliers that represented a target segment particularly well or particularly poorly, or at least differently from the other stimuli for that segment. In an attempt to better understand listeners' categorization patterns, this section describes the accuracy of language decisions made for each word in a stimulus group across the three listener groups. Means and trends from the raw data are discussed.

4.3.3.1 Phonemic stimuli

All listener groups categorized nonce words with Spanish /r/ and English /ɹ/ very successfully. As can be seen in Figures 4.7 (Spanish /r/) and 4.8 (English /ɹ/), the groups performed nearly at ceiling across all words in these categories.

From Figure 4.9 it is clear that responses to English /θ/ were less accurate overall and there is considerable variation among the eight stimuli. At the extremes are *sathuh* (/saθə/) and *thichuh* (/θitʃə/). The bilingual groups categorized *sathuh* very accurately (Early: 92.8%; Late: 93.7%), while *thichuh* was exceptionally difficult for all three groups. For this word, Monolinguals were correct in 44.4% of trials and Late Bilinguals in 37.4% of trials, but the Early Bilinguals responded accurately only 13.2% of the time. The two words with initial /θ/ were especially difficult for Monolingual and Early Bilingual listeners, and while the Late Bilinguals also struggled with *thichuh*, their accuracy for *thisuh* was not particularly low (78.4%).

Figure 4.7: Mean accuracy for each Spanish /r/ nonce word, by listener group, in Experiment 1.

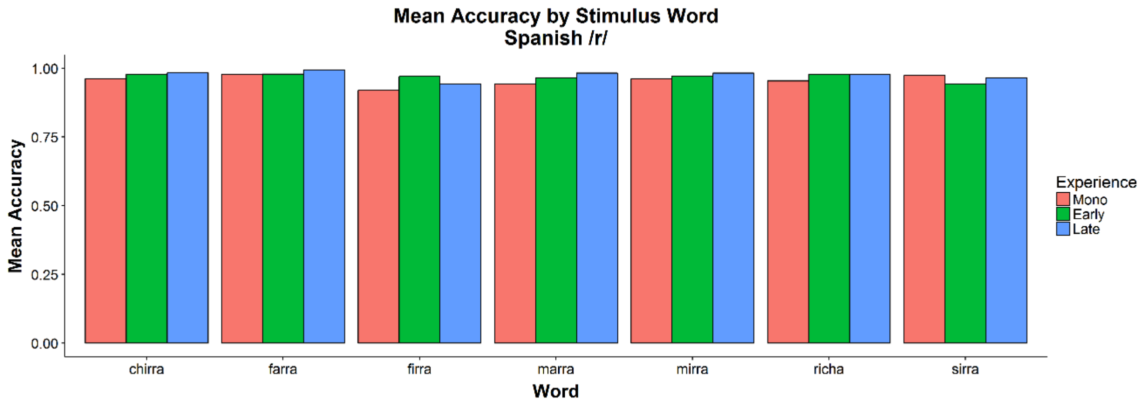


Figure 4.8: Mean accuracy for each English /ɹ/ nonce word, by listener group, in Experiment 1.

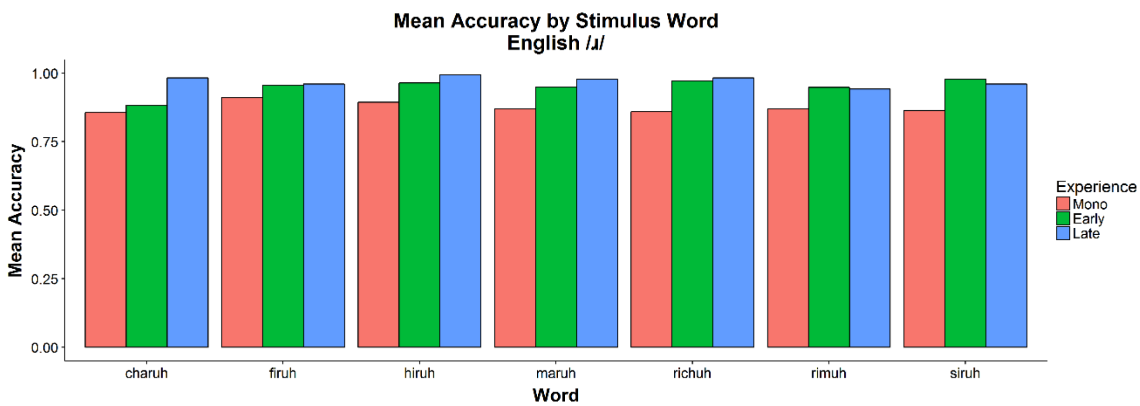
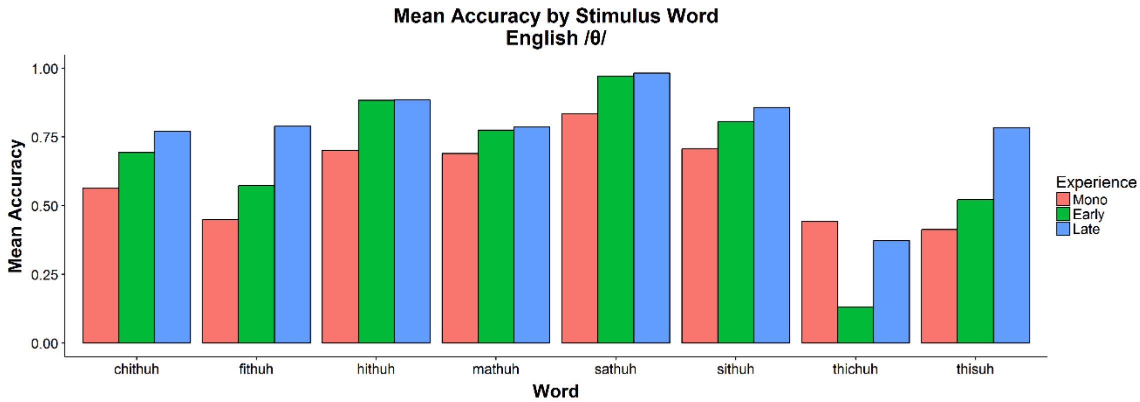


Figure 4.9: Mean accuracy for each English /θ/ nonce word, by listener group, in Experiment 1.



4.3.3.2 Phonetic stimuli

The nonce words with Spanish [l] were categorized accurately and consistently by the three listener groups, and there was little variability among the eight words in this category; see Figure 4.10. Listener performance in response to English [ɫ] was much more variable. The accuracy plot in Figure 4.11 suggests that listeners had much more difficulty with [ɫ]-initial words than with [ɫ]-medial words. *Lichuh* was especially difficult for listeners, with Monolinguals responding accurately in 48.1% of trials, Early Bilinguals in 21.4% of trials, and Late Bilinguals in 31.3% of trials. While the Monolinguals were significantly less accurate than the Late Bilinguals across English [ɫ] stimuli, this difficult word, *lichuh,* is the only English [ɫ] word for which the Monolinguals outperformed the Late Bilinguals.

Figure 4.10: Mean accuracy for each Spanish [l] nonce word, by listener group, in Experiment 1.

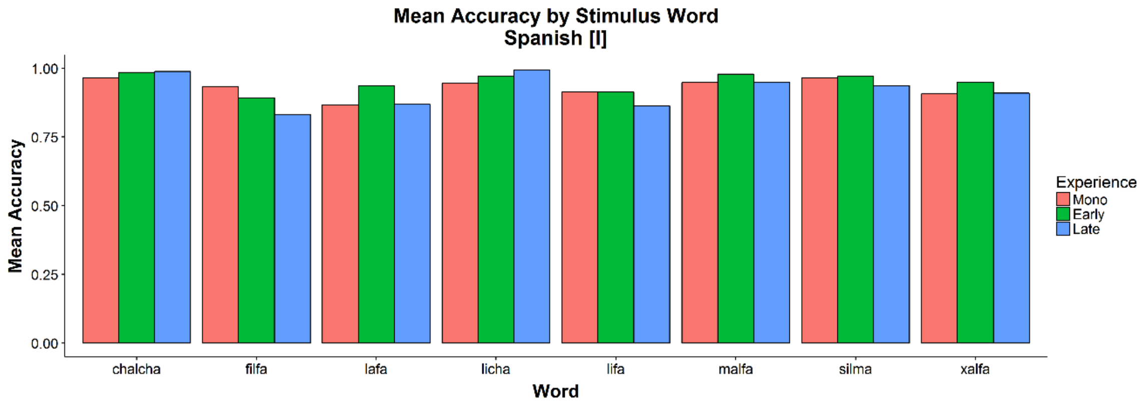
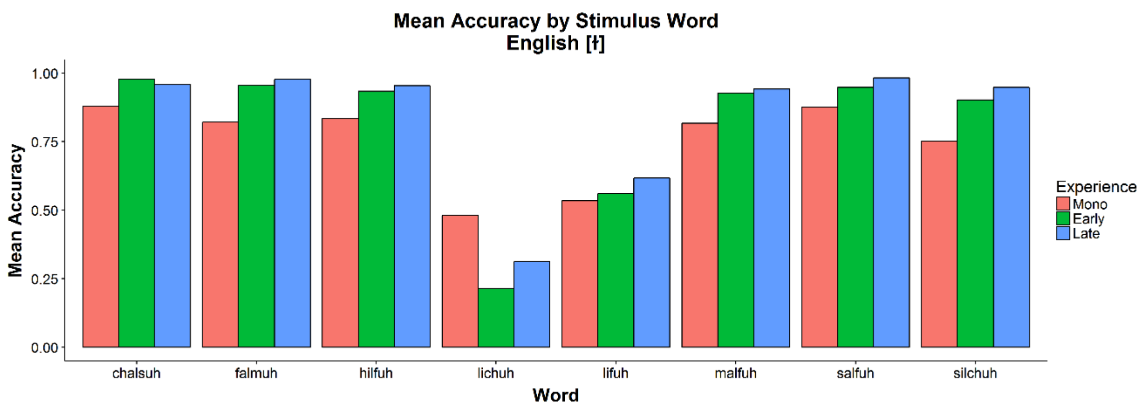


Figure 4.11: Mean accuracy for each English [ɫ] nonce word, by listener group, in Experiment 1.



Categorization accuracy was consistent across the eight stimuli of Spanish [u]; see Figure 4.12. All listener groups struggled with English [ʉ]; see Figure 4.13. Unlike the pattern observed for the difficult phoneme English /θ/, here there was only limited variation across the eight words. Notably, however, is the stimulus *chuchuh*, for which all three groups – especially the bilingual groups – registered low accuracy. For this word, Monolinguals were correct in about half of trials (49.0%), Early Bilinguals were correct 31.6% of the time, and Late Bilinguals 32.0% of the time.

Figure 4.12: Mean accuracy for each Spanish [u] nonce word, by listener group, in Experiment 1.

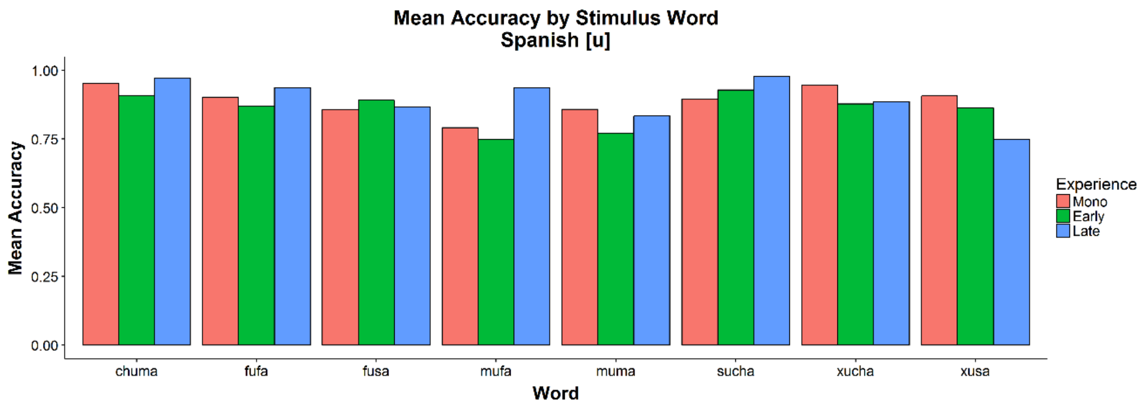
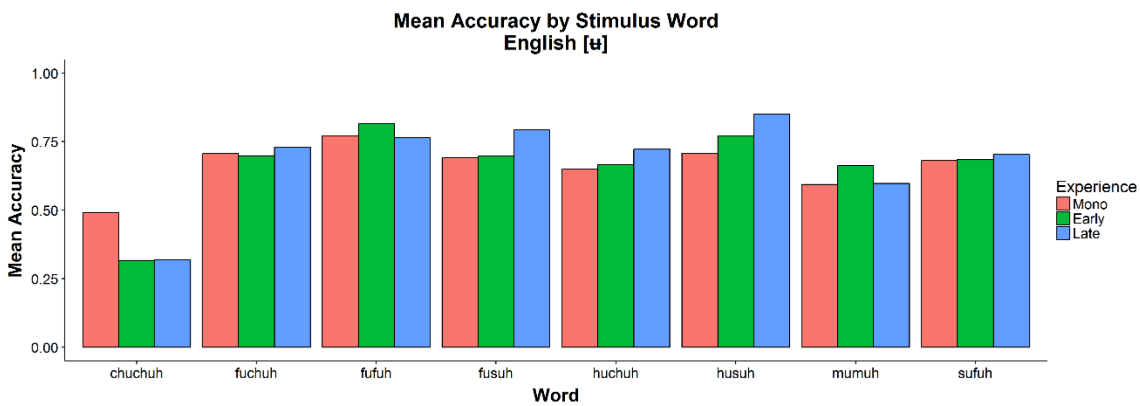


Figure 4.13: Mean accuracy for each English [ʉ] nonce word, by listener group, in Experiment 1.



The analyses by stimulus word indicate that listeners in all groups categorized the Spanish stimuli very consistently. Responses to the English segments were much more variable for all groups: for English /θ/, English [ɫ], and English [ʉ], some nonce words were categorized relatively accurately, and listeners found other nonce words extremely difficult to categorize. For words with English /θ/ and English [ɫ], the most challenging stimuli included these sounds word-initially, and nonce words with these segments word-internally were categorized more accurately. This pattern may suggest that perceptually difficult cues are more easily perceived when they do not occur in the word-initial position, and this was also the case for syllable identification in noise in Cutler et al. (2004). However, for the difficult English [ʉ], which only occurred as the nucleus of the first syllable, accuracy was consistently low across words. A more systematic study of how salience varies with position in a word may be warranted to understand the effects of context on these perceptually difficult segments.

4.4 DISCUSSION

The current study tested the sensitivity of monolingual and bilingual adults to language-specific segments in a nonce-word categorization task to determine which sounds listeners are most sensitive to and how language experience influences listeners' sensitivity. Overall, there was considerable variation in accuracy within the two stimulus types, phonemic and phonetic. Listeners very consistently categorized Spanish /r/ accurately but were much less likely to succeed with the English phoneme /θ/ and the English phonetic cue [ʉ]. The Spanish phonetic variants were categorized more accurately than their English counterparts, and listeners were more successful with the /l/ variants than with the /u/ variants. The categorization decisions of the Monolinguals, Early Bilinguals, and Late Bilinguals were very similar, although Late Bilinguals categorized the English phonemic cues and the English phonetic variant [ɫ] more accurately than the other groups. The Early Bilinguals looked much more like the Monolinguals than they resembled the Late Bilinguals with respect to accuracy, despite knowing both English and Spanish. The variable results across phonemic and phonetic cues – modeled on Flege's new and similar L2 phones – may not be the most useful way of characterizing which segments listeners are most sensitive to in a cross-language speech perception task using word-length stimuli. The results of Experiment 1 indicate that segments from the same class (i.e., phonemic or phonetic) can vary widely in how closely they are associated with their respective language, and at least part of this variation may be related to inherent differences in salience. The theorized relationships between native and non-native categories presented in earlier cross-language speech perception models may be too general since they fail to account for potentially universal differences in salience among sounds that may be unrelated to a listener's L1 or L2. This study also provides evidence that there may be fundamental differences in early and late bilinguals' access to language-cue associations. Reaction times to each stimulus revealed additional differences across listener groups: the Monolingual listeners responded faster than the Early Bilinguals, who in turn were faster than the Late Bilinguals. In this study, it was via response time differences, and not in their categorization accuracy, that Early Bilinguals could be distinguished from Monolinguals.

The consistency of categorization accuracy across the three listener groups suggests that language background was less important than cue salience in this task. Phonemic cues were mostly categorized more accurately than phonetic cues, supporting the parallel distinction made between new and similar phones in the SLM proposed by Flege (1987, 1995). In this model, second language learners create independent categories for sounds judged to be new to the L2 and not present in the L1, which facilitates the production and perception of such sounds. Phones that are recognized as similar to existing L1 segments are discriminated less well if no new category is established for them. The phonemes in the present task may be like the SLM's new phones, even for the Monolinguals who have not acquired Spanish, and as such they are more recognizable as language-specific sounds (Best, 1991), which led to more accurate categorization. The phonetic cues are parallel to the SLM's similar phones, sounds for which listeners recognize differences between L1 and L2 categories more slowly. Listeners categorized the phonetic cues similarly, with no differences between Monolinguals and Early Bilinguals, and variable differences between the bilingual groups – the Early listeners trended towards better accuracy of Spanish [l] than the Late Bilinguals, but the Late group outperformed the Early Bilinguals for English [ɫ]. In Best's (1991) perceptual assimilation model, these non-L1 sounds are mapped to categories in the L1, e.g. the Late Bilinguals' English [ɫ] to their Spanish [l]. This causes more competition in deciding between English or Spanish for stimulus language identity in the nonce-word categorization task. The variation in responses within the phonemic and phonetic categories, and across languages and listener groups, is explored below.

Phonemic cues: success with rhotics, trouble with English /θ/

The accuracy and RT analyses of the language-specific phonemic cues suggest that language background is less important than cue salience: listeners across the groups performed at ceiling in Spanish /r/ trials, were less accurate in response to English /ɹ/, and struggled to categorize stimuli with English /θ/. The Monolinguals and Early Bilinguals were both more sensitive to the association of /r/ with Spanish than of /ɹ/ with English, while the Late Bilinguals did not differ in their accuracy for these segments. This similarity between Monolingual and Early Bilinguals is striking because the Early Bilingual listeners, unlike the Monolinguals, speak Spanish and thus have a contrasting phonological system against which to judge the language specificity of English /ɹ/. The Early Bilinguals are also unable to make this association to the same degree as the Late Bilinguals, who also speak both languages. For the third phonemic cue, English /θ/, all groups performed worse than on the other phonemes, but the Late Bilinguals were significantly more accurate than the other groups. The Late Bilinguals' greater success than other groups with the English phonemes reflects an ability to contrast one language's phonological system with another, thus highlighting segments unique to each language, like the English /θ/. However, as was the case for English /ɹ/, it is not clear why Early Bilinguals are less sensitive to /θ/ than the Late Bilinguals. Differences among the listener groups will be considered below.

The difficulty listeners had categorizing nonce words with English /θ/ may be related to this segment being challenging more generally. Fricatives and interdentals in particular are acquired late by English-learning children (Clark, 2009; Dodd et al., 2003), and even native-English-speaking adults may confuse /θ/ for other English phones more often than confusions happen for other segments (Cutler et al., 2004). Therefore the difficulty categorizing English /θ/ may not be exclusively attributable to a low association of this segment with representations of English. However, despite general perception issues for /θ/, it is surprising that the Monolinguals – for whom this is a phoneme in their only language – were not more sensitive to this cue. Native English listeners in Cutler et al. (2004), for example, more accurately identified /θ/ than native Dutch listeners, despite the low perceptual salience of this sound. An examination of the differences in accuracy across the eight words in the English /θ/ category reveals variable performance by stimulus, with some words, like *sathuh*, being relatively easily categorized by all listeners, and /θ/-initial words being much more difficult. The contrast in performance between /θ/-initial and /θ/-medial stimuli further suggests differences in salience due to phonological context. Even so, the Late Bilinguals were still able to make use of this sound in their decisions; see, for example, *thisuh* which they categorized much more accurately than the other groups. This supports the possibility that salience alone cannot explain accuracy for English /θ/ words. The pattern observed in the present study likely reflects both difficulty perceiving English /θ/ for all groups, but also, importantly, a less clear association between /θ/ and English for the Monolingual and Early Bilingual groups. This weaker link contrasts with these listeners' stronger association of /r/ with Spanish (cf. linguistic "stereotypes" in Labov, 1971).

The Late Bilinguals' success with English /θ/ is despite the likelihood that many Late Bilinguals knew that /θ/ exists in Peninsular Spanish (cf. Sections 3.1 and 4.2.1). This awareness could be expected to cause confusion and thus fewer accurate responses for the Late Bilinguals, but instead this group proved to be the most reliable at judging nonce words with /θ/ as English. Part of the Late Bilinguals' accuracy pattern may be explained by their reaction times: Monolinguals performed significantly faster than Early Bilinguals (see Kovelman et al., 2008, for a similar pattern), and the Late Bilinguals responded more slowly than the Early Bilinguals. The Late Bilinguals' longer RTs compared to the other listener groups may indicate increased activation of English and Spanish representations during the decision-making process, or that their representations include more acoustic detail, and these features were accessed during their slower responses. It may also be that the language-specific representations for the Late Bilinguals are no different than for the faster Early Bilinguals, and that instead the Late Bilinguals spent more time deciding on the appropriate response to each stimulus, leading to better accuracy, due to non-linguistic factors, such as being somewhat older than the Early group and having an interest in volunteering for linguistic studies. It is unclear whether the performance of the Early Bilinguals would have improved with more time – whether they would have been able to better access and compare the representations of each language – or if the Late Bilinguals' accuracy would have worsened if they were forced to respond faster. Either case would indicate comparable language-specific representations, but the current results suggest there may be real differences in the representations of the bilingual groups.

Phonetic cues: variability among segments, especially for English sounds

The categorization of phonetic cues also indicates that sensitivities to the selected sounds were largely the same across listener groups. Listeners categorized Spanish phonetic cues better than English ones, and the /l/ variants better than the /u/ variants. While there may be support from related literature about why /θ/ is generally a challenging segment, it is unclear why English and /u/ variants should be more difficult than Spanish and /l/ phones. Surprisingly, the bilingual groups responded more quickly on inaccurate trials with English [ʉ] than on accurate trials. Here, unlike with the phonemic cues, there are also fewer distinctions between the bilingual groups with respect to their RTs. Whereas the Early Bilinguals' RTs were intermediate between Monolinguals' and Late Bilinguals' for language-specific phonemes, with the phonetic cues the Early Bilinguals were only significantly faster than the Late Bilinguals for the very salient Spanish [l] and these groups responded statistically the same for the other sounds. There were also few significant differences in accuracy across the groups, although the Late Bilinguals outperformed the Monolinguals for English [ɫ] words. The Early Bilinguals did not significantly differ from the Monolinguals for any of the phonetic segments.

The difficulty listeners from all backgrounds experienced in accurately categorizing English phonetic cues and both languages' variants of /u/ warrants further investigation. The English [ɫ] is more velarized, i.e. produced with the tongue further back in the oral cavity, than the Spanish [l], while the English [ʉ] is fronted compared to the Spanish [u]. The difference in success between English and Spanish phonetic cues is therefore unlikely to be due to some single acoustic property, e.g. if higher values of F2 are easier to perceive than lower F2, since the English variants differ in opposite directions from the Spanish cues. It may be that listeners hear more variation in English input than in

Spanish input between lighter or darker /l/ and between more or less fronted /u/ across dialects, speakers, and phonological contexts.

However, there are several reasons that variation in input alone is insufficient to explain the results. First, it would be surprising if monolingual English listeners were sensitive to the greater consistency of these segments being lighter and backed in Spanish, given their lack of exposure to Spanish. Second, if the variability present in English input motivated this discrepancy in accuracy, an entirely different categorization pattern should be expected. A light [l] or a backed [u] may be from either Spanish or English, since these light and backed variants exist in many dialects of English and are more likely in certain phonological contexts. In fact, there is evidence of this pattern in the analysis of the words comprising the English [ɫ] stimuli. The English category /l/ is darker than the Spanish category (Recasens, 2004, 2012), but this difference is greatest for /l/ in coda position. All three groups of listeners responded with lower accuracy to the English [ɫ]-initial words *lichuh* and *lifuh*, which may seem to have more Spanish-like features, than to the darker, coda [ɫ] stimuli. But this pattern is only observed in one direction: English stimuli may be categorized as Spanish, but Spanish stimuli are not as likely to be categorized as English. That is, the lighter Spanish [l] and backed Spanish [u] are consistently categorized as Spanish, even though these variants should be acceptable English productions in some contexts, e.g. [l] word-initially. While there was variability in accuracy across English [ɫ] words, categorization patterns across English [ʉ] words are largely consistent. This may suggest that the variability of English [ɫ] in natural input is greater or more salient to listeners, than variability in English [ʉ]. Another possible difference between the cues is the consonant versus vowel distinction. The consistency of (inaccurate) responses to English [ʉ] across nonce words may also suggest that listeners are less able to use the variation in vowels in language-decision tasks. This explanation is partly supported by Monolinguals and Early Bilinguals categorizing Spanish [u] less accurately than Spanish [l], but the difficulty of English cues more generally must still be accounted for.

Co-occurring acoustic cues

In order to understand how listeners use phonetic detail and language-specific phonological categories in context, word-length stimuli were created for the categorization task. While every effort was made to create nonce words that were equally plausible in both languages, except for the language-specific target segment, the naturally-produced stimuli used here inevitably carried additional cues indicative of language. The phonotactic restrictions of Spanish may have meant that the CVCV stimuli were simply more Spanish-like than English-like, even though this word structure is permitted in English. The Spanish-ness of these stimuli is supported by the reactions of participants in two pilot studies; in the first pilot, theoretically congruous stimuli that overlapped between English and Spanish in all segments, e.g. /tʃima/, were categorized as Spanish significantly more than English, and in the second pilot (cf. Section 4.2.2), listeners reported confusion about whether words were English or English-accented Spanish. In the present study, listeners from all three language backgrounds were able to overcome this potential bias towards Spanish for English /ɹ/ and [ɫ] words, which were successfully categorized as belonging to English. Therefore, some cues are unambiguously associated with English, even if the word structure is less common in the language than it is in Spanish. Furthermore, Monolinguals might not be expected to suffer from such a potential bias, since they do not have representations of Spanish phonotactics, but instead their categorization patterns were in line with the bilingual groups'.

The difficulties that persisted for English /θ/ and [ʉ] are especially interesting in light of another characteristic of the nonce-word stimuli used here: since the words were naturally produced, there were multiple phonetic cues to language in each token. As was mentioned in Section 3.2, the disyllabic nature of the nonce words meant that the unstressed vowel /a/ in the second syllable was reduced to [ə] in the English words; therefore, all the English nonce words contained both a language-specific target segment (e.g. /θ/) and the reduced vowel. Furthermore, the acoustic analyses of the /i/ and /a/ vowels in the first syllable of the nonce words indicate that there were also language-specific differences in the productions of these non-target segments, especially for /a/ (cf. Section 3.3). But again, despite these potential additional cues to language, listeners were not able to reliably categorize two of the English-specific segments, /θ/ and [ʉ]. This is again surprising because listeners were more accurate with the target segments /ɹ/ and [ɫ] which also occurred in combination with these additional vocalic cues. Given the more accurate performance of the Late Bilinguals than the other groups for English /θ/ it might be tempting to conclude that the Late Bilinguals were better able to capitalize on these supplementary language-specific cues than their peers. However, their accuracy did not significantly differ from the Monolinguals and Early Bilinguals in the English [ʉ] condition. If the Late Bilinguals were more sensitive to the English-ness of the nonce word filler vowels in the phonemic condition, where they outperformed their peers, it is unclear why they wouldn't have also been able to make use of the additional cue in the English [ʉ] words. Maybe the presence of three language-specific cues met some perceptual threshold, while the [ʉ] and schwa cues in the phonetic stimuli were insufficient to trigger an English response. Alternatively, it may be the case that listeners from all backgrounds were better able to use language-specific consonants in their decisions than vowels – whether the target segment English [ʉ] or the incidental filler vowel cues described here – even when the stimulus word contained multiple vocalic cues to language. Since these language-specific segments did not include any language-specific vowels as phonemic cues, further work is needed to determine whether listeners prioritized consonantal information or phonemic cues in their categorization decisions.

Bilinguals' mental representations

Similarities in the performance of the Monolinguals and Early Bilinguals indicates that these listeners do not differ in how sensitive they are to the English-ness and Spanish-ness of language-specific segments. A stronger version of such a proposal might be that these groups do not differ in the level of detail encoded in their language representations, and this despite the Early Bilinguals' having learned Spanish at home before English. This is not to say that the Early Bilinguals would not have shown evidence of Spanish proficiency in other tests, such as production or phoneme identification tasks. Similarly, the nonce-word categorization task may be unlike typical language situations that adult bilinguals find themselves in: whereas young children may regularly encounter new words that they must integrate into one language system or the other, or must learn which community of speakers would accept a new word as the conventional form, adults may be less practiced at associating new word forms with a particular language. What the results of Experiment 1 do suggest is that the ability of Early Bilinguals to generalize about the properties of their native languages and in particular to associate phonological properties with each language is not distinct from Monolinguals' awareness of language-specific properties.

This pattern sets these early Spanish-English bilinguals apart from the early Spanish-Catalan bilinguals in Sebastián-Gallés et al. (2005), whose sensitivity to Catalan-specific contrasts was purportedly compromised by their early exposure to Spanish. Rather, the similarity between responses from Monolinguals and Early Bilinguals in the present study supports considering the bilinguals' current exposure to English in addition to their early exposure to Spanish (cf. Amengual, 2014, 2015, and the language assessment used therein: Gertken et al., 2014). The role of ongoing language exposure as a factor beyond age of acquisition is also supported in the production and perception work by Flege (1991, 1997, 1999, 2004), so future work on the association of language and segmental cues should consider dominance and exposure to each language as other possible factors influencing sensitivity. Like the early bilinguals in Flege's studies, the early Spanish-English bilinguals in the present experiment acquired both languages as children and live and study immersed in their (chronological) L2, English. As a result, these listeners do not appear to have language-specific phonological awareness greater than the English Monolinguals. On the other hand, the Late Bilinguals may have increased sensitivity to language-specific phonological properties due to the circumstances of their bilingualism – having learned a language after puberty and only more recently immersing themselves in an L2 environment.

The bilingual listeners' language dominance and exposure were not directly assessed in this study, but the Early Bilingual group may share more commonalities with Monolinguals than with Late Bilinguals. These traits may help in understanding the similarities in Monolinguals' and Early Bilinguals' categorization decisions and potentially why the Late Bilinguals often outperformed the other groups. Previous work in cross-language speech perception has not presented evidence of an increase in sensitivity of late learners to L2 cues, although some literature on L2 production suggests that late learners may produce L2 contrasts with even greater separation than monolinguals, due to more crowded phonological space (Flege, 1995, 2007). In order to account for the difference in the sensitivities of Early and Late Bilinguals in the present study, additional traits of the bilingual groups must be compared. Most of the Early Bilinguals (11 of 18) learned English when they began kindergarten, and language instruction at this age is likely to be much less explicit than the middle and high school foreign-language classrooms in which the Late Bilinguals learned English. Even where there are parallels in L2 teaching at these ages, the experience of English language learning is much more recent for the Late Bilinguals than for the Early Bilinguals, and attending foreign language classes, practicing the language, and laboring to master the rules of and achieve proficiency in the L2 may lead the Late listeners to a greater metalinguistic awareness about properties of the language (Dąbrowska & Street, 2006), including increased sensitivity to language-segment associations. Indeed, some advantages have been reported for bilingual children in the domain of phonological awareness (Bruck & Genesee, 1995; but see also Bialystok, 2001, for alternate interpretations of findings), and phonological awareness is especially relevant in the study of literacy development (e.g. Anthony & Francis, 2005). However, the work on phonological and metalinguistic awareness in adults is limited to literacy and disorders (e.g. Pennington et al., 1990) and so does not consider how this sensitivity may affect cross-language speech perception. The current findings suggest that the metalinguistic awareness of listeners who acquired an L2 in early childhood may decline into adulthood. Over time and as English proficiency increases, young bilingual listeners may lose their initial phonological sensitivity and may later categorize segments no differently than Monolingual adults who acquired their only language in infancy. The Late Bilinguals may then have increased sensitivity to language-specific phonological properties due to the circumstances of their bilingualism through formal language training and not necessarily due to age of acquisition.

There is another explanation for the Late Bilinguals' better accuracy: they responded significantly more slowly than other groups to a number of the target segments and so their accuracy may be due to greater care or deliberation in their decision-making as demonstrated by their RTs. Research comparing monolinguals and bilinguals has found that differences in RTs are related to the kinds of linguistic information retrieved during a task. When participants access semantic information about category membership (e.g., whether an object is naturally occurring or human-made), there are no differences in response times between monolinguals and bilinguals, presumably because no linguistic information needs to be accessed (Gollan et al., 2005). However in naming tasks, which require the activation of phonological and lexical levels of representations, early bilinguals name pictures in their L2 more slowly than monolinguals (Gollan et al., 2005, 2011; Ransdell & Fischler, 1987), although this was not the case for bimodal ASL-English bilinguals naming in English (Emmorey et al., 2013). Gollan and colleagues (2005, 2011; Emmorey et al., 2013) characterize this delay in lexical access as one of "weaker links" and is described in the frequency-lag hypothesis: since bilinguals use each of their languages less than a monolingual uses their only language, each of a bilingual's lexical representations is less frequently accessed and the access itself is less practiced than a monolingual's representation. The English naming latency difference that was absent for the ASL-English bilinguals and English monolinguals is then the result of a difference between bimodal and unimodal bilinguals: ASL-English bilinguals can codemix and use both languages simultaneously (i.e., signing while they speak) and so using their non-dominant language, ASL, does not exclude the use of their dominant language, English. In this way, bimodal bilinguals have stronger links to dominant-language lexical representations than may be the case for bilinguals of two spoken languages, who retrieve one to the exclusion of the other when producing a word. The Early Bilinguals in the present study may then have responded with RTs intermediate between the Monolinguals and Late Bilinguals because the Early Bilinguals' access to Spanish representations is significantly weaker than these links are for the Late Bilinguals. While the tasks used by Gollan and colleagues were lexical decision or naming tasks, which necessarily involve accessing lexical representations, the stimuli in Experiment 1 were all nonce words, and so by definition had no representation in either language. Since there were differences in RTs between Monolinguals and Early Bilinguals, as there were for Gollan and colleagues' lexical decision and naming tasks but not for semantic categorization, it appears that the participants in Experiment 1 did indeed process the nonce words linguistically, despite their lack of semantic content. The difference between Early and Late Bilinguals' RTs in Experiment 1 thus supports the application of the frequency-lag hypothesis to another domain of linguistic representation – the links between language and language-specific phonological categories and pronunciations.

There are now two possible motivations for the differences in accuracy and RT across the three listener groups: circumstances and recency of language learning, and differences in the strengths of language-segment association due to language use over the lifespan. Both recency of learning and link strength can be envisioned as fluid characterizations that change over time; at an earlier stage of development – before English overtook Spanish as the Early Bilinguals' primary language – the Early Bilinguals might have looked more like the Late Bilinguals, in terms of both recency of language learning and strength of language-segment associations. This possibility could be probed by testing young listeners who will grow up to be the Early Bilinguals: children from Spanish-speaking families who have just begun learning English in school. Since there is no clear developmental reason motivating differences in performance across the segments (with the exception of English /θ/), Early Bilingual children who have recently been immersed in an English-speaking environment might be expected to perform like the adult Late Bilinguals and therefore be more sensitive to difficult sounds than monolingual English-speaking children.

Conclusion

The results of the nonce-word categorization task indicate many differences within the phonemic and phonetic stimulus categories modeled on the new and similar sounds described by Best (1991, 1995) and Flege (1987, 1995). In fact, categorization patterns were largely independent of listener language background, with all three groups exceling in and struggling with the same subset of segmental cues. These findings also show similarities across listeners who acquired English early, in parallel with the work of Flege et al. (1999) and Mack (1989) on early bilinguals' phoneme discrimination. Early Spanish-English bilinguals' sensitivity to the associations between segments and language does not significantly differ from the sensitivity of monolingual English listeners, and this sets their performance apart from the Spanish-Catalan bilinguals in Sebastián-Gallés et al. (2005), whose early exposure to and proficiency in Spanish is reported to have degraded their sensitivity to L2 contrasts. However, for the Early Bilinguals in the present study, their Spanish ability also did not improve the accuracy of their language classification decisions. This is in contrast to the performance of the late Spanish-English bilinguals, who categorized many segments more accurately (although also more slowly) than Monolinguals and Early Bilinguals.

The results of Experiment 1 raise many questions about how listeners perceive and use language-specific cues. Among the issues that warrant further investigation are 1) the possible speed-accuracy trade-off characterizing the differences in the bilingual groups' categorization decisions and 2) how changes in language experience over time influence the development of language representations. These questions are investigated in the subsequent studies in this dissertation. Experiment 2 further explores the categorization accuracy of the Early and Late Bilinguals in speeded and delayed language categorization tasks. If the improved performance of the Late Bilinguals in Experiment 1 is due to the increased time spent making language decisions, forcing listeners to respond more quickly may produce categorization patterns more like the Early Bilinguals. If the Early Bilinguals do have access to language representations as detailed as those of the Late Bilinguals, delaying the Early listeners' responses may improve their performance. The second issue of the stability of linguistic representations throughout development is explored in Experiment 3, in which six-year-olds complete an adapted version of Experiment 1 to determine when adult-like sensitivities to these language-specific cues emerge. Children with varying exposure to Spanish and English are also compared to further test how proficiency in a language shapes linguistic representations.

5. Experiment 2: The relationship between accuracy and reaction time in adults' language categorization decisions

5.1 INTRODUCTION

Experiment 2 probes the speed-accuracy trade-off potentially responsible for the differences in accuracy between the Early and Late Bilingual listeners in Experiment 1. In the first study, Early Bilinguals responded more slowly than the Monolingual listeners but faster than the Late Bilinguals, but the Early Bilinguals' accuracy patterns did not show a benefit of slower reaction times, possibly due to weaker connections between languages and language-specific segments. The accuracy of the Monolinguals and Early Bilinguals did not differ, and the Late Bilinguals were significantly more accurate than the other groups for phonemic cues and in some comparisons with the /l/ variants. This raises the question of what is the cause of the difference in accuracy of the Early and Late Bilinguals. Were the Late Bilinguals more accurate because they took more time to respond? Or was their performance due to more detailed linguistic representations that led to stronger associations between segmental cues and the respective languages? By the same token, were Early Bilinguals less accurate than Late Bilinguals because they responded more quickly? Or do the Early Bilinguals have less detailed phonetic representations of English and Spanish than do the Late Bilinguals?

Experiment 2 collected the categorization decisions of Monolingual, Early Bilingual, and Late Bilingual listeners in a timed task. Trials proceeded as in Experiment 1, but in Experiment 2 participants could not register a response until a specified delay had passed. The three delay periods were chosen to force listeners to respond faster, somewhat slower, and much slower than the range of responses to correct trials in Experiment 1. If the linguistic representations of Early and Late Bilinguals do not fundamentally differ, the longer response delays would potentially allow the Early Bilinguals to respond more accurately than they did in Experiment 1. A decrease in accuracy of the Late Bilinguals would then also be expected in the speeded trials. If Late Bilinguals have more detailed language-specific representations, stronger language-cue associations, or better access to either than the Early Bilinguals, the changes in reaction times should not affect the difference in accuracy between the bilingual groups. The following research question is addressed in Experiment 2, and the following hypotheses are tested:

2. Will listener accuracy change as the result of having more or less time to access language-specific phonological information?
   a. If early and late bilinguals have access to fundamentally equivalent information about each language's phonological categories, the accuracy of both groups will improve and become more similar when more time is allowed to make the categorization decision. When responses are forced to be faster, both groups will respond less accurately and more similarly to each other.
   b. If one group of bilinguals has encoded or has greater access to phonetic detail in their representations of English and Spanish, neither bilingual group is expected to perform differently than they did in Experiment 1, no matter how much time is allowed for a response.

5.2 Methodology

5.2.1 Participants

As for Experiment 1, listeners for Experiment 2 were recruited through the Department of Linguistics subject pool and received course credit (n=16) or responded to an advertisement in the University of Texas Events Calendar and were paid $10 (n=37).

Listeners either completed the same language history questionnaire as in Experiment 1 (Chan, 2014), if they had previously participated in a study with the UT Sound Lab, or a newly adapted questionnaire gathering the same information. Listeners' responses were used to classify them as Monolingual, Early Bilingual, or Late Bilingual according to the same criteria as in Experiment 1. See Table 5.1 for a summary of participant characteristics. Nineteen participants who did not fit into one of these groups were tested but excluded from the sample; these participants were recruited through the subject pool and were tested anyway in accordance with the Institutional Review Board protocol. Excluded participants studied Spanish in college, grew up speaking a language other than Spanish and English at home (e.g. Korean), began learning English after kindergarten, or grew up speaking English outside the U.S.[8]

Fourteen Monolingual listeners (7 females) were tested. Monolingual participants ranged in age between 18 and 29, with a mean age of 21. Ten of the 14 Monolinguals had exposure to Spanish, typically in classes in middle and/or high school. All listeners with some exposure to Spanish reported very low proficiency in the language.

There were 18 listeners in the Early Bilingual group (13 females), and their ages ranged from 18 to 43, with a mean age of 23 years. Fifteen of the Early Bilinguals were born and raised in the United States, and the remaining three participants were born in Mexico and moved to the U.S. by three years old. All listeners in the Early Bilinguals group had learned Spanish at home since birth. Four participants also learned English at home since birth (three of the U.S.-born participants, one of the Mexican-born participants). The

[8] But see Section 5.3.1 for the inclusion of some of these participants in an analysis of missing responses.

remaining 14 participants began learning English by the time they started elementary school.

Table 5.1: Demographic information and language background for participants in Experiment 2.

	Monolinguals	Early Bilinguals	Late Bilinguals
N	14	18	21
mean age	21	23	28
age range	18-29	18-43	19-39
females	7	13	15
mean age (in years) when learned English	0	2.9	8
mean age (in years) when learned Spanish	11.4	0	0
mean age (in years) when moved to U.S.	NA	0.4	21

In the Late Bilingual group there were 21 listeners (15 females); all were born and raised in Latin America and moved to the U.S. after age 13. Late Bilingual participants were between 19 and 39 years old, with a mean age of 28 years, and were from Mexico (n=12), Colombia (n=6), Argentina (n=1), Costa Rica (n=1), and Venezuela (n=1). Late Bilinguals moved to the U.S. between 13 and 32, with a mean age of arrival of 21. All listeners had learned only Spanish at home since birth. In this sample, seven listeners had attended a primary school where English was the language of at least 50% of the instruction, and seven listeners also received at least half of their secondary instruction in

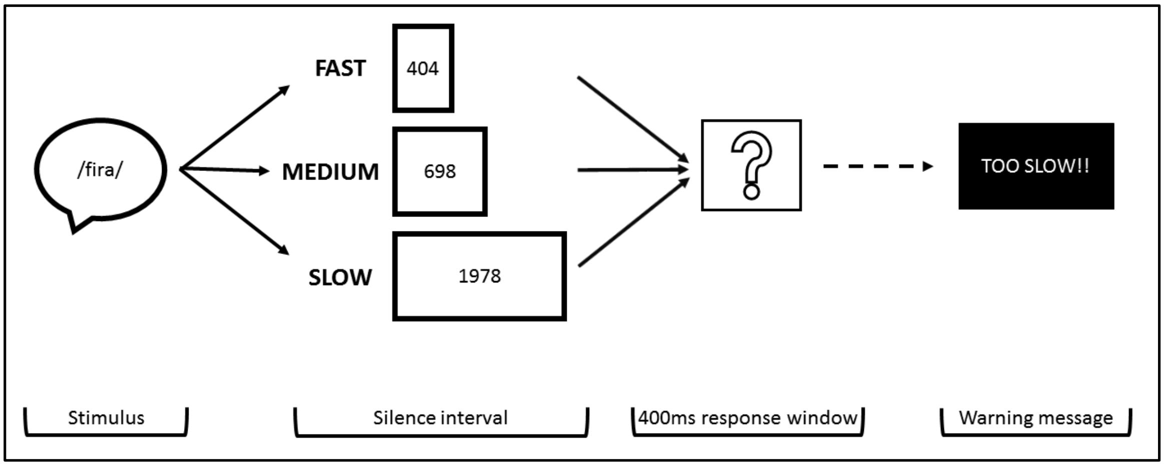
English. Four Late Bilinguals fell into both categories, and had received considerable English language instruction in both primary and secondary school.

5.2.2 Procedure

The testing procedure in Experiment 2 was the same as in Experiment 1: listeners completed the speeded nonce-word categorization task in the UT Sound Lab in a sound-attenuated booth seated in front of the same computer running the same software as in Experiment 1. Participants were oriented to the two buttons on the button box, labeled "ENG" and "SPAN" and counterbalanced across listeners, and completed the same number of test trials as in Experiment 1 (56 words x 8 blocks for 448 trials). The 56 words were randomly presented within each block. In Experiment 2, the eight blocks were divided into four response categories that varied the speed with which a participant was required to wait to respond to a stimulus. Each of these four variations included two repetitions of each stimulus (two blocks), but for convenience the four variations will be referred to as the four blocks. The first block replicated the instructions and limitless response timing of Experiment 1 (the "Self-Paced" block), and all listeners completed the Self-Paced block first.

The three timed blocks were presented after the Self-Paced block. In a pilot study, the different response times were randomized in a block, so that the silence interval occurred before a response window, which was signaled with a tone, and in which listeners did not know which length of the silence interval would be used in a given trial. The results of the pilot indicated that it was extremely difficult for listeners to successfully register responses when wait time was randomized. This design was based on the variable stimulus presentation times used in Diederich & Busemeyer (2006) and McElree et al. (2006); however, both studies used visually presented stimuli, and so their range of times was actually a range of stimulus *presentation* times, not of wait times. The use of auditory stimuli here makes varying the length of presentation times impossible, without fundamentally altering the stimulus itself, and so asking listeners to *wait* indeterminate amounts of time before responding may require different attentional resources than those used to focus on a stimulus for more or less time. Instead, the three different wait times were presented in blocks and the order of the remaining three blocks was counterbalanced across participants, yielding 12 different block presentation orders. Before a new block, listeners read instructions that repeated those for Experiment 1 and the Self-Paced block, but which additionally told participants "you will wait [X] seconds before you can respond. After you hear a word, listen for the tone. As soon as you hear the tone, respond immediately. If you respond too slowly you will see a warning on the screen" and "Respond as accurately as possible, but be sure to wait for the tone to respond!" The particular number of seconds, indicated with "[X]" in the sample above, are explained below. After reading these instructions, participants then completed a practice block of 10 real words (5 English, 5 Spanish), saw the instructions about the wait time and tone again, and then began the test trials. For each of the three timed blocks, listeners heard the stimulus, heard a specified length of silence after the offset of the word, and then heard a 50ms tone that indicated the beginning of the 400ms response window. If listeners did not respond during the 400ms response window, the screen went black and "TOO SLOW!!" appeared in white letters. Then next trial began immediately. See Figure 5.1 for an illustration of the structure of the trials in each block.

Figure 5.1: Schema of stimulus presentation, wait times, and response window for each block in Experiment 2.



Three silence interval lengths were chosen from across the distribution of response times from the Monolinguals' correct responses in Experiment 1. These wait times were chosen to force listeners, especially the Late Bilinguals, to make responses faster than they would naturally (i.e., than they would without a time limit, as in Experiment 1), and to make listeners, especially the Early Bilinguals, respond more slowly than they would naturally. The mean RT (50th percentile) in correct trials for Monolinguals was 404ms, and this was chosen as the fastest wait time ("Fast" block). Using the 50th percentile ensured that this wait time would be difficult for all listener groups, since in 50% of trials the Monolingual listeners would need more time for an accurate response. Almost 55% of Early Bilinguals responded longer than 404ms in correct trials in Experiment 1, and 66% of Late Bilinguals needed more than 404ms as well. The second wait length was 698ms, the 75th percentile of the Monolingual responses to correct trials in Experiment 1 ("Medium" Block). Seventy-one percent of Early Bilingual and 64% of Late Bilingual correct responses were made by 698ms in Experiment 1. The final wait lengths was 1978ms, which is the 97.5th percentile of the RTs for Monolinguals in correct trials in

Experiment 1 ("Slow" Block). For both Early and Late Bilinguals, 97% of correct responses were made by 1978ms. These three wait lengths were described as "0.4 seconds," "0.7 seconds," and "2 seconds" in the instructions presented to listeners.

5.3 Results

As in Experiment 1, categorization decisions and RTs were recorded; however, since the design of Experiment 2 controlled the time listeners waited before registering a categorization decision for three of the four blocks, the data were analyzed in three ways. First, missed trials are analyzed in Section 5.3.1; these are trials in the Fast, Medium, and Slow blocks in which a listener responded outside the designated response window or not at all (n=4321). These missed trials were omitted from the subsequent analyses, which were concerned with categorization accuracy, leaving the 81.1% of trials (n=18575) for which a response was registered. Next, in Section 5.3.2, the accuracy of listeners during the Self-Paced block are compared to the results of Experiment 1. Finally, Section 5.3.3 addresses changes in accuracy across the three timed blocks.

Accurate decisions were defined as those in which nonce words with an English-specific phone were categorized as English and in which nonce words with a Spanish-specific phone were categorized as Spanish. As before, *racha* and *rachuh* trials were excluded from the analysis, per Section 3.2 and footnote 5. RTs were calculated for the Self-Paced block by subtracting the length of the .wav file from the RT calculated by E-Prime, which reflected the time between trial onset and button press. Twenty trials (0.1%) were removed for having responses registered before the end of the stimulus word. Trials in the Self-Paced block that were more than three standard deviations above or below a participant's mean were classified as outliers and removed from the final data set (n=92; 0.5%). For the Self-Paced responses, RTs were log-transformed from milliseconds, which normalized the distribution and reduced the distance of the few scores (n=11; 0.1%) longer than 5000ms from the mean.

5.3.1 Missed trials

Due to the design of the experiment, responses to trials in the Fast, Medium, and Slow blocks were only recorded during the 400ms response window indicated by a 50ms tone. If listeners responded while the stimulus played, during the measured silence (404ms, 698ms, or 1978ms), or after the 400ms response window, no response was recorded and the participant saw the “TOO SLOW!!” message. This message appeared even in cases when, in fact, the listener had responded too quickly – during the word or during the silence interval. However, listeners were given 10 practice trials to get used to the timing of the response in the block and then again saw the instructions, which emphasized the need to wait for the tone, before the test trials began. Despite seeing the instructions about waiting for the tone twice for each block (and six times overall) and completing several dozens of test trials, many listeners struggled to respond during the designated 400ms window. Experiment 2 was designed to be challenging, for the sake of forcing listeners to respond outside their normal RT, but it was expected to challenge the different listener groups to greater and lesser extents. Given the distribution of RTs to correct trials in Experiment 1, Monolinguals were expected to be the least challenged by the wait lengths, with fewer Early Bilinguals expected to naturally respond at these wait lengths, and with even fewer Late Bilinguals naturally responding at these rates; see Section 5.2.2. However, if the Bilingual groups were less likely to respond at faster speeds, it would be unclear if this effect was due to the listeners’ statuses as bilinguals – that bilinguals in general respond more slowly in language categorization tasks – or if the effect was because these particular bilinguals had competing representations of English and Spanish, and this competition and activation of both languages representations made them unable to respond more quickly.

Table 5.2: Percentage of trials with missing responses in Experiment 2.

	Monolinguals	Early Bilinguals	Late Bilinguals	Other Bilinguals
Fast (404ms)	9.6 (29.5)	40.0 (48.9)	31.0 (46.3)	17.2 (37.7)
Medium (698ms)	7.5 (26.4)	34.0 (47.4)	25.0 (43.3)	22.5 (41.8)
Slow (1978ms)	5.2 (22.3)	28.1 (45.0)	26.6 (44.2)	12.2 (32.8)
Total missed trials	7.5 (26.3)	33.9 (47.3)	27.5 (44.7)	17.3 (37.8)

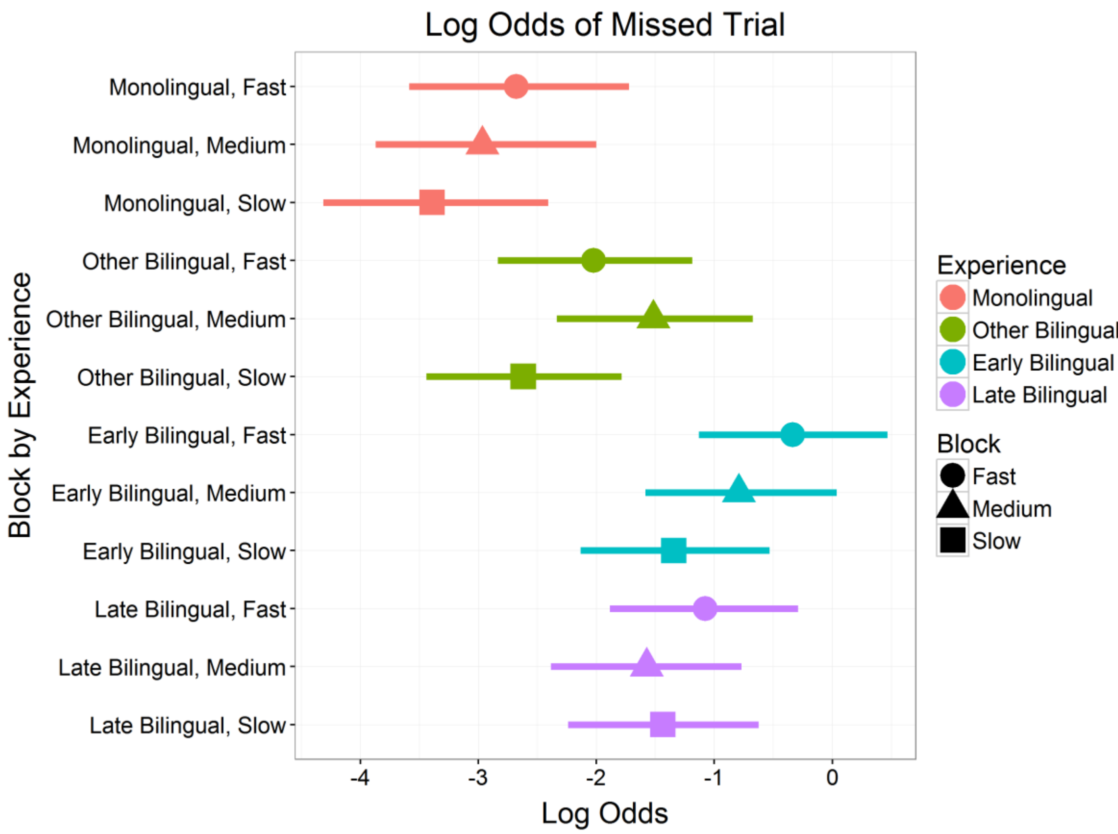
This possibility was tested by including a fourth listener group, Other Bilinguals, in the following analysis of missed trials. These 11 listeners met the same criteria as the Early Bilinguals of interest in the main study, except that instead of Spanish these listeners had heard another language at home since birth. All received course credit for their participation. If these listeners suffered more missed trials than Monolinguals, listeners' bilingualism may be the cause of the longed RTs and would suggest that the existence of competing representations across languages – no matter the languages – makes the task difficult to complete. If these listeners instead were able to respond at the same rates as Monolinguals, the Early and Late Bilinguals' missed trials may be due to the competition of English and Spanish representations in particular when making categorization decision. See Table 5.2 for the mean percent of missed trials for each listener group in each block. Standard deviations are included in parentheses, and the total percentage of missed trials is reported for the three timed blocks and excludes the first, Self-Paced block.

Table 5.3: Summary of mixed effects logistic regression model fitting missed trials in Experiment 2.

Predictor	*Mean*	*Posterior SD*	*95% CI*	*p value*
Intercept (Other Bilingual, Fast Block)	-2.023	0.416	(-2.833, -1.185)	
Monolingual	-0.655	0.627	(-1.907, 0.603)	<0.10
Early Bilingual	1.687	0.602	(0.442, 2.895)	<0.0001
Late Bilingual	0.945	0.558	(-0.126, 2.022)	<0.05
Medium Block	0.506	0.097	(0.315, 0.699)	NS
Slow Block	-0.596	0.110	(-0.812, -0.376)	<0.10
Monolingual * Medium	-0.795	0.172	(-1.135, -0.456)	<0.05
Monolingual * Slow	-0.118	0.189	(-0.503, 0.245)	<0.01
Early Bilingual * Medium	-0.962	0.135	(-1.224, -0.699)	<0.001
Early Bilingual * Slow	-0.412	0.147	(-0.704, -0.131)	<0.10
Late Bilingual * Medium	-1.003	0.130	(-1.254, -0.742)	<0.001
Late Bilingual * Slow	0.238	0.141	(-0.033, 0.509)	<0.10

The distribution of missing responses was analyzed using a Bayesian mixed effects logistic regression model with listener language group (four levels: Monolingual, Early Bilingual, Late Bilingual, Other Bilingual), target segment (seven levels: /r, ɹ, θ/ and [l, u, ɫ, ʉ]), and block (three levels: Fast, Medium, Slow) as fixed effects, and participant and stimulus word as random intercepts. Note that the Self-Paced block was excluded from analysis since it was impossible to miss a trial when there was no time restriction on the response. The models were fit via a Markov Chain Monte Carlo procedure using STAN (Gelman et al., 2015). Model comparison was performed using the Deviance Information Criterion (DIC; Spiegelhalter et al., 2002). A model with main effects of listener group and block improved the fit of a model with listener group alone, and model fit was further improved by including an interaction between listener group and block. Adding segment did not significantly improve the model further. The model is summarized in Table 5.3. The model intercept reflects the likelihood of a missed trial for the Other Bilingual group in the Fast block. The fitted log odds of failing to respond during the response window are plotted in Figure 5.2. Error bars represent the 95% Bayesian credible intervals.

Figure 5.2: Log odds of missed trial, by block and listener group in Experiment 2.



The four listener groups differed in how the three timed blocks affected the number of trials they missed. The Monolinguals responded mostly consistently across blocks, and there was no significant difference in missed trials between Fast and Medium blocks (β=0.289, posterior SD=0.141, p=0.27) or between Medium and Slow blocks (β=0.425, posterior SD=0.158, p=0.19), although there was a trend for there to be more missed trials in the Fast block than in the Slow block (β=0.714, posterior SD=0.153, p<0.10). For the Early Bilinguals, there was no difference in rate of missed trials between Fast and Medium blocks (β=0.456, posterior SD=0.92, p=0.13), but the difference between Medium and Slow blocks (β=0.552, posterior SD=0.099, p<0.10) approached significance. There were also more missed trials in the Fast block than in the Slow block for Early Bilinguals (β=1.008, posterior SD=0.096, p<0.01). For the Late Bilinguals, there were no significant differences between blocks (Fast vs. Medium: β=0.497, posterior SD=0.089, p=0.12; Fast vs. Slow: β=0.358, posterior SD=0.088, p=0.18; Medium vs. Slow: β=0.139, posterior SD=0.092, p=0.36). Other Bilingual listeners missed as many trials in the Fast block as in the Medium block (β=0.506, posterior SD=0.097, p=0.11), and there was a trend for them to miss significantly fewer in the Slow than in the Fast block (β=0.596, posterior SD=0.110, p<0.10). Other Bilinguals missed more trials in the Medium block than in the Slow block (β=1.102, posterior SD=0.108, p<0.01).

In the most challenging block, the Fast block, the Early Bilinguals registered more missed trials than any of the other groups and the Monolinguals fewer than the other groups. Monolinguals were significantly less likely than the Early Bilinguals or Late Bilinguals to fail to respond during the 400ms response window (vs. Early: β=2.342, posterior SD=0.621, p<0.0001; vs. Late: β=1.600, posterior SD=0.624, p<0.01). The difference between Monolinguals and Other Bilinguals approached significance, with Monolinguals more likely to register a response (β=0.655, posterior SD=0.627, p<0.10). The Early Bilinguals missed more trials than the Late Bilinguals (β=0.742, posterior SD=0.563, p<0.05) and Other Bilinguals (β=1.687, posterior SD=0.602, p<0.0001), and the Late Bilinguals missed more trials than the Other Bilinguals (β=0.945, posterior SD=0.558, p<0.05).

These patterns among groups remained in the Medium block, in which listeners had to wait 698ms to respond to stimuli. The Monolinguals had significantly fewer missed trials than the other groups (vs. Early Bilinguals: β=2.176, posterior SD=0.624, $p<0.0001$; vs. Late Bilinguals: β=1.393, posterior SD=0.629, $p<0.01$; vs. Other Bilinguals: β=1.450, posterior SD=0.630, $p<0.01$). Early Bilinguals registered more missed trials than Late Bilinguals (β=0.783, posterior SD=0.562, $p<0.05$) and Other Bilinguals (β=0.726, posterior SD=0.597, $p<0.05$), and in this block there was no difference between Late and Other Bilinguals (β=0.058, posterior SD=0.561, $p=0.44$).

In the Slow Block, most of these patterns persisted. Monolinguals again had the fewest missed trials (vs. Early Bilinguals: β=2.049, posterior SD=0.629, $p<0.0001$; vs. Late Bilinguals: β=1.956, posterior SD=0.632, $p<0.0001$; vs. Other Bilinguals: β=0.773, posterior SD=0.639, $p<0.10$), and the Other Bilinguals registered a response more often than the Early or Late Bilinguals (vs. Early: β=1.275, posterior SD=0.604, $p<0.01$; vs. Late: β=1.183, posterior SD=0.566, $p<0.01$). In the Slow block there was no difference between Early and Late Bilinguals (β=0.092, posterior SD=0.563, $p=0.40$).

5.3.2 Self-paced decisions

The accuracy results of the Self-Paced block in Experiment 2 are compared with those from Experiment 1 to ensure replicability of the effects reported in Chapter 4, before exploring variation in accuracy by block. The Experiment 1 responses were limited to the first two blocks of the eight total repetitions that each listener heard, since listeners in Experiment 2 completed the equivalent of these two blocks at their own pace before having to respond at specific wait lengths. This new data set – Blocks 1 and 2 from Experiment 1 and Self-Paced responses from Experiment 2 – was analyzed using a mixed effects logistic regression model with the *lme4* package (v1.1-7) in R (v3.2.2). Listener language group (three levels: Monolingual, Early Bilingual, Late Bilingual), target segment (seven levels: /r, ɹ, θ/ and [l, u, ɫ, ʉ]), and experiment (1 or 2) were tested as fixed effects, and listener and stimulus word were included as random intercepts. As in Experiment 1, a model with an interaction between listener group and target segment significantly improved the fit of a model without the interaction (χ^2=71.595, df=12, p<0.0001). The addition of experiment as a predictor did not improve model fit (χ^2=0.111, df=1, p=0.74). The lack of effect of experiment indicates that the accuracy results did not differ between self-paced trials in the two studies.

Table 5.4: Mean accuracy of each listener group for each stimulus type in each of the four blocks in Experiment 2.

(A) Monolinguals

Block		Self-Paced	Fast (404ms)	Medium (698ms)	Slow (1978ms)
Phonemic Cues	Spanish /r/	98.0 (14.2)	90.8 (29.1)	95.6 (20.7)	97.3 (16.3)
	English /ɹ/	88.1 (32.4)	91.8 (27.6)	89.1 (31.3)	95.1 (22.7)
	English /θ/	54.4 (49.9)	71.1 (45.4)	69.5 (46.2)	75.1 (43.3)
Phonetic Cues	Spanish [l]	94.5 (22.8)	86.2 (34.6)	92.8 (25.9)	88.7 (31.7)
	English [ɫ]	74.8 (43.5)	76.2 (42.7)	80.7 (39.6)	84.8 (36.0)
	Spanish [u]	85.1 (35.7)	89.1 (31.3)	90.0 (30.1)	88.7 (31.8)
	English [ʉ]	65.6 (47.6)	73.9 (44.0)	72.8 (44.6)	74.5 (43.7)

Table 5.4, cont.

(B) Early Bilinguals

Block		Self-Paced	Fast (404ms)	Medium (698ms)	Slow (1978ms)
Phonemic Cues	Spanish /r/	97.2 (16.5)	93.8 (24.1)	98.3 (13.1)	94.6 (22.6)
	English /ɹ/	91.2 (28.3)	87.2 (33.6)	92.3 (26.8)	92.9 (25.8)
	English /θ/	56.7 (49.6)	70.3 (45.8)	74.5 (43.7)	75.7 (43.0)
Phonetic Cues	Spanish [l]	96.5 (18.4)	85.4 (35.5)	90.4 (29.6)	92.4 (26.5)
	English [ɫ]	74.5 (43.7)	73.6 (44.2)	78.1 (41.4)	78.2 (41.4)
	Spanish [u]	88.8 (31.5)	85.1 (35.7)	91.7 (27.6)	89.7 (30.5)
	English [ʉ]	52.2 (50.0)	59.5 (49.2)	54.2 (50.0)	63.0 (48.4)

Table 5.4, cont.

(C) Late Bilinguals

Block		Self-Paced	Fast (404ms)	Medium (698ms)	Slow (1978ms)
Phonemic Cues	Spanish /r/	97.1 (16.8)	97.0 (17.2)	94.8 (22.3)	97.6 (15.5)
	English /ɹ/	95.6 (20.5)	93.3 (25.0)	95.8 (20.2)	98.1 (13.6)
	English /θ/	75.6 (43.0)	86.0 (34.7)	76.1 (42.7)	87.2 (33.4)
Phonetic Cues	Spanish [l]	92.0 (27.2)	91.6 (27.8)	95.9 (19.8)	91.8 (27.5)
	English [ɫ]	82.3 (38.2)	85.3 (35.5)	86.2 (34.6)	91.2 (28.4)
	Spanish [u]	88.4 (32.0)	90.2 (29.8)	90.3 (29.7)	93.6 (24.5)
	English [ʉ]	70.0 (45.9)	77.5 (41.8)	76.3 (42.6)	80.3 (39.9)

5.3.3 Accuracy analysis

Mean accuracy rates for each stimulus type for each listener group in each block are reported in Table 5.4, and standard deviations are presented in parentheses. Missing trials were removed from the final data set, since there was no accuracy information available for these trials. The Self-Paced trials were included in the analysis, but since accuracy for these trials did not differ from Experiment 1 (see the previous section) they will not be reported on here. As in Experiment 1, the accuracy results were analyzed using a Bayesian mixed effects logistic regression model with listener language group (three levels: Monolingual, Early Bilingual, Late Bilingual), target segment (seven levels: /r, ɹ, θ/ and [l, u, ɫ, ʉ]), and block (four levels: Self-Paced, Fast, Medium, Slow) as fixed effects, block order (six counterbalancing orders) as a covariate, and participant and stimulus word as random intercepts. An interaction between listener group and target segment was anticipated, given the results of Experiment 1. If additional time to respond allowed the Early Bilinguals' accuracy to improve, or if less time caused the Late Bilinguals' accuracy to decrease, a three-way interaction of block, listener group, and segment would reveal differences in accuracy for the difficult segments across the blocks for the bilinguals. As in Section 5.3.1, the models were fit via a Markov Chain Monte Carlo procedure in STAN, and DIC was used for model comparison. Participant and stimulus word were included as random effects on the intercept for all models. The null model with only the two random effects was significantly improved by adding listener group and target segment, as in Experiment 1. The model fit was further improved by adding block as a fixed effect, but the addition of block order did not improve model fit. Next interactions were tested. Also as in Experiment 1, the interaction between listener group and target segment was significant. Adding an interaction between block and segment significantly improved fit, but the interactions between block and listener group did not improve fit. This suggests that the effect of block was uniform across listener groups. The final model, including interactions of segment with listener group and segment with block, is summarized in Table 5.5. The intercept represent the log odds of an accurate response for Monolinguals hearing a nonce word with Spanish /r/ in the Fast block.

Table 5.5: Summary of mixed effects logistic regression model fitting accuracy results in Experiment 2.

Predictor	*Mean*	*Posterior SD*	*95% CI*	*p value*
Intercept (Monolingual, Spanish /r/, Fast Block)	3.146	0.479	(2.193, 4.091)	<0.0001
English /ɹ/	-0.665	0.546	(-1.708, 0.413)	<0.10
English /θ/	-2.003	0.520	(-3.024, -0.999)	<0.0001
Spanish [l]	-0.793	0.517	(-1.811, 0.213)	<0.05
English [ɫ]	-1.539	0.516	(-2.564, -0.524)	<0.0001
Spanish [u]	-0.797	0.536	(-1.908, 0.234)	<0.05
English [ʉ]	-1.868	0.517	(-2.896, -0.864)	<0.0001
Early Bilinguals	0.019	0.440	(-0.824, 0.924)	NS
Late Bilinguals	0.076	0.440	(-0.737, 0.978)	NS
Medium Block	0.545	0.292	(-0.028, 1.121)	NS
Slow Block	0.618	0.305	(0.022, 1.231)	<0.10
Block SP	1.003	0.316	(0.380, 1.619)	<0.05
English /ɹ/ * Early	-0.087	0.334	(-0.732, 0.576)	<0.10
English /θ/ * Early	0.120	0.304	(-0.463, 0.720)	<0.0001
Spanish [l] * Early	0.027	0.329	(-0.613, 0.677)	<0.05
English [ɫ] * Early	-0.243	0.308	(-0.833, 0.366)	<0.0001
Spanish [u] * Early	-0.056	0.318	(-0.665, 0.574)	<0.05
English [ʉ] * Early	-0.801	0.300	(-1.384, -0.206)	<0.0001
English /ɹ/ * Late	0.631	0.350	(-0.054, 1.311)	NS
English /θ/ * Late	0.839	0.302	(0.257, 1.434)	<0.01
Spanish [l] * Late	-0.030	0.328	(-0.667, 0.610)	<0.05
English [ɫ] * Late	0.461	0.312	(-0.142, 1.067)	<0.01
Spanish [u] * Late	-0.038	0.318	(-0.667, 0.599)	<0.05
English [ʉ] * Late	0.026	0.300	(-0.567, 0.614)	<0.0001
English /ɹ/ * Medium	-0.310	0.374	(-1.041, 0.427)	NS
English /ɹ/ * Slow	0.153	0.402	(-0.631, 0.938)	NS
English /ɹ/ * Self-Paced	-0.846	0.387	(-1.608, -0.087)	NS
English /θ/ * Medium	-0.716	0.327	(-1.354, -0.088)	<0.0001
English /θ/ * Slow	-0.381	0.339	(-1.047, 0.286)	<0.0001
English /θ/ * Self-Paced	-1.860	0.344	(-2.545, -1.187)	<0.0001
Spanish [l] * Medium	0.212	0.364	(-0.520, 0.909)	NS
Spanish [l] * Slow	-0.307	0.366	(-1.041, 0.401)	NS
Spanish [l] * Self-Paced	-0.078	0.377	(-0.817, 0.664)	NS
English [ɫ] * Medium	-0.195	0.343	(-0.871, 0.478)	<0.01
English [ɫ] * Slow	0.061	0.356	(-0.627, 0.768)	<0.05

Table 5.5, cont.

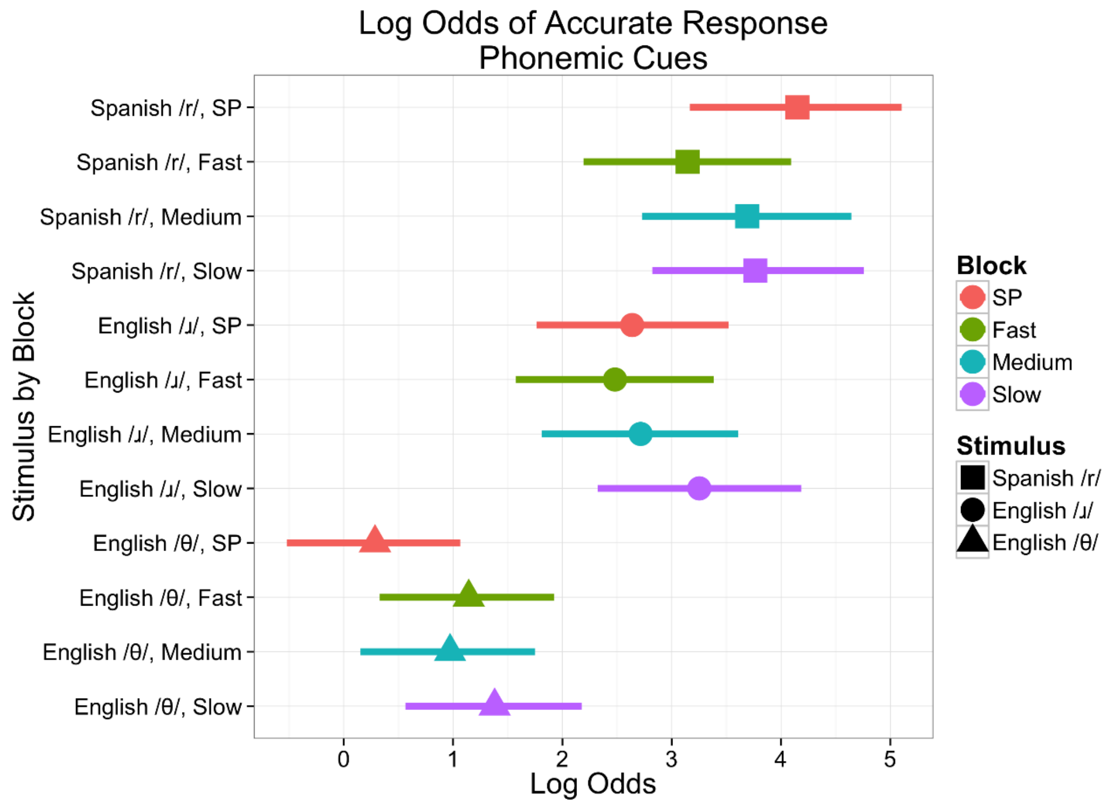
English [ɫ] * Self-Paced	-1.009	0.350	(-1.684, -0.323)	<0.0001
Spanish [u] * Medium	-0.238	0.347	(-0.921, 0.458)	NS
Spanish [u] * Slow	-0.329	0.364	(-1.060, 0.386)	NS
Spanish [u] * Self-Paced	-1.008	0.361	(-1.698, -0.301)	<0.05
English [ʉ] * Medium	-0.603	0.322	(-1.232, 0.042)	<0.0001
English [ʉ] * Slow	-0.454	0.336	(-1.105, 0.193)	<0.0001
English [ʉ] * Self-Paced	-1.383	0.345	(-2.076, -0.694)	<0.0001

Random effects	*Variance*
Listener	0.978
Stimulus Word	0.821

Phonemic cue accuracy, across blocks

As in Experiment 1, the interaction between target segment and listener group was significant, and the results of Section 5.3.2 indicate that there was no systematic difference in listener accuracy by segment or listener group across Experiment 1 and the Self-Paced block of Experiment 2. As such, this analysis will focus on the influence of the timed blocks and the interaction of block and segment in the present results. Given the structure of the model output, the betas, posterior standard deviations, and *p*-values reported below are associated with the Monolingual listeners, but the effects do not differ in the other groups. The fitted log odds of accuracy for each target segment and block are plotted in Figure 5.3 for phonemic cues and Figure 5.4 for phonetic cues. Again, these log odds were calculated for the Monolinguals, but are representative of the other groups as well. Error bars represent the 95% Bayesian credible intervals, and the Self-Paced block is abbreviated "SP."

Figure 5.3: Log odds of accuracy for phonemic cues, by segment and block, in Experiment 2.



Responses to nonce words containing the Spanish /r/ were categorized similarly in the three timed blocks. The Fast block was not significantly different from the Medium block (β=0.545, posterior SD=0.292, *p*=0.12), and there was a trend for more accurate responses to Spanish /r/ in the Slow versus the Fast block (β=0.618, posterior SD=0.305, *p*<0.10). There were no differences in accuracy between the Medium and Slow blocks (β=0.073, posterior SD=0.322, *p*=0.45).

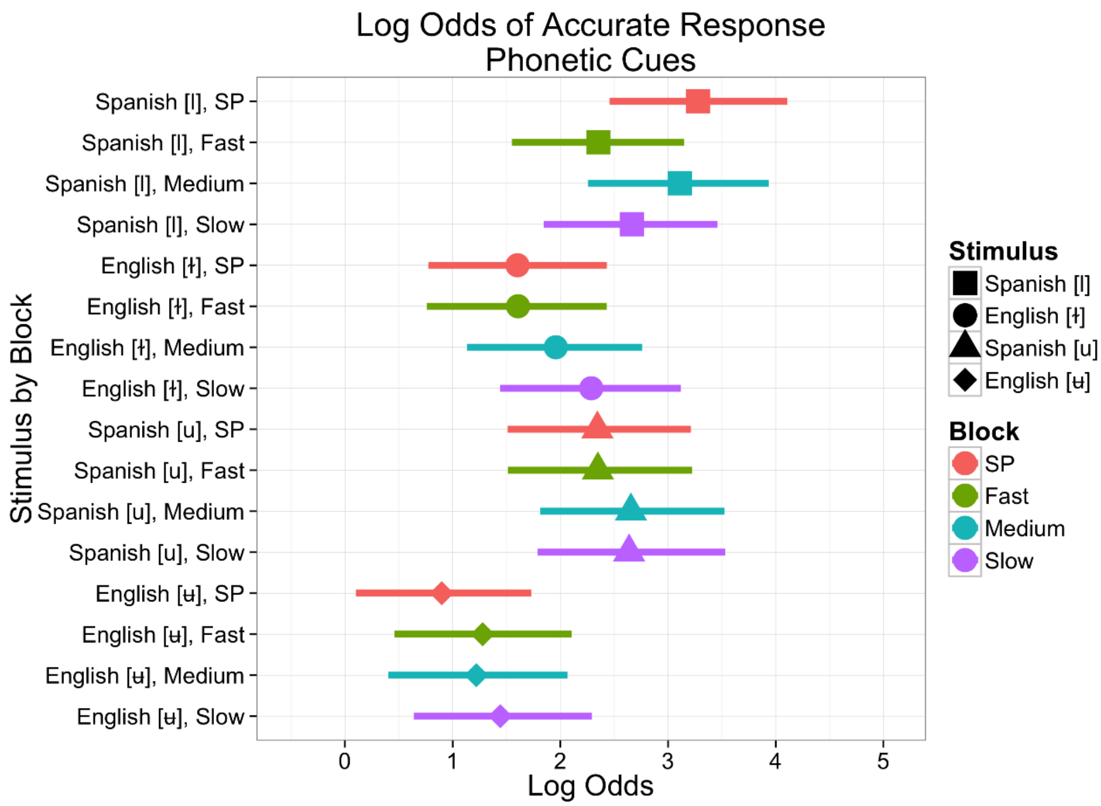
Accuracy was also largely consistent across blocks for English /ɹ/. The Fast and Medium blocks did not differ (β=0.234, posterior SD=0.236, *p*=0.31), and nor did the Medium and Slow blocks (β=0.537, posterior SD=0.262, *p*=0.12), but listeners were more likely to respond accurately in the Slow block than in the Fast block (β=0.772, posterior SD=0.262, $p<0.05$).

For nonce words with English /θ/, there were no significant differences among the timed blocks (Fast vs. Medium: β=0.171, posterior SD=0.150, $p=0.34$; Fast vs. Slow: β=0.237, posterior SD=0.156, $p=0.16$; Medium vs Slow: β=0.408, posterior SD=0.156, $p=0.16$).

Phonemic cue accuracy, within blocks

The effect of wait length had little effect on the relative accuracy of each target segment. In all three timed blocks, Spanish /r/ was categorized more accurately than English /θ/ (Fast: β=2.003, posterior SD=0.520, $p<0.0001$; Medium: β=2.718, posterior SD=0.532, $p<0.0001$; Slow: β=2.384, posterior SD=0.538, $p<0.0001$), and English /ɹ/ more accurately than English /θ/ (Fast: β=1.338, posterior SD=0.491, $p<0.01$; Medium: β=1.743, posterior SD=0.492, $p<0.0.0001$; Slow: β=1.872, posterior SD=0.507, $p<0.0001$). Accuracy of Spanish /r/ compared to English /ɹ/ varied by block. In the Fast block, responses to Spanish /r/ trended towards being significantly more likely to be correct than responses to English /ɹ/ (β=0.665, posterior SD=0.546, $p<0.10$), while in the Medium block the difference was significant (β=0.975, posterior SD=0.553, $p<0.05$) and in the Slow block there was no difference (β=0.512, posterior SD=0.582, $p=0.13$).

Figure 5.4: Log odds of accuracy for phonetic cues, by segment and block, in Experiment 2.



Phonetic cue accuracy, across blocks

For stimuli with the Spanish [l], responses in the Medium block tended to be more accurate than in the other blocks. Listeners responded significantly less accurately in the Fast block than in the Medium block (β=0.756, posterior SD=0.213, p<0.05), and there was no difference between the Fast and Slow blocks (β=0.311, posterior SD=0.200, p=0.24). There was also no significant difference between the Medium and Slow blocks (β=0.445, posterior SD=0.223, p=0.16).

Listeners trended towards more accurate responses to nonce words with English [ɫ] in the Slow block, and accuracy in the other blocks did not differ. There was no difference in accuracy between the Fast and Medium blocks (β=0.350, posterior SD=0.176, p=0.20) or Medium and Slow blocks (β=0.329, posterior SD=0.180, p=0.22). The difference between the Fast and Slow blocks approached significance, with a greater likelihood of correct responses in the Slow block (β=0.680, posterior SD=0.178, p=0.05).

For both variants of /u/, there were no differences in accuracy across the blocks. Performance for Spanish [u] stimuli was statistically the same among the blocks (Fast vs. Medium: β=0.307, posterior SD=0.198, p-0.24; Fast vs. Slow: β=0.290, posterior SD=0.198, p=0.25; Medium vs. Slow: β=0.017, posterior SD=0.200, p=0.48). For nonce words with English [ʉ], there were also no significant differences in accuracy among the blocks (Fast vs. Medium: β=0.059, posterior SD=0.141, p=0.44; Fast vs. Slow: β=0.164, posterior SD=0.142, p=0.35; Medium vs. Slow: β=0.223, posterior SD=0.141, p=0.29).

Phonetic cue accuracy, within blocks

Two patterns in relative accuracy among phonetic cues remained stable across the timed blocks. In all three timed blocks, Spanish [l] was categorized as accurately as Spanish [u] (A: β=0.005, posterior SD=0.466, p=0.49; B: β=0.454, posterior SD=0.473, p=0.16; C: β=0.026, posterior SD=0.467, p=0.48). Across all blocks, Spanish [u] was categorized more accurately than English [ʉ] (A: β=1.070, posterior SD=0.464, p<0.01; B: β=1.436, posterior SD=0.469, p<0.001; C: β=1.196, posterior SD=0.467, p<0.01).

For the other two comparisons, the effect of Block on accuracy varied. For Spanish [l] and English [ɫ], the Spanish variants were more likely to be categorized accurately in the Fast (β=0.746, posterior SD=0.469, p<0.05) and Medium (β=1.152, posterior SD=0.477, p<0.01) blocks, but this difference was neutralized in the Slow block (β=0.378, posterior SD=0.475, p=0.19). English [ɫ] and English [ʉ] were equally difficult in the Fast block (β=0.329, posterior SD=0.480, p=0.21), but in blocks with longer wait lengths

English [ɫ] was categorized significantly more accurately than English [ʉ] (Medium: $\beta=0.738$, posterior SD=0.480, $p<0.05$; Slow: $\beta=0.845$, posterior SD=0.486, $p<0.05$).

The results of the accuracy analyses by block and target segment are summarized in Tables 5.6 and 5.7. Table 5.6 summarizes the accuracy of the segments in each block, and Table 5.7 summarizes the relative of accuracy in the blocks for each segment. The “=” indicates differences that were not significant, the “>” and “<” are used for significant differences, and “»” and “«” represent differences that approached significance.

Table 5.6: Summary of accuracy results for target segment comparisons in Experiment 2.

(A) Phonemic Cues

	Target Segments
Fast Block	Spanish /r/ » English /ɹ/ > English /θ/
Medium Block	Spanish /r/ > English /ɹ/ > English /θ/
Slow Block	Spanish /r/ = English /ɹ/ > English /θ/

(B) Phonetic Cues

<table>
<tr><th></th><th>Cross-language</th><th>Cross-segment</th></tr>
<tr><td>Fast Block</td><td rowspan="2">Spanish [l] > English [ɫ]
Spanish [u] > English [ʉ]</td><td>Spanish [l] = Spanish [u]
English [ɫ] = English [ʉ]</td></tr>
<tr><td>Medium Block</td><td rowspan="2">Spanish [l] = Spanish [u]
English [ɫ] > English [ʉ]</td></tr>
<tr><td>Slow Block</td><td>Spanish [l] = English [ɫ]
Spanish [u] > English [ʉ]</td></tr>
</table>

Table 5.7: Summary of accuracy results for block comparisons in Experiment 2.

(A) Phonemic Cues

	Blocks
Spanish /r/	Fast « Slow Fast = Medium, Medium = Slow
English /ɹ/	Fast < Slow Fast = Medium, Medium = Slow
English /θ/	Fast = Medium = Slow

(B) Phonetic Cues

<table>
<tr><td></td><td>Blocks</td></tr>
<tr><td>Spanish [l]</td><td>Fast < Medium

Fast = Slow, Medium = Slow</td></tr>
<tr><td>English [ɫ]</td><td>Fast < Slow

Fast = Medium, Medium = Slow</td></tr>
<tr><td>Spanish [u]</td><td rowspan="2">Fast = Medium = Slow</td></tr>
<tr><td>Spanish [ʉ]</td></tr>
</table>

5.4 Discussion

Experiment 2 tested for changes in categorization accuracy by controlling how much time listeners had to decide to which language a nonce word belonged. This experiment was designed to address a potential speed-accuracy trade-off reflected in the accuracy rates and RTs of the Early and Late Bilinguals in Experiment 1: Early Bilinguals often responded significantly faster than Late Bilinguals, but Late Bilinguals were often more accurate than Early Bilinguals. In Experiment 2, listeners categorized the same nonce words as in Experiment 1 in four blocks that varied the time between stimulus and the response window. There was a significant interaction of target segment and listener group, as in Experiment 1, and there was also an interaction between target segment and block, with responses in the Slow block, with the longest response delay, trending towards better accuracy than in the Fast block, with the shortest response delay. Model comparisons revealed no interaction between listener group and block and no three way interaction: the differences among the groups, and among the groups for each segment, did not differ across the blocks.

Accuracy across segments and blocks

Listeners' categorization patterns for the seven target segments were the same in Experiment 1 and in the Self-Paced block in Experiment 2, but the structure of Experiment 2 influenced the accuracy with which some segments were categorized during the subsequent timed blocks. For at least one sound, the English phonemic cue /θ/, the timed blocks seemed to force greater attention to the categorization task, and accuracy rates in the Fast, Medium, and Slow blocks were all significantly greater than in the Self-Paced block, even when this meant being forced to respond after only 404ms (vs. Fast: $\beta=0.857$, posterior SD=0.142, $p<0.05$; vs. Medium: $\beta=0.686$, posterior SD=0.138, $p<0.05$; vs. Slow: $\beta=1.094$, posterior SD=0.143, $p<0.01$). Even so, listener accuracy in response to English /θ/ was still significantly worse than the other phonemic cues. While it remains true that /θ/ is a perceptually difficult segment, there are at least two factors that improve its perception and successful categorization: having more recently acquired the language in which /θ/ is a phoneme, as was the case for the Late Bilinguals in Experiment 1, and increasing the demands of and thus the attention required for the categorization tasks, which is reflected in the greater success in the Experiment 2 timed blocks. However, this benefit for increased attention did not also help listeners categorize the other difficult segment, English [ʉ], so there may be other explanations for the improved accuracy in the timed blocks for English /θ/.

The timed blocks also revealed the stability of some categorization patterns among the phonetic cues. In Experiment 1 it was observed that listeners categorized Spanish cues better than English cues and /l/ variants better than /u/ variants. Here, the greater accuracy in response to nonce words with Spanish [l] than to words with Spanish [u] was maintained across all wait lengths, suggesting that the perceptual saliency of /l/ versus /u/ is unrelated to the link between these segments and the language to which they belong and may be rooted in more general perceptual abilities. Similarly, the Spanish [u] was consistently categorized more accurately than English [ʉ], so having more or less time to make a decision does not improve how listeners use English [ʉ] in context. This also confirms that there is something more difficult perceptually about the English [ʉ], but unlike the English /θ/, language experience and more time to make a categorization decision do not seem to improve listeners' ability to identify it as an English-specific phone. As was discussed in Section 4.4, the phonetic variants of /u/ were the only vowels included in this project, so these patterns may reflect difficulties using vowels in context more generally and may not be specific to English [ʉ] or /u/ across languages. This may also explain why early Spanish-Catalan bilinguals in Amengual (2015), who were able to identify Catalan-specific vowel contrasts in isolation, were unable to identify mispronunciations related to vowels: vocalic information may be less salient to listeners than consonantal cues in the context of a word.

The difficulty categorizing English [ʉ], the /u/ variants more generally, and the mid-vowels in Amengual (2015) are at odds with the relative strength of vowels over consonants in other perception tasks. The intelligibility of vowel-only sentences (in which consonants have been replaced with noise) is better than consonant-only sentences (Fogerty & Kewley-Port, 2009; Kewley-Port et al., 2007), vowels are perceived more accurately in noise than consonants (Cutler et al., 2004), and the discrimination of consonants but not vowels declines with cortical electrical interference (Boatman et al., 1997). The results of the latter study also suggest that consonants and vowels are processed in different parts of the cortex. These studies conclude that the perception of consonants is more dependent on transitions and vocalic information, and so these cues must be available in the input and then integrated for identification. Kewley-Port et al. (2007) also points out that in the intelligibility studies that have found the preference for vowels, listeners can use top-down information to aid in perception. In studies that have focused on meaningful and nonce monosyllables, listeners rely more on bottom-up processing that begins with identifying segmental information, and looking up words in this way leads to an advantage for consonants (Owren & Cardillo, 2006). This is precisely the case in the present study: absent meaningful contextual information via lexical semantics, listeners were forced to rely on segmental information and so may have prioritized consonants to "look up" the language best associated with the segments of the nonce words.

While English [ʉ] was difficult to categorize across all blocks, the salience of other phonetic cues did vary by block. English [ɫ] was less accurately categorized than Spanish [l] but better than English [ʉ] in Experiment 1, and both patterns depended on block in Experiment 2. The difference in accuracy between Spanish and English /l/ variants was neutralized after the longest silence interval, in the Slow block, but the improved categorization of English [ɫ] over English [ʉ] was reduced to non-significant in the Fast response block. Given enough time, and in a task demanding increased attention, listeners were better able to draw on their knowledge of language-cue associations for the otherwise difficult English [ɫ]. It remains unclear, however, why English [ɫ] would have been more difficult than Spanish [l], or easier than English [ʉ], to begin with. The Spanish [l] was not subject to these block effects, so the questions of differences between English and Spanish phonetic cues persist after Experiments 1 and 2.

Forcing listeners, especially the Early Bilinguals, to make responses more slowly did not lead to more similar accuracy rates between Early and Late Bilinguals, e.g., in the Slow block. At the same time, making listeners like the Late Bilinguals respond more quickly also did not make the bilingual groups' responses more similar. The results of this task do not provide evidence that the differences in accuracy rates of Early and Late Bilinguals in Experiment 1 were related to their RTs; instead, there may be fundamental differences in the groups' access to language-specific phonological representations. These differences are likely related to the age at which the bilinguals acquired their L2 or to factors closely associated with L2 acquisition, namely the recency of acquisition and how explicit their knowledge of the L2 is, given how it was acquired. Furthermore, the analysis of the trials in which listeners were unable to make a response during the designated response window reveals additional differences between the language activation of Early and Late Bilinguals.

Missed trials and experimental design

The analysis of missed trials – trials in which the listener failed to respond during the 400ms response window – reveals several shortcomings of the design used in Experiment 2 but also indicates significant patterns and differences across listener groups.

A first disadvantage of the design is that listener responses via the button box were only recorded during the 400ms response window, so there is no record of the language categorization choice made at other points during the trial, for example, while the stimulus played or during the interval of silence, or if actually no response was made at all. This structure makes it impossible to analyze the accuracy of the missed trials, and whether listeners struggled to respond quickly enough (i.e., they responded *after* the 400ms response window) or if they could not achieve the appropriate timing of waiting until the tone played (i.e., they responded *before* the response window). An additional complicating factor is that only one error message was used, no matter the error. The "TOO SLOW!!" screen appeared if no response was registered during the 400ms response window, but if listeners responded too quickly – during the interval of silence – and then failed to re-enter their response after the tone, the same message appeared. However, in these cases, the listeners were actually too fast, and had not waited for the tone, so seeing the "TOO SLOW!!" warning may then have been confusing for listeners who responded during the silence. Since there were block-related differences in the categorization of some segments, it is evident that listeners were on-task and focused during the experiment, so the lost trials may have included information and categorization decisions that would have strengthened the effects reported here. This may be especially true for the bilingual groups, the Early Bilinguals in particular, who contributed the most missed trials. In the future, it would be more valuable to record all responses from all phases of the task and to provide more specific feedback on trials when a response was not entered during the response window.

Despite these limitations on the data provided in Experiment 2, it is unlikely that the large number of missed trials were motivated by misleading feedback alone. After completing the Self-Paced block, in which there were no restrictions on or mentions of time, participants reviewed the instructions about waiting for the tone ("As soon as you hear the tone, respond immediately…Wait for the tone to respond") six times: once at the beginning of the second block, once between the second block's practice and test trials, once at the beginning of the third block, once between the third block's practice and test trials, once at the beginning of the fourth block, and once again between the fourth block's practice and test. Furthermore, listeners completed 10 practice items at the start of each block and then over 100 trials at a given wait length (e.g. 698ms in the Medium block). Some of the listeners with missed trials failed to respond during the response window on as many as 100% of trials. It is unclear why after repeated instructions, practice, and experience some listeners were unable to complete the task at the necessary rhythm, especially since negative feedback was given after every missed trial. Even if the feedback was misleading (i.e., it indicated the response was too slow when in fact the participant had pressed a button too soon), it would be surprising if listeners didn't attempt to vary their pace or otherwise experiment with their response times to avoid the feedback. The lack of adaptation to the response window timing is especially unexpected since the tone played after the interval of silence, even if the listener had pressed the button during the stimulus word or silence interval. If a listener consistently responded during the interval of silence they would have had to have heard, and ignored, the tone in trial after trial. It is possible that some listeners simply misunderstood the task, ignored the tone entirely, and responded as quickly as possible, across all three timed blocks. It may also be the case that receiving the "TOO SLOW!!" warning triggered increased monitoring in the listeners, such that they might have been further slowed by the negative feedback in an effort to be both accurate in timely in the following trial. Given that the structure of Experiment 2 limited the record of responses to those made during the 400ms response window, the RT data does not allow for the testing of this possibility by comparing RTs after missed trials to RTs after completed trials. If increased monitoring after misses results in additional missed trials by further slowing participant responses, responses after missed trials may occur *after* the response window, while misses after completed trials may be more likely to occur *before* the response window, during the silent interval.

These possibilities likely explain part of the variation in the distribution of missing responses across the timed blocks, but there were also significant differences in missed trials across each of the three blocks and the listener groups. As would be expected, more missed trials occurred in the Fast block, which had the shortest wait length at 404ms, and the Slow block (1978ms) had the fewest missed trials. Interestingly, the design of the study was not equally confusing for all listeners: Monolinguals missed very few trials, while the bilingual groups – especially the Early Bilinguals – had significantly more trouble registering a response during the 400ms window. It is unlikely that one group, e.g., the Monolinguals, was more familiar with this kind of experimental design or the instructions provided, since both the Monolinguals and the Early Bilinguals were educated entirely in the U.S. and can be expected to have the same exposure to psychological studies and testing paradigms in schools. Difficulty responding on time instead seems to be related to bilingualism. This theory was tested by including data from other early bilinguals who grew up in the U.S. and spoke a language other than English or Spanish at home with their parents. The Other Bilinguals also failed to register a response significantly more often than the Monolinguals. Importantly, these groups have comparable exposure to Spanish: they have encountered it in their community and in the media, and maybe they have studied it in school, but none were proficient in it or had studied it in college. The fact that even the Other Bilinguals registered more missed trials than the Monolinguals suggests that

Spanish proficiency was not the sole cause of the effect and that bilinguals of any language background may respond more slowly to a language categorization task: accessing language-specific phonological knowledge may be more difficult when multiple linguistic representations exist and one language must be suppressed (cf. Kroll & Stewart, 1994), and this may be especially true in a task that required the activation of both languages for the Spanish-English bilinguals.

Speaking Spanish affected responses beyond bilingualism more generally. Even though the Early Bilinguals and Other Bilinguals share country of birth, age of English acquisition, and language of education, the Other Bilinguals registered significantly fewer missed trials than the Early Bilinguals. Listeners who grew up speaking two languages had more difficulty with the task than Monolinguals, but listeners who grew up speaking the two languages under study had significantly more missed trials due to having specific phonological representations for the languages accessed in the task. Furthermore, Early Bilinguals had even more trouble responding during the designated window than Late Bilinguals, who also speak English and Spanish. The additional difficulty for Early Bilingual listeners suggests that their languages were more activated, or the competition between them was greater, than for the Late Bilinguals. If Gollan et al.'s (2005, 2011) frequency-lag hypothesis can be extended to accessing language-specific phonological information, then the greater likelihood of missing the response window is due to the weaker links between each language and their respective segments for Early Bilingual listeners.

A final note is merited regarding the last outstanding comparison among listener groups, that between Monolinguals and Late Bilinguals. In Experiment 1, the Late Bilinguals also responded more slowly and more accurately than the Monolinguals, which suggested that the Late Bilinguals were more sensitive to the English-ness of the English segments than the Monolinguals but that they also needed more time to access this link between language and segment. Here in Experiment 2, the Late Bilinguals were significantly more likely to miss registering a response than the Monolinguals, so even though their awareness of English segments may be greater than the Monolinguals', Late Bilinguals take longer to access their English representations, accumulate sufficient information to categorize the stimulus as English, and then respond.

Three conclusions emerge by unifying the results of the missed trials analysis with the listener groups' performance in Experiments 1 and 2. First, Early and Late Bilinguals take longer than Monolinguals to access language-specific representations and then accumulate enough information to make an accurate categorization decision, which leads to longer RTs in the Self-Paced blocks of Experiments 1 and 2 and which means that both groups of Bilinguals are less able to reach this threshold for a decision on time in the timed blocks. Second, the nature of the language-segment associations themselves differ across the two Bilingual groups, with the information accumulated by the Early Bilinguals through their weaker language-segment links being less accurate or less language-specific than the information accessed by the Late Bilinguals. Finally, the difference in the nature of the Early and Late Bilinguals' representations can be explained by differences in when and how the groups acquired English.

Future work can more adequately address the difference between the groups by considering measures of the bilinguals' language dominance, codeswitching, and continued language use, since any combination of these and other language proficiency factors may be the source of the variation in responding on time and the strength of language-segment associations. If the Early Bilinguals are more balanced in language dominance than the Late Bilinguals, both languages may have been highly activated and harder to suppress to make a response than for listeners who have a clearly dominant language. Many of the Early Bilinguals grew up in communities in Texas for which codeswitching between Spanish and English is frequent, and so for these listeners both languages may be more activated at any given time than for listeners who more typically switch between English and Spanish contexts (Grosjean, 1989). The frequency-lag hypothesis would predict that links with each language would then be weaker for codeswitchers than for listeners whose language use is more imbalanced. With any of these possibilities, the strength of the link between Spanish and English representations and the relative activation of these representations varies between the two bilingual groups. These groups were chosen for having acquired English early or late, but the reason for the number of missed trials may in fact be due to factors correlated with, but perhaps not the same as, age of acquisition, and settling the cause of the difference between the difficulty experienced by Early and Late Bilinguals is left to future studies.

Conclusion

The accuracy patterns of listeners in Experiment 2 did not reflect significant differences according to how much time was allotted to make a categorization decision, but the categorization accuracy of many language-specific segments was susceptible to changes in response times. The results of Experiment 1 indicated a possible speed-accuracy trade-off for the Late Bilinguals over the Monolinguals and Early Bilinguals, but Experiment 2 indicates the Late Bilinguals' increased accuracy over the Early Bilinguals was not modulated by the longer and shorter response windows of the timed blocks. The lack of change in bilinguals' performance in this timed task confirms that the Early and Late Bilinguals have different degrees of access to phonological information about each of their languages: the Late Bilinguals are better able to distinguish Spanish and English phones when they are presented in context. This sensitivity may be due to having acquired their L2 more recently than the Early Bilinguals and having been taught about the L2 more explicitly than the Early listeners, and this finding may also represent an extension of the frequency-lag hypothesis to the links between languages and phonological representations. Among the differences that emerged across the language-specific target sounds, the English segments were particularly sensitive to changes in accuracy across the blocks, which benefited English /θ/ in all blocks and English [ɫ] in the longest block, while additional response time did not affect the accuracy of English [ʉ]. There were also dramatic effects of bilingualism in the ability of listeners to respond on time in the timed blocks at all, and the struggle to register a response during the narrow, timed window seems to be especially great for bilinguals who speak English and Spanish. This finding further supports the claim that there are fundamental differences in how English and Spanish are related, accessed, and activated in the minds of a bilingual.

6. Experiment 3: Children's perception of language-specific acoustic cues

6.1 Introduction

Experiment 3 explores the language categorization patterns of six-year-olds to test whether children show the same sensitivities to language-specific segmental cues as adults. In Experiment 1, Early and Late Bilingual adults differed in their accuracy and response time: Monolingual and Early listeners were less accurate than Late Bilinguals, even though both bilingual groups spoke English and Spanish, the target languages. Experiment 3 continues to probe the similarities in the performance of Monolinguals and Early Bilinguals by looking at the categorization patterns of these groups at an earlier stage of development. Will Monolingual and Early Bilingual children make the same language categorization decisions as each other, as was observed for the adults? Or will the Early Bilingual children's accuracy rates look more like the Late Bilingual adults', whose exposure to English was more recent and more formal, via foreign-language classes? Will the categorization patterns of Monolingual children reflect the same sensitivities as the Monolingual adults, since their language profile does not change between childhood and adulthood? Or will the results of Experiment 3 indicate that adults are more sensitive to language-cue associations than children, who are still learning the distributions of phonological properties in their language(s)?

Six-year-olds were selected for this task because previous work has shown they are able to differentiate their own local dialect from foreign-accented speech but are unable to reliably distinguish regional dialects from their own dialect (Floccia et al., 2009; Girard et al., 2008). This difference between regional and foreign accents indicates that children are sensitive to the language-specific patterns that set the cues present in foreign-accented speech apart from typical productions in their own dialect but are not yet able to use the acoustic differences between regional varieties. Testing this age group's ability to categorize language-specific phonemic and phonetic segments provides insights about the kinds of acoustic cues children are most sensitive to and how these cues compare to adult patterns. These findings also shed light on the results of the studies by Floccia and colleagues on differences in the perceptions of regional and foreign accents.

Listeners in Experiment 3 completed a version of the Experiment 1 task adapted for six-year-olds. Of particular interest was whether children would be sensitive to the same language-specific cues as the adults – in particular, would they struggle with English /θ/ and English [ʉ], as their adult counterparts did – and whether the monolingual children and the children who would grow up to be the Early Bilinguals would show the same accuracy patterns. Experiment 3 also compared listeners from a third language background, children from English-speaking families in a Spanish-English dual-language immersion program, who have considerable exposure to Spanish but are not (yet) proficient. Experiment 3 seeks to answer the following research questions, and the following hypotheses are tested:

3. Are six-year-olds able to use phonemic and phonetic segments to distinguish Spanish and English, and does language experience affect sensitivity to the cues?
    a. Children are expected to be less accurate than adults in the use of language-specific segments to make categorization decisions since their linguistic representations are still developing. However, like adults, children are expected to be more sensitive to phonemic cues than to phonetic cues.
    b. Children learning English and Spanish are expected to be more accurate than monolingual children due to having developing representations of both languages to compare.

4. How does language experience from childhood to adulthood change listeners' sensitivity to language-specific segments?
    a. The monolingual children are expected to have comparable accuracy to the monolingual adults, who they grow into, since both groups continue to have the representations of one language on which to base their judgments.
    b. The children who speak Spanish at home are expected to have comparable accuracy to the early bilingual adults, since these children will have the same linguistic profile as the early bilingual adults when they grow up.

6.2 METHODOLOGY

6.2.1 Participants

Monolingual and bilingual five- to seven-year-olds participated in this study. Monolinguals were recruited from a list of Austin families who had volunteered to participate in language studies, and bilinguals were recruited from the same volunteer list and from Wells Branch Elementary School in Round Rock. The parents of all children completed a language background questionnaire to gather information about the language(s) spoken at home and any other exposure the children may have to foreign languages, especially to Spanish. See Table 6.1 for a summary of participant characteristics. All children were awarded a book for their participation.

Table 6.1: Demographic information and language background for participants in Experiment 3.

	Monolinguals	Spanish at School only (SpanSchool)	Spanish at Home (SpanHome)
N	24	11	13
mean age	5;10	6;3	6;3
age range	5;0-7;10	5;6-7;3	5;6-7;6
females	8	5	4

The Monolingual children (n=24; 8 females) were from English-speaking families in which no other language was spoken, and their exposure to Spanish was minimal (e.g. from watching television programs like *Dora the Explorer*). All children in the Monolingual group were born and raised in the U.S. The mean age of the Monolingual children was 5;10, and they ranged in age from 5;0 to 7;10.

All children recruited from Wells Branch Elementary were enrolled in the school's dual-language program. Children who hear only Spanish at home are automatically enrolled in the program, and English-speaking families may elect to enroll their child. Because of these policies, the dual-language classes include children who hear only Spanish at home and may also include children who hear only English at home, as well as those who hear both languages. Children from Spanish-speaking families may enroll in the dual-language program whenever they enroll at the school, but children from English-speaking families must enroll in kindergarten; therefore, English-speaking children in first grade have necessarily already completed kindergarten in the dual-language system. In kindergarten and first grade dual-language classrooms, children are taught language arts in their native language (English or Spanish), content courses are divided between the languages, and the "language of the day" alternates between English and Spanish. The language background questionnaire provided additional information about the language(s) spoken by the children's parents at home, and this information was used to divide the dual-language children into English-speaking children whose Spanish exposure was limited to the school setting (SpanSchool; n=11; 5 females) and Spanish-speaking children who heard Spanish at school and also from at least one parent at home (SpanHome; n=13; 4 females). The mean age of the SpanSchool children was 6;3, ranging from 5;6 to 7;3. All children in the SpanSchool group heard only English at home. The mean age of the SpanHome group was 6;3, and their ages ranged from 5;6 to 7;6. All children in the SpanHome group heard Spanish from one or both of their parents. Seven of the 13 SpanHome children heard only Spanish at home, and the remaining six heard a mix: one heard both Spanish and English from the mother and only Spanish from the father, one heard both from the mother and only English from the father, one heard only Spanish from the mother and both from the father, two children heard both languages from both parents (three of the four parents were native speakers of Spanish who learned English as adults), and one child heard both languages from an aunt who lived with the family.

6.2.2 Procedure

Children were tested in a quiet room either in the UT Child Language Lab on the campus of UT Austin or at Wells Branch Elementary School in Round Rock, Texas. Children tested in the Child Language Lab were accompanied by a parent, and a researcher discussed with the parent the procedure of the study and the parent gave written informed consent for their child to participate. After consent was obtained, the parent completed a language history questionnaire describing their child's exposure to foreign languages. Then the experiment began in an adjoining room. Parents interested in watching their child complete the study were invited to watch from a room in the lab that broadcast the live recording from the testing room.

Parents of kindergarten and first grade students enrolled in the Spanish-English dual-language program at Wells Branch Elementary School received a packet of information describing the study, two copies of a consent form, and a language history questionnaire. Children whose parents returned a signed copy of the consent form and who completed the language history questionnaire were eligible to participate. Children were tested individually in a room in the school's office suite. Since office noise, class bells, and meetings were audible from the testing room at the school, the children wore headphones while completing the experiment.

Figure 6.1: Screen images presented to children in Experiment 3.

The experiment was presented on a touchscreen computer running Matlab R2013a (v8.1.0.604), and it included three phases: a familiarization phase, a practice phase, and the test phase. During familiarization, children were introduced to pictures of two adult male faces; one appeared on the left of the screen and the other on the right. The two "boys" in the pictures were early Spanish-English bilinguals from Texas who had participated in an unrelated study. Children were asked to touch each boy's picture to ensure they understood how the touchscreen worked. A successful touch triggered positive feedback, cheering and flashing stars; see Figure 6.1. On the left is the image of the two boys who children selected as having produced each stimulus. On the right is a still image of the positive feedback display: stars on a flashing yellow and blue background, which was accompanied by cheering. In the few cases when children touched the pictures too lightly or too quickly, they were asked to touch the picture again. After children successfully triggered the positive feedback for each picture, children were told they would hear each boy speak. First the picture on the right disappeared, leaving only the boy on the left, and a 10-second clip of the *The North Wind and the Sun* audio recording played in either English or Spanish for 10 seconds. See Section 3.3 for more information about the recording. Then the picture of the boy on the left disappeared and the boy on the right reappeared, and the other language recording played. The location of each picture was fixed (i.e. the picture on the left was always the same), and the boy on the left was always presented first, but the language associated with each picture was counterbalanced across participants. Perhaps surprisingly, no child protested or pointed out that the voice used for each boy was in fact the same.

After the familiarization with the touchscreen and the boys' languages, the practice phase began. Children were told that they would hear a word and would have to decide which boy said it. The design of the familiarization phase and the wording of the task itself ensured that children could complete essentially the same task as the adults but without having to refer to "language" or English and Spanish by name (cf. Akhtar et al., 2012). The participants heard six real words, three in English and three in Spanish, presented randomly, and were asked to indicate which boy said the word by tapping the appropriate image on the touch screen. To motivate children to complete additional trials, feedback was given after each response, with the cheers and stars for positive feedback and a blank, black screen accompanied by a monotone bell for negative feedback. After completing the six real-word trials, a "2" flashed on the screen indicating that the child had made it to "round 2," the test phase. In the test phase, participants heard the same 56 novel words from Experiment 1, 32 English and 24 Spanish, presented randomly, and they categorized these words as they did in the practice phase. Accuracy and reaction time were recorded, and the experiment was video recorded as well.

6.3 RESULTS

Language categorization decisions and reaction times were automatically recorded by Matlab. Decisions were coded as accurate if words containing the English-specific phoneme /ɹ/ or /θ/ or the English variant of [ɫ] or [ʉ] were classified as English and if words with the Spanish-specific phoneme /r/ or the Spanish variant of [l] or [u] were classified as Spanish. Trials with the Spanish stimulus *racha* /ratʃa/ and the English stimulus /ɹatʃə/ were excluded from the analysis (cf. Section 3.2 and footnote 5). RTs were calculated by subtracting the length of the stimulus .wav file from the time calculated by Matlab, which measured the difference between trial onset and picture tap. This ensured that the RTs analyzed here reflected the length of time for the child to make a categorization decision, after hearing the end of the stimulus word. RTs were log-transformed from milliseconds to normalize the distribution of responses for the regression analyses. The videos of each participant were coded by an undergraduate research assistant who marked problematic trials: trials were discarded if the child touched a picture before the stimulus had been presented, thus triggering feedback (n=62), or if the stimulus was not heard because the child or experimenter was talking or because other noise (e.g. from a sibling) interfered with the stimulus (n=47). A total of 109 trials (4.2%) were removed for one of these reasons. After this filtering, an additional four trials had RTs less than 300ms (n=4; 0.2%) and were discarded as spurious. RTs longer than 8s (n=118; 4.6%) were also discarded for not reflecting a child's perception of the stimulus. In total, 231 trials were discarded as problematic or outliers, 8.9% of the 2592 trials. The remaining 2364 trials (Monolinguals: 1195; SpanSchool: 539; SpanHome: 627) are analyzed here. Accuracy (correct, incorrect) and log-transformed RT were tested in separate regression analyses, which were analyzed using Bayesian inference with the *glmer2stan* package (v0.995) in R (v3.2.2) to interface with STAN via RStan (v2.8.2).

Table 6.2: Mean accuracy of each listener group for each stimulus type in Experiment 3.

		Monolinguals	SpanSchool	SpanHome
Phonemic Cues	Spanish /r/	72.9 (44.6)	74.3 (44.0)	87.7 (33.1)
	English /ɹ/	75.8 (43.0)	88.4 (32.3)	91.3 (28.4)
	English /θ/	49.2 (50.1)	52.5 (50.3)	66.3 (47.5)
Phonetic Cues	Spanish [l]	64.4 (48.0)	77.8 (41.8)	83.5 (37.3)
	English [ɫ]	58.1 (49.5)	60.5 (49.2)	85.1 (35.8)
	Spanish [u]	66.7 (47.3)	69.9 (46.2)	81.3 (39.2)
	English [ʉ]	43.2 (50.0)	51.3 (50.3)	63.2 (48.5)

6.3.1 Accuracy analysis

Mean accuracy rates and standard deviations for each stimulus type for each listener group are reported in Table 6.2. Children's accuracy was analyzed using a Bayesian mixed effects logistic regression model with listener language group (three levels: Monolingual, SpanSchool, SpanHome) and target segment (seven levels: /r, ɹ, θ/ and [l, u, ɫ, ʉ]) as fixed effects and participant and stimulus word as random intercepts. The models were fit via a Markov Chain Monte Carlo procedure using STAN (Gelman et al., 2015). Model comparison was performed using the Deviance Information Criterion (DIC; Spiegelhalter et al., 2002). A model with an interaction between the two fixed effects did not provide an improved fit over the model with the two main effects alone. See Table 6.3 for the model summary. The reference group, reflected in the model intercept, represents the accuracy of Monolinguals categorizing stimuli with the Spanish phonemic cue /r/. The fitted log odds of accuracy for each phonemic target segment and listener group are plotted in Figure 6.2 and for each phonetic target segment and listener group in Figure 6.3. The error bars represent the 95% Bayesian credible intervals.

Table 6.3: Summary of mixed effects logistic regression model fitting accuracy results in Experiment 3.

Predictor	*Mean*	*Posterior SD*	*95% CI*	*p value*
Intercept (Monolingual, Spanish /r/)	0.997	0.242	(0.512, 1.467)	<0.0001
English /ɹ/	0.386	0.319	(-0.229, 1.011)	<0.10
English /θ/	-1.139	0.303	(-1.722, -0.542)	<0.0001
Spanish [l]	-0.270	0.300	(-0.867, 0.320)	NS
English [ɫ]	-0.607	0.305	(-1.196, -0.007)	<0.01
Spanish [u]	-0.335	0.295	(-0.918, 0.245)	<0.10
English [ʉ]	-1.311	0.297	(-1.906, -0.737)	<0.0001
Spanish at School	0.310	0.203	(-0.083, 0.706)	=0.10
Spanish at Home	1.019	0.205	(0.622, 1.427)	<0.0001

Random effects	*Variance*
Listener	0.445
Stimulus Word	0.443

Phonemic cues

Overall, children most accurately categorized nonce words with Spanish /r/ and English /ɹ/ and struggled to categorize English /θ/. The lack of significant interaction indicates that the effect of segment is the same across listener groups. Given the structure of the model output, the betas, posterior standard deviations, and *p*-values reported below are associated with the Monolingual children, but the effects do not differ in the other groups. There was a trend for words with English /ɹ/ to be categorized more accurately than words with Spanish /r/ (β=0.386, posterior SD=0.319, p<0.10), and categorization of both English /ɹ/ and Spanish /r/ was significantly more accurate than of English /θ/ (vs. English /ɹ/: β=1.526, posterior SD=0.309, p<0.0001; vs. Spanish /r/: β=1.139, posterior SD=0.303, p<0.0001).

Figure 6.2: Log odds of accuracy for phonemic cues in Experiment 3.

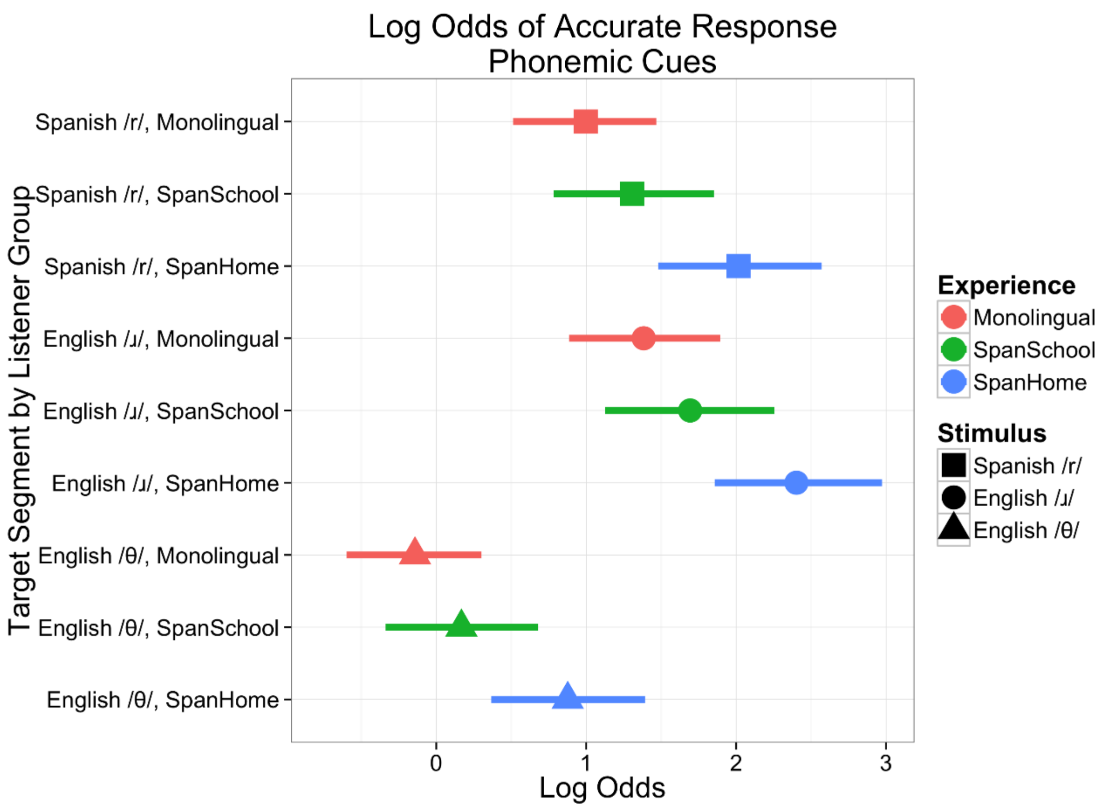
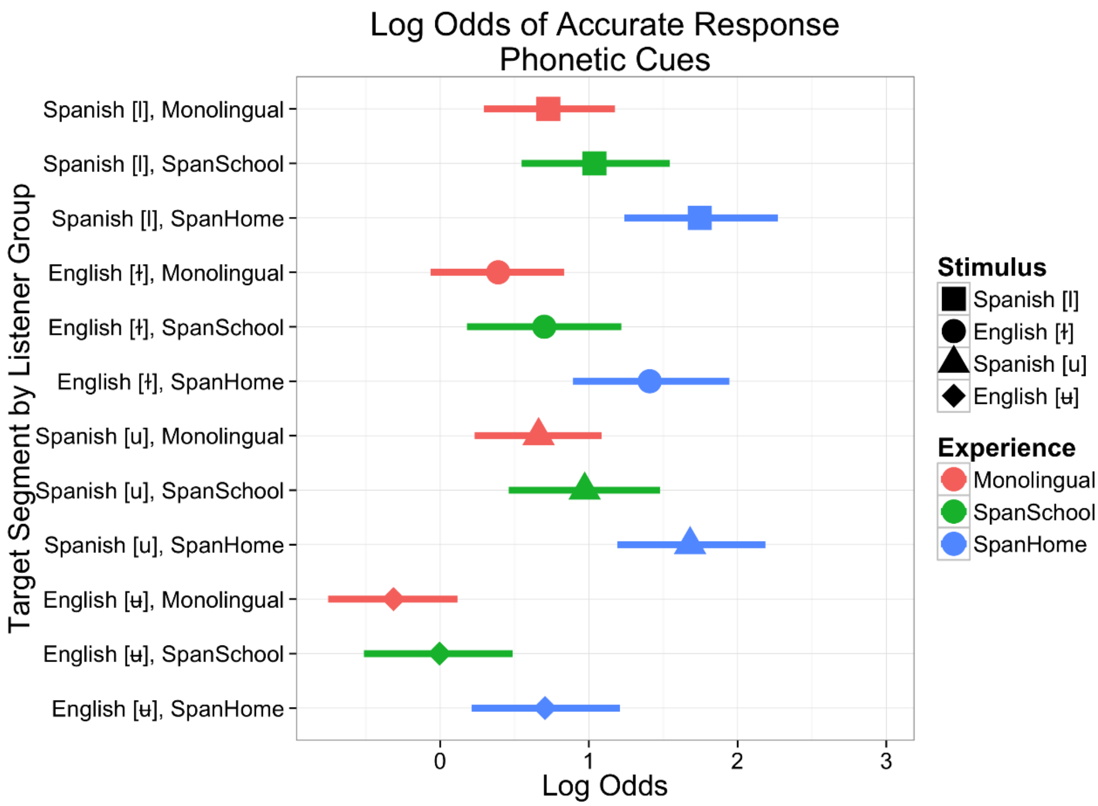


Figure 6.3: Log odds of accuracy for phonetic cues in Experiment 3.



Phonetic cues

Regarding categorization performance for phonetic cues, the language-specific variants of /l/ and the Spanish [u] were categorized accurately at very similar rates, while children were unable to reliably categorize English [ʉ]. As above for the phonemic cues, the statistics reported here reflect the differences in the Monolingual group but are representative of the other groups. There was a trend for Spanish [l] to be more accurately categorized than English [ɫ] (β=0.337, posterior SD=0.286, p<0.10), while the Spanish [u] was categorized significantly more accurately than English [ʉ] (β=0.976, posterior SD=0.277, p<0.0001). There was no difference in accuracy between the Spanish variants [l] and [u] (β=0.065, posterior SD=0.279, p=0.39), but the difference between the English phonetic segments was significant, with [ł] categorized more accurately than [ʉ] ($\beta=0.703$, posterior SD=0.279, $p<0.001$).

Listener language background

The SpanSchool children trended towards better accuracy than the Monolinguals, and the SpanHome group was more accurate than their peers. Here, due to the lack of significant interaction, the betas, posterior standard deviations, and *p*-values reported below are for responses to the Spanish phoneme /r/, but again, the effect of listener group does not differ across the segments. The SpanSchool listeners were marginally more accurate than Monolingual children ($\beta=0.310$, posterior SD=0.203, $p=0.10$), and the children who heard Spanish at home were significantly more accurate than the Monolinguals ($\beta=1.019$, posterior SD=0.205, $p<0.0001$) and the SpanSchool group ($\beta=0.709$, posterior SD=0.242, $p<0.01$).

Tables 6.4 and 6.5 summarize the accuracy analyses for the target segments and listener groups, respectively. As in previous chapters, the “=” indicates differences that were not significant, the “>” and “<” are used for significant differences, and “»” and “«” represent differences that approached significance.

Table 6.4: Summary of accuracy results for target segment comparisons in Experiment 3.

(A) Phonemic Cues

	Target Segments
Accuracy	English /ɹ/ » Spanish /r/ > English /θ/

(B) Phonetic Cues

	Cross-language	Cross-segment
Accuracy	Spanish [l] » English [ɫ] Spanish [u] > English [ʉ]	Spanish [l] = Spanish [u] English [ɫ] > English [ʉ]

Table 6.5: Summary of accuracy results for listener group comparisons in Experiment 3.

	Listener groups
Accuracy	Monolinguals « SpanSchool < SpanHome

6.3.2 Reaction time analysis

Mean RTs and standard deviations for each stimulus type and listener group are reported in Table 6.6. Log-transformed RTs were analyzed using a Bayesian mixed effects linear regression model with listener language group (three levels: Monolingual, SpanHome, SpanSchool), target segment (seven levels: /r, ɹ, θ/ and [l, u, ɫ, ʉ]), and accuracy (correct, incorrect) as fixed effects and participant and stimulus word as random intercepts. These models were also fit via a Markov Chain Monte Carlo procedure using STAN, as described above. A model with the three main effects and no interactions was selected as the best fit for the data using DIC. See Table 6.7 for the summary of the model. The reference group, reflected in the model intercept, represents the log RT of inaccurate responses by Monolinguals categorizing stimuli with the Spanish phonemic cue /r/. The fitted log RT for each phonemic target segment and listener language group are plotted in Figure 6.4 and for each phonetic target segment and listener language group in Figure 6.5. For clarity of presentation given the number of conditions, Figures 6.4 and 6.5 present only RTs for accurate responses. The error bars represent the 95% Bayesian credible intervals.

Table 6.6: Mean RT (in milliseconds) in correct (A) and incorrect (B) trials for each listener group and stimulus type in Experiment 3.

(A) Correct trials

		Monolinguals	Early Bilinguals	Late Bilinguals
Phonemic	Spanish /r/	1991.3 (1144.7)	1916.7 (998.6)	2118.6 (1483.6)
	English /ɹ/	1916.3 (1214.2)	1988.7 (1337.5)	1655.3 (1032.6)
	English /θ/	2149.6 (1192.8)	2542.6 (1568.4)	1863.3 (1275.8)
Phonetic Cues	Spanish [l]	2169.7 (1333.7)	2101.6 (1610.8)	1864.9 (1364.7)
	English [ɫ]	2294.4 (1650.8)	2677.7 (2006.1)	2014.4 (1673.4)
	Spanish [u]	2169.9 (1431.6)	2115.3 (1518.5)	1857.4 (1081.8)
	English [ʉ]	2364.8 (1407.8)	2399.6 (1315.9)	2167.6 (1620.7)

(B) Incorrect trials

		Monolinguals	Early Bilinguals	Late Bilinguals
Phonemic Cues	Spanish /r/	2206.3 (1524.2)	2137.2 (1443.0)	1735.6 (1115.0)
	English /ɹ/	1949.1 (1042.8)	2196.7 (1496.8)	2107.0 (1562.3)
	English /θ/	2425.2 (1697.7)	2198.2 (1407.3)	2616.9 (1841.5)
Phonetic Cues	Spanish [l]	2170.9 (1462.9)	2694.1 (1903.1)	1759.0 (1322.9)
	English [ɫ]	2364.1 (1378.4)	1835.7 (818.1)	2752.4 (2223.5)
	Spanish [u]	2281.6 (1330.5)	2237.2 (1248.8)	3088.3 (1838.4)
	English [ʉ]	2142.2 (1229.6)	2445.1 (1466.6)	2703.7 (1734.1)

Table 6.7: Summary of mixed effects linear regression model fitting reaction time in Experiment 3.

Predictor	*Mean*	*Posterior SD*	*95% CI*	*p value*
Intercept (Monolingual, Spanish /r/, incorrect)	7.515	0.089	(7.343, 7.689)	<0.0001
English /ɹ/	-0.081	0.048	(-0.176, 0.013)	NS
English /θ/	0.059	0.048	(-0.038, 0.153)	NS
Spanish [l]	-0.004	0.048	(-0.098, 0.092)	NS
English [ɫ]	0.047	0.047	(-0.048, 0.138)	NS
Spanish [u]	0.042	0.047	(-0.051, 0.132)	NS
English [ʉ]	0.088	0.047	(-0.007, 0.181)	NS
Spanish at School	0.029	0.149	(-0.266, 0.321)	NS
Spanish at Home	0.007	0.140	(-0.265, 0.284)	NS
Response correct	-0.053	0.023	(-0.098, -0.008)	NS

Random effects	*Variance*
Listener	0.398
Stimulus Word	0.048

Phonemic cues

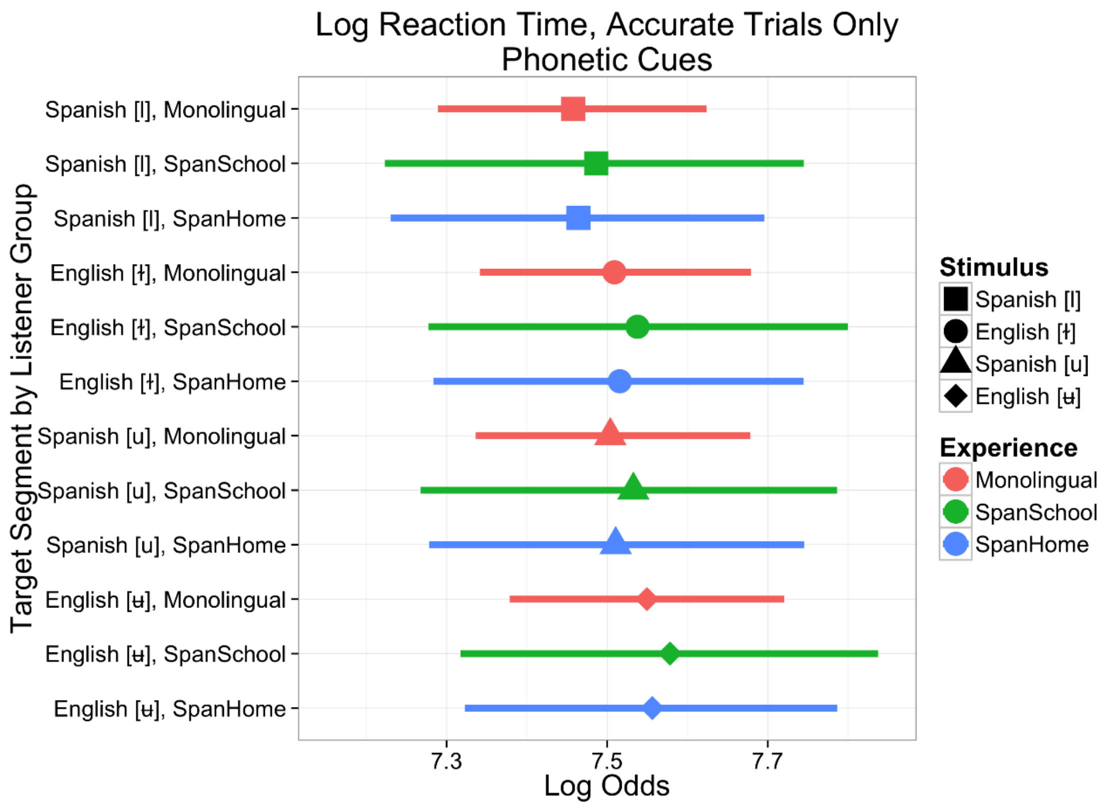
As was done for the accuracy analysis above, the betas, posterior standard deviations, and *p*-values reported here are associated with the Monolingual children, for accurate trials, but the effects do not differ in the other groups. The RTs for nonce words with phonemic cues were similar for all segments, and there was no significant difference between RTs for Spanish /r/ and English /ɹ/ (β=0.081, posterior SD=0.048, *p*=0.18) or Spanish /r/ and English /θ/ (β=0.059, posterior SD=0.048, *p*=0.25). Responses to English /ɹ/ trended towards being significantly faster than to English /θ/ (β=0.140, posterior SD=0.47, *p*<0.10).

Figure 6.4: Log reaction time for phonemic cues in Experiment 3.

Phonetic cues

As above, the statistics reported here reflect the differences in the Monolingual group on accurate trials but are representative of the other groups. The RTs for phonetic cues did not vary by segment. There was no significant difference between the Spanish and English /l/ variants (β=0.051, posterior SD=0.46, *p*=0.28) or /u/ variants (β=0.046, posterior SD=0.044, *p*=0.30). Responses to the Spanish segments [l] and [u] did not differ (β=0.046, posterior SD=0.045, *p*=0.30), and there was also no difference between English [ɫ] and English [ʉ] (β=0.041, posterior SD=0.044, *p*=0.32).

Figure 6.5: Log reaction time for phonetic cues in Experiment 3.



Listener language background

There was no difference in RT between any pair of listener groups (Monolinguals vs. SpanSchool: β=0.029, posterior SD=0.149, *p*=0.37; Monolinguals vs. SpanHome: β=0.007, posterior SD=0.140, *p*=0.47; SpanSchool vs. SpanHome: β=0.022, posterior SD=0.169, *p*=0.42). These statistics reflect accurate responses to the Spanish phoneme /r/, but again, the effect of listener group does not differ across the segments.

Effect of accuracy on reaction time

The difference in RT between accurate and inaccurate trials for Monolinguals in response to nonce words with Spanish /r/ was not significant (β=0.053, posterior

SD=0.023, *p*=0.27), and is representative of the lack of RTs differences in accuracy across the other listener groups for other segments.

The RT analyses are summarized for the target segments in Table 6.8, listener groups in Table 6.9, and trial accuracy in Table 6.10. Non-significant differences are represented with “=”, significant differences with “>” and “<”, and “»” and “«” indicate differences that approached significance.

Table 6.8: Summary of RT results for target segment comparisons in Experiment 3.

(A) Phonemic Cues

	Target Segments
RTs	English /ɹ/ « English /θ/ Spanish /r/ = English /ɹ/, Spanish /r/ = English /θ/

(B) Phonetic Cues

	Cross-language	Cross-segment
RTs	Spanish [l] = English [ɫ] Spanish [u] = English [ʉ]	Spanish [l] = Spanish [u] English [ɫ] = English [ʉ]

Table 6.9: Summary of RT results for listener group comparisons in Experiment 3.

	Listener groups
RTs	Monolinguals = SpanSchool = SpanHome

Table 6.10: Summary of RT results for trial accuracy comparisons in Experiment 3.

	Trial accuracy
RTs	Correct = Incorrect

6.3.3 Variation across nonce words

As was the case in Experiment 1, children's categorization patterns in Experiment 3 represent decisions made for eight stimuli per target segment. To investigate possible nonce-word-specific effects motivated by the filler segments and the different locations of the target segments (word-initial or word-internal), this section describes how well listeners from the three language backgrounds categorized each word in a stimulus group. Means and trends from the raw data are discussed.

6.3.3.1 Phonemic stimuli

The nonce word stimuli with Spanish /r/ and English /ɹ/ were categorized mostly consistently by the SpanHome listeners, but the Monolingual and SpanSchool children were more variable; see Figure 6.6 for Spanish /r/ and Figure 6.7 for English /ɹ/. There is no clear effect of word or where the target segment occurs in the nonce word. The Monolingual and SpanSchool children sometimes performed similarly, as in their responses to *marra*, the most difficult of the Spanish /r/ words, while for other words the Monolinguals outperformed the SpanSchool listeners (e.g. *chirra*), and for still others the pattern was reversed (e.g. *richa* and *chiruh*). Responses to *richa* (Spanish /r/) and *rimuh* (English /ɹ/) were mostly strong, perhaps owing to the target segments occurring word-initially. There is no clear effect of word. However, the other /ɹ/-initial stimulus, *richuh,* was more difficult than *rimuh* for all groups and also more difficult than other nonce words with /ɹ/.

Figure 6.6: Mean accuracy for each Spanish /r/ nonce word, by listener group, in Experiment 3.

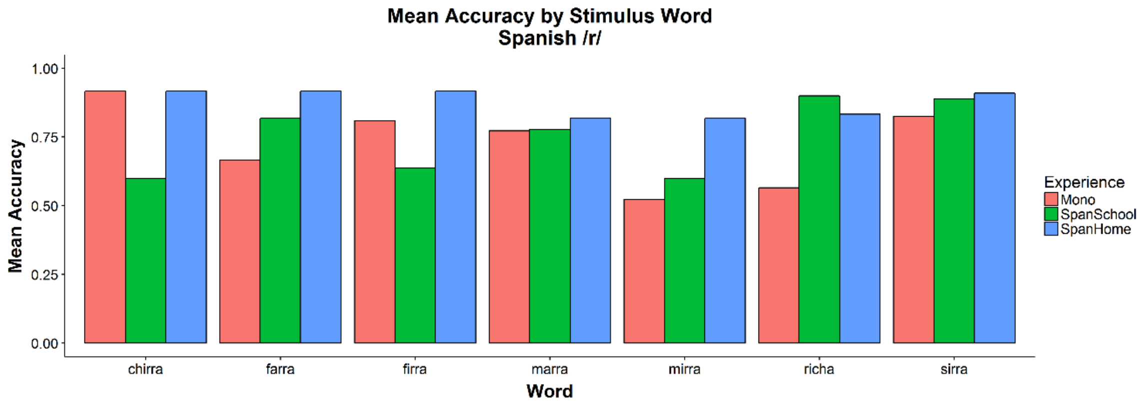
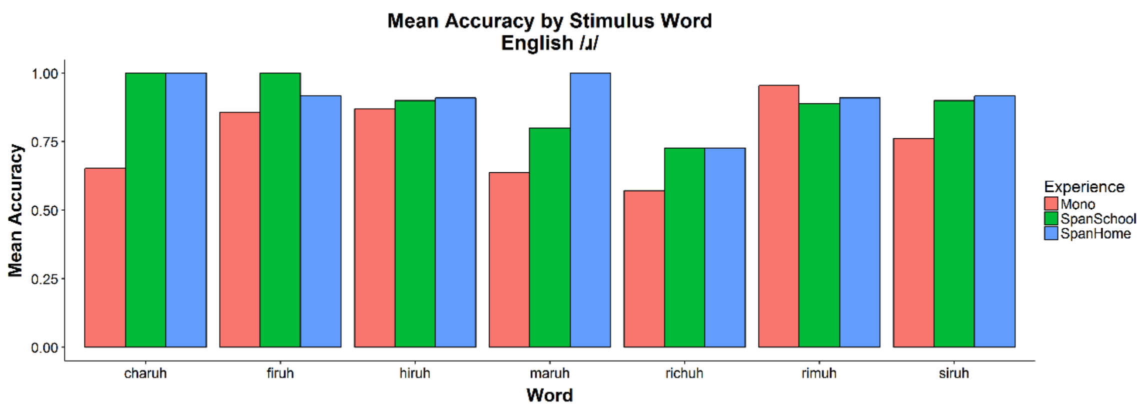
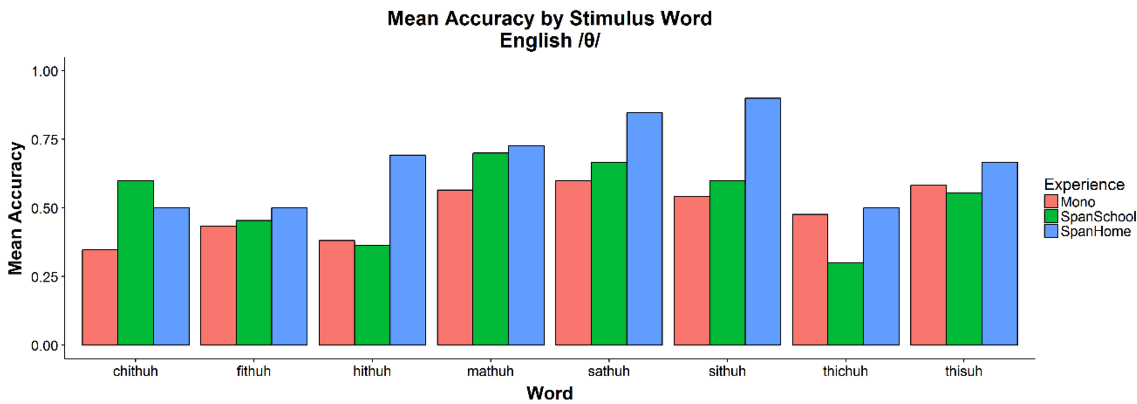


Figure 6.7: Mean accuracy for each English /ɹ/ nonce word, by listener group, in Experiment 3.



For the English /θ/ stimuli, in Figure 6.8, there was again considerable variability in accuracy across the words, as well as lower rates of accuracy overall. All groups seemed to perform worse on the /θ/-initial words *thichuh* and *thisuh*. This is reminiscent of the adult pattern; the adults had difficulty accurately categorizing /θ/-initial words. However, for the children, there were other stimuli to which they responded just as badly, e.g. *chithuh* and *fithuh*. It appears that the lack of salience of /θ/ in word-initial position is not alone responsible for the poor responses among child listeners; rather /θ/ may simply be even more perceptually difficult for children, across the board, than it is for adults.

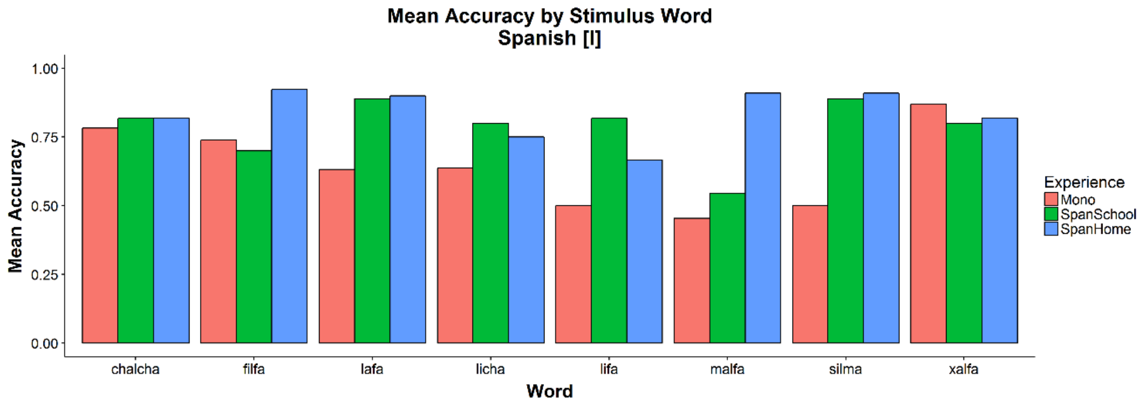
Figure 6.8: Mean accuracy for each English /θ/ nonce word, by listener group, in Experiment 3.



6.3.3.2 Phonetic stimuli

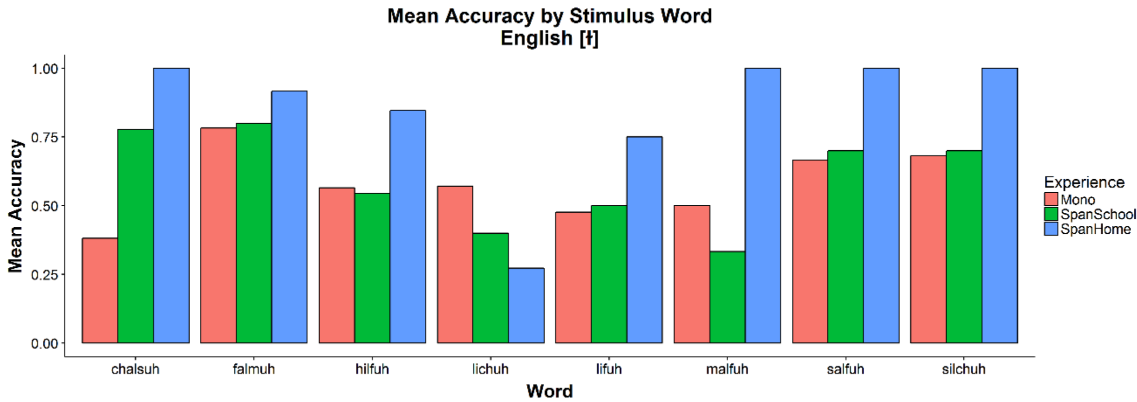
The accuracy of responses to stimuli with Spanish [l] are in Figure 6.9. The SpanSchool and SpanHome groups responded consistently to Spanish [l] stimuli, although SpanSchool struggled with *malfa* more than with other words, patterning instead with the Monolinguals. The Monolinguals struggled with *lifa* and *silma* as well – stimuli that included both [l]-initial and [l]-medial words – so no clear pattern emerges for where the segment occurs.

Figure 6.9: Mean accuracy for each Spanish [l] nonce word, by listener group, in Experiment 3.



For the words with English [ɫ], in Figure 6.10, accuracy rates also vary. *Lichuh, lifuh*, and *malfuh* were among the most difficult words for listeners, especially the Monolingual and SpanSchool children. Again, these more difficult words include stimuli with the target segment word-initially and word-medially, so it is not clear to what extent phonological context influenced responses. Furthermore, some of the "difficult" words for one group, like *chalsuh* and *malfuh* for the Monolinguals, received highly accurate responses from the other groups: SpanSchool and SpanHome categorized *chalsuh* accurately, and the SpanHome group had no problem categorizing *malfuh*.

Figure 6.10: Mean accuracy for each English [ɫ] nonce word, by listener group, in Experiment 3.



In Figure 6.11, the accuracy of responses to words with Spanish [u] is plotted. For most stimuli, the three groups' performances were comparable. However, for the stimulus *fusa*, the three listener groups are quite distinct.

Figure 6.11: Mean accuracy for each Spanish [u] nonce word, by listener group, in Experiment 3.

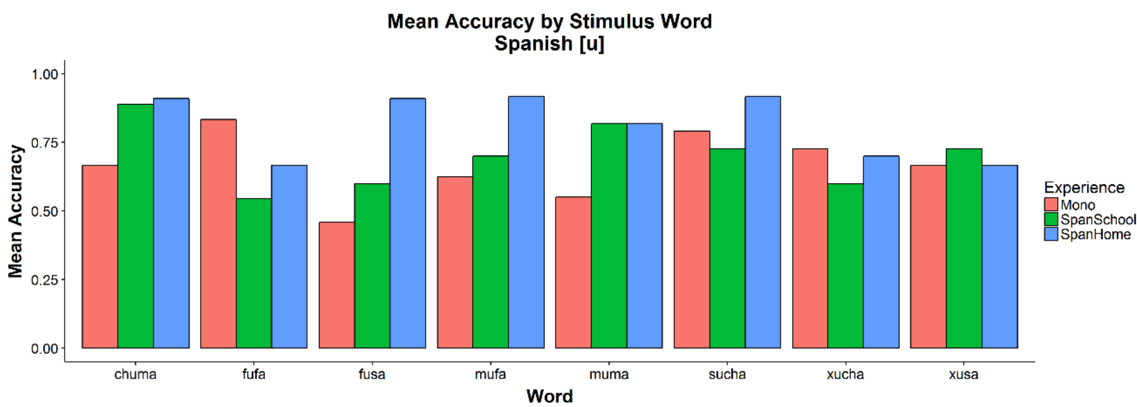
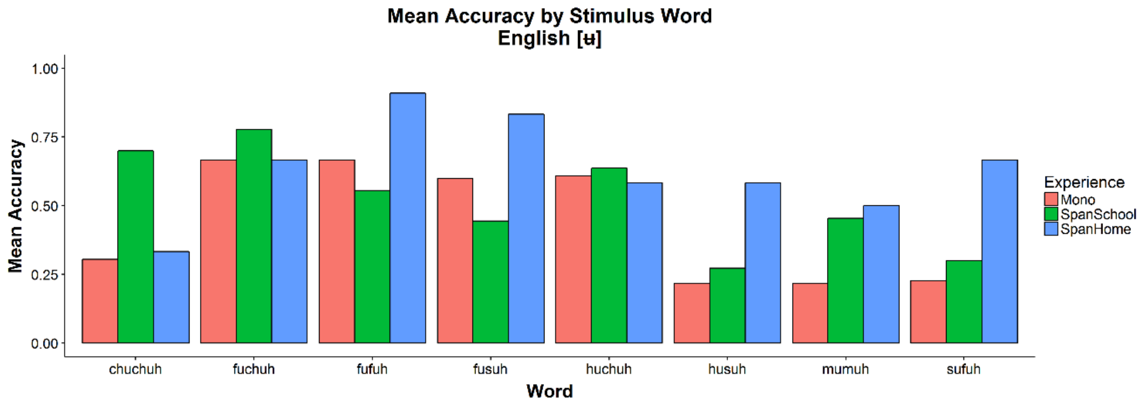


Figure 6.12: Mean accuracy for each English [ʉ] nonce word, by listener group, in Experiment 3.



Monolingual and SpanSchool children responded with lower accuracy across the English [ʉ] stimuli, while the SpanHome listeners varied between very strong and very weak categorization accuracy; see Figure 6.12. At the low end were the accuracy scores in response to *husuh* – 21.7% for Monolinguals and 27.3% for SpanSchool – to which the SpanHome children responded accurately in a moderate 58.3% of trials. The SpanHome group struggled with *chuchuh*, at 33.3% accurate, but soared to 90.9% accuracy for *fufuh.* The stimuli *husuh, mumuh,* and *sufuh* were more difficult for all three groups, but even here the inherent differences among the groups' accuracies are apparent.

Overall, children's responses across the eight words of a target segments were highly variable and none of the trends apparent in the adult analysis in Experiment 1 were evident here. In particular, there does not seem to be an effect of where the target sound is located in the word – word-initially or word-internally – for the more difficult segments, although the beginnings of this pattern may be visible. For example, words that begin with English [ɫ] are categorized more inaccurately than some other English [ɫ] stimuli, but there are also nonce words with English [ɫ] that are categorized as poorly as (or more poorly than) the [ɫ]-initial words. The same is true for the categorization of English /θ/ words.

Between childhood and adulthood, listener sensitivity to difficult segments in some positions (e.g., word-internally) may improve dramatically, thus raising the accuracy rates for some of the segment-internal stimuli, while sensitivity to these sounds word-initially remains low even into adulthood.

6.4 Discussion

In Experiment 3, six-year-olds with varying exposure to English and Spanish categorized nonce words as having been produced by an English speaker or a Spanish speaker. This study was realized to make comparisons between the adult categorization decisions from Experiment 1, reported in Chapter 4, and the sensitivities of children who are still acquiring their language(s) and learning phonological regularities. Children who heard Spanish at home with their families and who were learning English in a dual-language program (SpanHome) performed significantly more accurately than monolingual English-learning children and more accurately than children from English-speaking families who were learning Spanish in the dual-language program (SpanSchool). The difference in accuracy between the two groups of children from English-speaking families – the Monolinguals and SpanSchool listeners – was marginally significant (p=0.10), with the SpanSchool group performing somewhat more accurately than Monolinguals. The three groups categorized the Spanish phoneme /r/ and the English phoneme /ɹ/ accurately and comparably, and both these phonemes were categorized significantly better than the English phoneme /θ/. Responses to language-specific phonetic cues were similar for nonce words containing Spanish [l], English [ɫ], and Spanish [u]. Responses to the English phonetic cue [ʉ] were significantly less accurate than to the Spanish variant [u] or to the other English phonetic segment, [ɫ]. Children's RTs were more variable than adults' overall, probably at least in part because of differences in the task; here children responded to pictures on a touchscreen in a more interactive task than the adults' button box responses. Children's RTs did not differ by accuracy, listener group, or target segment, although there was a trend for responses to English /ɹ/ to be faster than to English /θ/.

Categorization decisions: children and adults

Overall, the language decisions made by children and adults were similar. For both age groups, and all language backgrounds within the age groups, Spanish /r/ and English /ɹ/ were categorized more accurately than English /θ/, which was among the most difficult segments to categorize. Likewise, the English segment [ʉ] was the worst categorized of the phonetic cues for all listener groups among both children and adults. These commonalities among the ages and listener groups suggest that the difficulties listeners experienced categorizing English /θ/ and English [ʉ] are rooted in general ease of perceptibility, independent of language exposure or age of acquisition. See Section 4.4 for additional discussion about the properties of these sounds and proposals for why listeners struggled to categorize them; the bottom line might be that there are developmental and perceptual reasons for the low accuracy rates for stimuli with English /θ/, but an explanation for English [ʉ] is not readily known. One possibility is that there are fundamental differences in how listeners use consonants and vowels in lexical access, although it is unclear if the nonce words in the present study would have triggered the same kind of activation as has been reported for lexical retrieval (Owren & Cardillo, 2006; cf. Kewley-Port et al., 2007). The findings of Experiment 3 add to the case that there is something inherently different about the perception of English [ʉ] in nonce words, and in the absence of evidence like that available for /θ/ in other perception studies, a tentative conclusion is that the perceptual difficulty of /u/ relates to its being the only vocalic segment tested in this project.

Although the two age groups' performance overlapped in the most difficult segments, there were differences between the groups as well. While the adults found Spanish /r/ somewhat easier to categorize than English /ɹ/ – this was the pattern for Monolinguals and Early Bilinguals – English /ɹ/ trended towards higher accuracy than Spanish /r/ for the children. Among the language-specific phonetic cues, there were more distinctions in accuracy for the adults than for the children. The six-year-old listeners essentially made two distinctions: English [ʉ], which they categorized poorly, and the other phonetic cues, which they categorized better. The adults preferred Spanish phonetic segments over English ones, and /l/ variants over /u/ variants, creating three groups: Spanish [l], which was categorized well; English [ɫ] and Spanish [u], which adults categorized reliably although less well than Spanish [l]; and the poorly-categorized English [ʉ]. The fact that Spanish [l] was more salient for adults than for children may indicate that something about adults' additional exposure to this sound strengthened its association with Spanish. Since the Spanish [l] was best categorized among the phonetic cues even among the Monolingual adults, the exposure that influenced responses may include hearing this segment in Spanish-accented English, despite listeners' low exposure to the Spanish language itself.

Children's responses to the eight words within each target segment category were typically quite variable, unlike the adult analysis by stimulus word, so few generalizations can be made in an attempt to understand children's responses based on where each segment occurs in a word. Part of the variation in the children's responses may be due to differences in the adult and child versions of the experiments, in that children responded once to each stimulus word, whereas adults heard each stimulus eight times. One tenuous pattern among the children's responses to English [ɫ] stimuli is that words beginning with English [ɫ] were categorized less accurately than stimuli with the sound medially. This is like the adult pattern, which reflects that a less velarized variant of [ɫ] occurs word initially in English. However, it should also be noted that Monolingual and SpanSchool children's responses to *malfuh*, an [ɫ]-medial word with the more velarized production, were more inaccurate than to the [ɫ]-initial stimuli. Phonological context alone does not explain children's responses to words with English [ɫ].

The similarities, and differences, in the categorization decisions of children and adults suggest that associations between segment and language are largely fixed by the time a child enters elementary school. This may be especially true for phonemic cues. Children at this age have learned to generalize variable productions of segments and words to higher level phonemic and lexical representations (e.g. Nathan et al., 1998; Schmale et al., 2011). In exchange for the efficiency in mapping different productions to common categories, children may also be less able to use phonetic variation that doesn't result in a phonemic change in word-level tasks. Rather than 'ignore' phonetic variation, adult listeners have learned to both generalize over variation and associate social meaning (e.g. region of origin or non-native speaker) with variation. The present study suggests children may be particularly sensitive to phonemic segments. The phonemic/phonetic distinction may in fact be the principle difference between foreign accents and regional accents, and their differential performance with each category could explain why children of this age can recognize the foreign accents but not regional accents (Floccia et al., 2009; Girard et al., 2008). Foreign-accented speech may contain larger deviations from the local dialect, perhaps including more phonemic than phonetic differences. This may also be what sets Indian English apart from a child's local dialect, even though Indian English itself is a regional variety (Wagner et al., 2013). Likewise, regional accents may differ more in sounds that vary along a continuum, like the phonetic cues here: a six-year-old English-speaking child may then only be on the cusp of being able to differentiate the lighter [l] from the velarized [ɫ] for the purposes of speaker classification.

With additional exposure to variation in speech, children must eventually learn to attend to the social meanings of variation as adults do, and they will also strengthen the associations between these phonetic realizations and higher level phonemic and lexical representations. Through a listener's life, this variation in the input will include exposure to a wider range of English productions (e.g., the more and less velarized variants of English /l/), as well as passive exposure to Spanish, minimally through exposure to Spanish-accented speech. Such exposure may be responsible for the increased salience of Spanish phonetic cues for adult listeners. However, some sounds remain inherently more difficult than others, and even the level of exposure that adults have to their native language(s) seems insufficient to substantially raise awareness of the language association of these difficult sounds or increase sensitivity to them.

Differences in language experience across age groups

In addition to categorization differences by segment, variation across listener groups was also of interest. For practical reasons, the listener groups could not be the same at both age points: at six years old, the Late Bilingual adults are Spanish monolinguals living in Latin America. Instead three distinct language exposure profiles were chosen for comparison among child and adult listeners. The focus here was the Monolingual children, who grow into the Monolingual adults, and the SpanHome children, who mature into the Early Bilinguals. The results of Experiment 1 indicate few differences between Monolingual and Early Bilingual adults: there was a trend for Monolinguals to be more accurate than Early Bilinguals in categorizing Spanish [u], and Monolinguals typically responded faster than Early Bilinguals. However, among the Monolingual and SpanHome children – the younger versions of the Monolingual and Early Bilingual adults – the difference between the groups was clear. The SpanHome listeners responded significantly more accurately than the Monolingual children across all segments; that is, the six-year-old children who heard Spanish at home and who were learning English in school were significantly more sensitive to language-segment associations than their Monolingual peers. This is also the difference that was observed between the improved performance of Late Bilingual adults compared to the Monolingual and Early Bilingual adults. But by the time those more accurate Spanish-speaking children are college-aged, they perform statistically the same as the English monolinguals. The change in sensitivity from the SpanHome children to the Early Bilingual adults is crucial to understanding how language experience affects the associations between language and segment – or perhaps, access to these associations. Since adult-like sensitivities to the segmental cues presented in this task are largely in place by six years of age, the question becomes what the change is in these early bilingual listeners' language experience that results in adult performance indistinct from Monolinguals and less like Late Bilinguals.

One possible explanation for the change in performance is the SpanHome group's increased exposure to and proficiency in English over the next 14 or so years. While both the Early and Late Bilingual adults spoke the same pair of languages, their experiences with each language were quite distinct. The Early Bilinguals had been educated entirely in English, had been immersed in an English-speaking or English- and Spanish-speaking society (in the case of listeners from near the Texas-Mexico border) since at least the start of elementary school, and may not have been literate in Spanish. The Late Bilinguals had more recently been immersed in English, most had begun English-language education after puberty, they were literate in both languages, and they had been educated in Spanish in primary and secondary school. There were likely differences in language dominance between the bilingual groups as well, although dominance was not evaluated directly. In this respect, the SpanHome children may be more like the Late Bilingual adults than the Early Bilingual adults in their more recent transition from Spanish-speaking families to the English-language community, within the last year or two with the beginning of elementary school. This change from predominantly Spanish exposure to predominantly English exposure between six years old and college-aged may fundamentally alter the nature of Early Bilinguals' language representations and language-segment associations, at least with respect to Spanish, the first acquired but possibly less dominant language. It may also be the case that the change in community and community language exposure heightened listeners' sensitivity to differences in the English and Spanish speaking communities, which may include the acoustic correlates present in speech.

Alternatively, the main reason for the difference in performance between the Early and Late Bilingual adults may be the language learning process more generally. While the SpanHome children will acquire English and become fluent speakers, like the Early Bilingual adults, Experiment 3 may have tested them during a stage in which their metalinguistic awareness is particularly heightened, in these early years of language acquisition. See Section 4.4 for additional discussion of differences in English exposure and learning among the Early and Late Bilingual adults.

Another possible interpretation of the results of Experiments 1 and 3 is not that the SpanHome children became more like Monolinguals as they aged, but that the Monolingual children's sensitivities improved over time and became more like the SpanHome children, who would have then remained mostly consistent with age. This would also result in the similar accuracy rates observed for the Monolingual and Early Bilingual adult groups. There are minimally two reasons to doubt this explanation. The first is that the Monolingual children's language experience does not change fundamentally between elementary school and college; their exposure to variation in English, including exposure to foreign-accented English, increases, but it is not clear why this would result in improved performance for Monolinguals while the change in language exposure for the Early Bilingual group would then have to be argued to not have changed their accuracy. There is a second reason to think the change between children and adults is unlikely to be due to an improvement in Monolingual performance alone. If Monolinguals' accuracy improved to the SpanHome rates by adulthood, there are now three changes in performance that must be motivated (an improvement in Monolingual accuracy from childhood to adulthood, the better accuracy of SpanHome children than of Monolingual children, and the better accuracy of Late Bilingual adults than other adults), while positing that there was no real change in accuracy between the SpanHome children as they grow into Early Bilingual adults, or at least that the change was so small that the Monolingual listeners made up for this gain into adulthood as well. To explain this scenario, it would also be necessary to conclude that the Early listeners' significant changes in exposure to English influenced their categorization accuracy less than the Monolingual listeners' continued development of English. Instead, a more succinct explanation describes the differences between children's and adults' accuracy in terms of changes in language exposure: the root of the higher accuracy rates of the SpanHome children and the Late Bilingual adults is likely the SpanHome children's increase in exposure to English and the parallels between the Late Bilinguals' current language profile and the language profile of the SpanHome children.

The categorization decisions of the SpanSchool listeners warrant a final note about the influence of language exposure on accuracy. These children from English-speaking families had been enrolled in the dual-language immersion program at their school for seven to eight months, in the case of kindergarteners, or a year and seven to eight months, in the case of first graders, and the same is true of the SpanHome children. However, the SpanSchool children's exposure to Spanish is nearly totally dependent on the consistent use of Spanish by their teachers, whereas the SpanHome children have been exposed to English at least passively in the community and through media since birth. Since most of the dual-language immersion teachers are themselves early bilinguals, there may be important differences in the Spanish input to which the SpanSchool children are exposed from their teachers, and the exposure the SpanHome children get from their mostly late-bilingual parents. For example, there may be differences in the Spanish productions of the early bilingual teachers and the late bilingual parents (e.g., Flege et al., 1997, 1999), and the early bilingual teachers may be more likely to codeswitch or use a child's dominant language since the teachers themselves may be more balanced in Spanish and English than parents who learned English as adults. The acoustic properties and variation in the Spanish input the SpanSchool and SpanHome children receive may have contributed to the SpanSchool group's accuracy being significantly lower than the SpanHome children's accuracy: in order to categorize nonce words significantly more accurately than Monolinguals, the SpanHome children need exposure to Spanish greater than a year or year and a half, or may need more consistent exposure than that available to them when class time is divided between the two languages.

The SpanSchool children also arguably do not understand much of the Spanish they hear – many were unable to match high-frequency labels (e.g. *cachorro* 'puppy') with representative objects in an unrelated task – but their considerable exposure to the language in the classroom led to a marginally significant improvement in accuracy over Monolingual children (p=0.10). Continued consistent exposure to the phonology of Spanish, which may also lead to fluency in the language, would be expected to make these English-speaking listeners even more accurate, even before they became highly proficient in Spanish. Note also that since accuracy rates of the SpanSchool children were trending towards significantly higher than the Monolingual children, what these children have learned of Spanish phonology through exposure to the language itself seems to be greater than what the Monolingual children or adults may be aware of through occasional exposure to Spanish-accented English.

Conclusion

The categorization patterns of children in Experiment 3 demonstrate similarities between children and adults in the salience of the language-specific segmental cues but differences across the listener groups at each age point. Child listeners struggled with the same cues as adults, namely English /θ/ and [ʉ], but showed less sensitivity to distinctions among phonetic segments. Whereas the Monolingual and Early Bilingual adults made statistically indistinct categorization decisions, the Monolingual and SpanHome children were statistically different: the SpanHome children categorized language-specific cues more accurately than the Monolingual children. Many parallels are apparent between the language experience of the SpanHome children at this age and the Late Bilingual adults, even though the SpanHome children grow into the Early Bilinguals. The change in accuracy between the SpanHome children and the Early Bilingual adults likely represents the influence of language exposure and potentially of language dominance on access to language-cue associations and further suggests that the recency of changes in language exposure may heighten listener sensitivity to language-specific cues.

7. General Discussion

This dissertation tested listener sensitivity to a range of language-specific cues in a cross-language speech perception task as a function of listener language exposure. Previous work has investigated how monolinguals and early bilinguals identify and discriminate L2 contrasts that are absent from or differ in the L1. Results indicate that bilingual listeners are quite comparable to monolinguals in their ability to discriminate L2 sounds, especially the earlier the L2 was acquired and the less the L1 is used, although some work has also found that late bilinguals are success in the discrimination tasks as well. However, few tasks require listeners to access language representations, since segments are typically presented in isolation. This approach makes it difficult to generalize the salience of these contrasts from auditory discrimination to linguistic representations, and such an association of language with segment might facilitate lexical access (Flege, 2007) and similar connections are made for each language and lexemes (e.g., Gollan et al., 2005, 2011). More recent work with mispronunciation tasks confirms that listeners may struggle to use discriminable contrasts in context. The broader goal of this dissertation was to more completely evaluate how monolinguals and bilinguals use language-specific cues in context, and in particular three questions were asked: are there differences in listeners' perceptions of the language-specificity of segmental cues, can these differences be modulated by listener language experience, and what can the sensitivity to different cues tell us about bilinguals' mental representations?

The studies included in the project presented listeners with nonce words that contained one of seven language-specific segmental cues that either represented phonemic contrasts – phonological categories specific to each language – or phonetic distinctions – categories that exist in both languages but which are realized differently at the subsegmental level. Models of non-native speech perception and second language acquisition make related divisions between more and less familiar sounds, based on a listener's L1. The phonemic cues included one Spanish-specific segment, the Spanish /r/, and two English-specific phones, /ɹ/ and /θ/. Two phonetic categories were included, /l/ and /u/; the Spanish [l] is 'clear,' produced with the tongue further forward than the 'dark' English [ɫ], and the Spanish [u] is produced with the tongue further back than the fronted English [ʉ]. Listeners from different language experience backgrounds categorized these nonce words as sounding more like Spanish or more like English in self-paced (Experiments 1 and 3) or timed (Experiment 2) tasks. Listeners included English monolingual, early Spanish-English bilingual, and late Spanish-English bilingual adults (Monolinguals, Early Bilinguals, and Late Bilinguals; Experiments 1 and 2), and monolingual English-speaking children and children enrolled in a Spanish-English dual language program who heard Spanish only at school (and were from English-speaking families) or heard Spanish at home from their families as well (Monolinguals, SpanSchool, and SpanHome; Experiment 3). By presenting language-specific cues in context, these studies offered the opportunity to learn about the relative contributions of sounds from listeners' L1s and L2s to linguistic representations and to define when and how language experience strengthens or weakens the connections between phonological categories of both languages.

In Experiment 1, adults categorized the nonce-word stimuli at their own pace, and differences in categorization accuracy and RT emerged both across segments and across listener groups. Monolinguals, Early Bilinguals, and Late Bilinguals patterned similarly across the seven language-specific cues; Spanish /r/, English /ɹ/, and the phonetic /l/ variants were consistently categorized well, while all groups struggled with English /θ/ and English [ʉ] in particular. There were no differences in accuracy rates between the

Monolinguals and Early Bilinguals, but the Late Bilinguals categorized some English segments better than the other groups. These results show that Early Bilinguals' representations of English are very much like those of the Monolinguals', at least in a perception task in which phonological context is provided, which deviates from some of the findings of reduced sensitivity to L2 contrasts in other early bilingual populations (Sebastián-Gallés & Soto-Faraco, 1999). However, the difference in accuracy between the Early and Late Bilinguals indicates that Late Bilinguals had some advantage over the Early listeners in sensitivity to language-specific, especially to English-specific, phonological contrasts. It is proposed that some factor related to age of acquisition may explain this sensitivity, and the explicit L2 training received after puberty may have heightened the Late Bilinguals' awareness of the English-ness and Spanish-ness of the cues used in this task. In Experiment 1, the Late Bilinguals generally responded more slowly than the Early Bilinguals, who responded more slowly than the Monolinguals, so the improved accuracy in the Late Bilinguals group may be the result of a speed-accuracy trade-off. This possibility was tested in Experiment 2.

Experiment 2 used the same stimuli and basic design of Experiment 1, but to explore the relationship between accuracy and RT, listeners completed the categorization task in four blocks: first responses were self-paced, as in Experiment 1, and then listeners were forced to wait 404ms (Fast), 698ms (Medium), and 1978ms (Slow) after the nonce word to register their response. These timed responses were based on Monolinguals' response times in correct trials in Experiment 1, and were chosen to elicit categorization responses at times much faster (Fast), somewhat faster (Medium), and slower (Slow) than Early and Late Bilinguals' natural RTs in Experiment 1. There was no difference in segmental accuracy across groups between Experiments 1 and 2. There was a trend for improved accuracy in the Slow block, but the different wait lengths did not interact with the three listener groups; that is, the groups performed the same relative to each other in all three timed blocks. The constancy of listeners' accuracy in the speeded and delayed blocks suggests that the categorization differences between Early and Late Bilinguals is the result of differences in linguistic representations or listeners' access to language-specific phonological information. This finding provides support for the extension of the frequency-lag hypothesis (Gollan et al., 2005, 2011; Emmorey et al., 2013) to the association of phonological categories with their respective languages. The frequency-lag hypothesis predicts that the links between lexical form and the language from which the lexical form comes are weaker in bilinguals than in monolinguals; this is because bilinguals retrieve and produce forms from across two languages, so any specific link to a lexeme is accessed less frequently for a bilingual than for a monolingual, whose links are relatively stronger for having repeatedly accessed the form in their only language.

A related finding of Experiment 2 is the frequency with which the bilingual groups were unable to register a response during the 400ms response window that began after the interval of silence that varied in length across the blocks. There were certainly features of the experimental design that could have made interpreting the feedback on missed trials difficult, but this difficulty was significantly related to listener language background. Bilingual listeners – even those from a supplementary group who spoke a language other than Spanish at home – were disproportionately affected by the timed blocks and registered significantly more missed trials than the Monolinguals. This may further support the frequency-lag hypothesis: due to weaker links between phonological categories and languages, bilinguals needed additional time to accumulate sufficient information to make a decision, and this process often took longer than was allowed by the restricted response window. Among the bilingual groups, Spanish-English bilinguals missed more trials than other bilinguals, and the Early Bilinguals had more trouble responding on time than the Late Bilinguals. The robustness of these patterns indicates that competition between representations of English and Spanish was especially high for the Early Bilinguals, whose languages may be more balanced or more constantly activated than for the Late Bilinguals. If the Early Bilinguals are more likely to codeswitch or otherwise frequently switch between languages, this may have also contributed to high levels of activation and competition of both languages during the timed task.

In Experiment 3, six-year-olds completed a child-friendly version of Experiment 1, and as in the previous studies three listener groups were tested to consider a range of language experiences. Categorization accuracy patterns, and differences, closely resembled the accuracy of adults in Experiments 1 and 2, although children were somewhat less sensitive to differences in the phonetic cues. Two of the listener groups were of particular interest: the Monolingual children paralleled the Monolingual adults in their language profile, and the SpanHome children – who spoke Spanish at home and had begun learning English in school – would grow to be the Early Bilingual adults. While the Monolingual and Early Bilingual adults did not differ in their categorization accuracy in Experiments 1 and 2, there were significant differences between the Monolingual and SpanHome children across all segments: the SpanHome children were consistently more accurate than their Monolingual peers. However, this increased sensitivity to the language-specificity of segmental cues for the SpanHome children vanishes by adulthood.

An explanation for this decline of sensitivity into adulthood may lie in how new languages are processed, compared to how native or longer-spoken languages are stored and retrieved. New languages are processed differently, in distinct areas of the brain, than earlier acquired languages; later learned languages show greater neural activity in regions of the brain responsible for control and planning and show more variability in where they are processed across individuals (Abutalebi, 2008; Perani et al., 2003). As a speaker's proficiency improves and the processing of a new language becomes more automated, the neural networks responsible for processing the L2 begin to overlap with those used for the L1 (Abutalebi, 2008). It may be the case that the learner's metalinguistic awareness is particularly high before proficiency in the L2 approximates L1 proficiency and before the neural networks overlap. The regions recruited for L2 processing may contribute to metalinguistic awareness since they include the frontal lobe, which is responsible for control and planning (Abutalebi, 2008). This additional attention to the L2 could also highlight the ways in which the L2 differs from the L1, increasing awareness of the L1 properties as well, including phonological differences. However, it is unclear whether and how phonological awareness, or metalinguistic awareness more generally, changes across the lifespan as language proficiency develops. Research indicates greater metalinguistic awareness in bilingual children (Bruck & Genesee, 2005; Johnson & Wilson, 2002) and even in children only exposed to a second language (Akhtar et al., 2012), but as mentioned earlier, similar work on adults' awareness is very limited (e.g. to phonological awareness and literacy, as in Anthony & Francis, 2005).[9] The results of the present study provide

---

[9] There is evidence from research on the phonological processing of listeners with dyslexia that phonological representations across dyslexic and non-dyslexic individuals may actually be comparable while the *access* to these representations is compromised (Boets et al., 2013). The Boets et al. (2013) study provides evidence that the connection between the left inferior frontal gyrus (IFG) and the left superior temporal gyrus (STG) is degraded for listeners with dyslexia. It remains to be seen whether non-native speakers, without dyslexia, may similarly have reduced connectivity between the IFG and STG, at least in the earlier stages of learning. The connection between the left IFG and the left STG may then be strengthened with increased proficiency and as the L2 begins to be processed by the neural networks responsible for the L1.

support for the case that listeners who have more recently acquired a language are processing both languages differently than they will when processing becomes more automated. As a result, the listeners in this project who had learned their L2 recently were more accurate than other groups: the Late Bilinguals in Experiments 1 and 2 and for the SpanHome children in Experiment 3 were better able to associate English segments with English and Spanish segments with Spanish[10].

Categorization accuracy also varied across the seven target segments. These language-specific sounds were selected as example of two classes of phones relevant to cross-language speech perception. In the current study, these were called phonemic and phonetic cues and these categories were modeled on the new versus similar distinction defined by Flege (1987, 1995) for second language learners and on two of the contrasts described by Best (1991, 1995) for non-native speech perception. Phonemic cues map to separate native categories (for a monolingual; cf. Best, 1991) or are recognized as distinct from any native language sounds (for second language learners; cf. Flege, 1995) and are thus more highly salient to listeners and are acquired earlier. The phonetic cues are more difficult for listeners to distinguish and are acquired later; they map to a single native category and are perceived as the same as a native-language phone. The categorization accuracy of the seven target segments was relatively consistent across ages, language backgrounds, and task demands. This consistency supports the phonemic/phonetic (and new/similar, and separate category/same category) contrast being strongly rooted in the

[10] This pattern may also be true for the SpanSchool children, who heard only English at home but were enrolled in the Spanish-English dual language immersion program in school. These listeners showed a non-significant trend (p=0.10) for better accuracy than the Monolingual children, and the findings of Akhtar et al. (2012) might suggest that this effect could be strengthened with additional participants or a different task.

general saliency of sounds, and being modified in only a limited way by language experience. Furthermore, the ability of listeners to use phonemic and phonetic cues in context in language decision tasks may likewise reflect general perceptual salience and not the connection between language and cue in particular. However, when differences in accuracy did arise among the listener groups, it was always in the direction of the listeners who acquired their L2 most recently being more sensitive to language specificity, especially in the case of English cues. Some explanations exist in related literature for the difficulties listeners had categorizing English /θ/, one of the two most challenging cues, but no such theories are available to account for why English [ʉ], and the /u/ variants more generally, were hard to perceive or to link to English.

The variations in salience of the different kinds of segments studied here have direct implications for work on the processing and perception of foreign-accented speech. Both children and adults are able to identify and process foreign-accented speech, and the adults are especially facilitated if they have previous exposure to the foreign accent (Floccia et al., 2009; Girard et al., 2008; Vieru et al., 2011; Wagner et al., 2013; Witteman et al., 2013). However, adults (Clopper & Pisoni, 2004, 2007) but not children (Floccia et al., 2009; Girard et al., 2008) are also able to distinguish regional varieties from the dialect, again, especially in cases of adult listeners with exposure to the accent. The difficulty children have with foreign accents and more divergent native varieties like Indian English may in fact be related to the presence of phonemic and phonetic cues in the speech. The results of the three experiments presented here suggest that listeners have more difficulty with phonetic cues than with phonemic cues, and that listeners are not able to reliably use some cues in language decision tasks. The difference between more and less divergent accents may be that perceptually similar varieties are more likely to vary only in phonetic contrasts while the more distinct accents – foreign and otherwise – may contain more phonemic differences than regional dialects. This proposal could be tested by more directly comparing children's and adults' sensitivities to phonemic and phonetic cues; given the differences in regional- and foreign-accent perception between children and adults, it might be expected that children are less sensitive to phonetic cues than adults. Whether accents along the continuum described by Wagner et al. (2013) do vary in the frequency of phonemic and phonetic deviations also merits further inquiry.

This project also has important implications for variation in the speech perception abilities of early and late bilinguals. The findings from all three experiments reveal differences in the strength of language-cue associations for the different groups, and these associations shape how the relationship between bilinguals' language representations is conceptualized. Most previous studies demonstrating that early bilinguals are not significantly different from monolinguals have contrasted early learners with monolingual speakers of the bilinguals' L2, typically English (e.g. Flege et al., 1999; Mack, 1989). Early bilinguals' performance in their first language (here, Spanish) is not typically explored in cross-language speech perception tasks, and this is why it was so important in this project to include cues particular to both Spanish and English. The results of the applied tasks comprising this dissertation confirm earlier work using isolated segments: Early Bilingual listeners categorized nonce words like Monolinguals, despite their early exposure to Spanish and in contrast with the findings of Sebastián-Gallés and colleagues (1999, 2005). However the differences between Early and Late Bilinguals' accuracies indicate that there are differences in bilinguals' access to language representations – access that cannot be improved with longer response times (Experiment 2) but that has evolved in fundamental ways since childhood (Experiment 3). The change in Early Bilinguals' exposure to English in the years between kindergarten and college is likely a cause of the decreased access to language-specific representations, even if the acoustic detail in the representations themselves does not differ in quality from the Spanish-dominant Late Bilinguals' representations. The recency of L2 learning may also drive the greater accuracy of the SpanHome children who grow into the Early Bilingual adults, and this single explanation can account for the Late Bilingual adults' success as well.

In future work it will continue to be important to consider the contributions of consonants and vowels in the context of a word, as discussed earlier, since listeners may respond to these sound categories differently in isolation and in context. To this end it will be necessary to involve language pairs for which there are more language-specific contrasts and a wider variety of segments to be studied than those available for English and Spanish. All phonemic cues used here were consonants, with a necessary but limiting overreliance on the differences in rhotics across the languages. Similarly, the mispronunciation studies in Spanish and Catalan by Sebastián-Gallés et al. (2005) and Amengual (2014, 2015) were also restricted in scope, with vowels being the point of comparison across the languages. Contrasting a language pair that differs more significantly in both consonants and vowels at the phonemic and phonetic levels would provide the evidence needed to further test the conclusions drawn from these results.

The findings of this project also suggest that the recency of learning a language and how the language was learned (more or less explicitly) may affect the degree to which listeners are able to associate segments with their respective language and thus how they use language-specific cues in a categorization task. Differences in these dimensions seem to be especially important for improving listener accuracy for the most difficult, least salient segments. Recency and kind of language learning were confounded with age of acquisition for the adults in the present study, but the differences in children's accuracy lends support to a recency effect heightening sensitivity to language-specific acoustic cues. The strategic inclusion of additional listener groups, including learners who acquired their L2 in more and less formal contexts (e.g., in a classroom versus by moving to a new language community), would provide additional evidence to support or refute how language instruction and recency affect listener sensitivity to language-specific segments.

References

Abutalebi, J. (2008). Neural aspects of second language representation and language control. *Acta Psychologica*, 128, 466-478.

Akhtar, N., Menjivar, J., Hoicka, E., & Sabbagh, M. A. (2012). Learning foreign labels from a foreign speaker: the role of (limited) exposure to a second language. *Journal of Child Language*, 39, 1135-1149.

Albareda-Castellot, B., Pons, F., & Sebastián-Gallés, N. (2011). The acquisition of phonetic categories in bilingual infants: new data from an anticipatory eye movement paradigm. *Developmental Science* 14 (2), 395-401.

Allen, J.S., & Miller, J.L. (2004). Listener sensitivity to individual talker differences in voice-onset-time. *Journal of the Acoustical Society of America*, 115 (6), 3171-3183.

Amengual, M. (2012). Interlingual influence in bilingual speech: Cognate status effect in a continuum of bilingualism. *Bilingualism: Language and Cognition*, 15 (3), 517-530.

Amengual, M. (2014). The perception and production of language-specific mid-vowel contrasts: Shifting the focus to the bilingual individual in early language input conditions. *International Journal of Bilingualism*. Advance online publication. doi: 10.1177/1367006914544988.

Amengual, M. (2015). The perception of language-specific phonetic categories does not guarantee accurate phonological representations in the lexicon of early bilinguals. *Applied Psycholinguistics.* Advance online publication. doi: 10.1017/S0142716415000557.

Anthony, J.L., & Francis, D.J. (2005). Development of Phonological Awareness. *Current Directions in Psychological Science*, 14 (5), 255-259.

Aslin, R.N., Pisoni, D.B., Hennessy, B.L., & Perey, A.J. (1981). Discrimination of Voice Onset Time by Human Infants: New Findings and Implications for the Effects of Early Experience. *Child Development*, 52 (4), 1135-1145.

Baker, R.E., Bonnasse-Gahot, L., Kim, M., Van Engen, K.J., and Bradlow, A.R. (2011). Word durations in non-native English. *Journal of Phonetics,* 39 (1), 1–17.

Balukas, C., & Koops, C. (2015). Spanish-English bilingual voice onset time in spontaneous code-switching. *International Journal of Bilingualism*, 19 (4), 423-443.

Bates D., Maechler, M., Bolker, B., & Walker S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7. Retrieved from http://CRAN.R-project.org/package=lme4.

Best, C.T. (1991). The Emergence of Native-Language Phonological Influences in Infants: A Perceptual Assimilation Model. *Haskins Laboratories Status Report on Speech Research*, SR-107/108, 1-30.

Best. C.T. (1995). A Direct Realist View of Cross-Language Speech Perception. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, ed. W. Strange (Timonium, MD: York Press). 171-204.

Bialystok, E. (2001). 10. Metalinguistic aspects of bilingual processing. *Annual Review of Applied Linguistics*, 21, 169-181.

Boatman, D., Hall, C., Goldstein, M.H., Lesser, R., & Gordon, B. (1997). Neuroperceptual differences in consonant and vowel discrimination: As revealed by direct cortical electrical interference. *Cortex*, 33, 83-98.

Boersma, P., & Weenink, D. (2012). Praat: doing phonetics by computer [Computer program]. Version 5.3.32, retrieved 17 October 2012 from http://www.praat.org/.

Boets, B., Op de Beeck, H.P., Vandermosten, M., Scott, S.K., Gillebert, C.R., Mantini, D., Bulthé, J., Sunaert, S., Wouters, J., & Ghesquière, P. (2013). Intact But Less Accessible Phonetic Representations in Adults with Dyslexia. *Science*, 342 (6163), 1251-1254.

Bosch, L., Costa, A. and Sebastián-Gallés, N. (2000). First and second language vowel perception in early bilinguals. *European Journal of Cognitive Psychology* 12 (2), 189-221.

Bosch, L., & Sebastián-Gallés, N. (2003). Simultaneous Bilingualism and the Perception of a Language-Specific Vowel Contrast in the First Year of Life. *Language and Speech* 46 (2-3), 217-243.

Bradlow, A. (1995). A comparative acoustic study of English and Spanish vowels. *Journal of the Acoustical Society of America*, 97 (3), 1916-1924.

Bradlow, A., Clopper, C., Smiljanic, R., & Walter, M.A. (2010). A perceptual phonetic similarity space for languages: Evidence from five native language listener groups. *Speech Communication*, 52, 930-942.

Bruck, M., & Genesee, F. (1995). Phonological awareness in young second language learners. *Journal of Child Language*, 22, 307-324.

Bullock, B.E., & Toribio, A.J. (2009). Trying to hit a moving target: On the sociophonetics of code-switching. In *Multidisciplinary approaches to code switching,* eds. L. Isurin, D. Winford, & K. de Bot (Amsterdam: John Benjamins). 189-206.

Carlson, M.T., Goldrick, M., Blasingame, M., & Fink, A. (2015). Navigating conflicting phonotactic constraints in bilingual speech perception. *Bilingualism: Language and Cognition*. Advance online publication. doi: 10.1017/S1366728915000334.

Chan, C.L. (2014). NUsub-db: Northwestern University Subject Database [Web Application]. Department of Linguistics, Northwestern University. https://babel.ling.northwestern.edu/nusubdb2/.

Clark, E.V. (2009). *First language acquisition*. Cambridge: Cambridge University Press.

Clopper, C.G., & Pisoni, D.B. (2004). Homebodies and army brats: Some effects of early linguistic experience and residential history on dialect categorization. *Language Variation and Change*, *16*, 31–48.

Clopper, C.G., & Pisoni, D.B. (2007). Free classification of regional dialects of American English. *Journal of Phonetics*, *35*, 421–438.

Clopper C. G., Pisoni D. B., and de Jong K. (2005). Acoustic characteristics of the vowel systems of six regional varieties of American English. *Journal of the Acoustical Society of America,* 118, 1661–1676.

Costa, A., & Caramazza, A. (1999). Is lexical selection in bilingual speech production language-specific? Further evidence from Spanish-English and English-Spanish bilinguals. *Bilingualism: Language and Cognition*, 2 (3), 231-244.

Creel, S.C. (2012a). Phonological similarity and mutual exclusivity: on-line recognition of atypical pronunciations in 3-5-year-olds. *Developmental Science*, 15 (5), 697-713.

Creel, S.C. (2012b). Preschoolers' Use of Talker Information in On-Line Comprehension. *Child Development*.

Cutler, A., Weber, A., Smits, R., & Cooper, N. (2004). Patterns of English phoneme confusions by native and non-native listeners. *Journal of the Acoustical Society of America*, 116, 3668–3678.

Dąbrowska, E., & Street, J.A. (2006). Individual differences in language attainment: Comprehension of passive sentences by native and non-native English speakers. *Language Sciences*, 28, 604-615.

Diederich, A., & Busemeyer, J.R. (2006). Modeling the effects of payoff on response bias in a perceptual discrimination task: Bound-change, drift-rate-change, or two-stage-processing hypothesis. *Perception & Psychophysics*, 68 (2), 194-207.

Diehl, R.L., & Walsh, M.A. (1989). An auditory basis for the stimulus-length effect in the perception of stops and glides. *Journal of the Acoustical Society of America*, 85 (5), 2154-2164.

Dodd, B., Holm, A., Hua, Z., & Crosbie, S. (2003). Phonological development: a normative study of British English-speaking children. *Clinical Linguistics & Phonetics*, 17 (8), 617-643.

Eimas, P.D., Siqueland, E.R., Jusczyk, P., & Vigorito, J. (1971). Speech Perception in Infants. *Science*, 171 (3968), 303-306.

Emmorey, K., Petrich, J.A.F., & Gollan, T.H. (2013). Bimodal Bilingualism and the Frequency-Lag Hypothesis. *Journal of Dead Studies and Deaf Education*, 18 (1), 1-11.

Escudero, P. (2006). The phonological and phonetic development of new vowel contrasts in Spanish learners of English. In *English With a Latin Beat: Studies in Portuguese/Spanish-English Interphonology, Studies in Bilingualism*, Vol. 31, eds. B.O. Baptista, and M.A. Watkins (Amsterdam: John Benjamins), 149–161.

Fennell, C.T., & Waxman, S.R. (2010). What Paradox? Referential Cues Allow for Infant Use of Phonetic Detail in Word Learning. *Child Development* 81 (5), 1376-1383.

Flege, J.E. (1984). The detection of French accent by American listeners. *Journal of the Acoustical Society of America,* 76 (3), 692-707.

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, 47-55.

Flege, J.E. (1991). Age of learning affects the authenticity of voice onset time (VOT) in stop consonants produced in a second language. *Journal of the Acoustical Society of America,* 89, 395–411.

Flege, J.E. (1995). Second Language Speech Learning: Theory, Findings, and Problems. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, ed. W. Strange (Timonium, MD: York Press). 233-277.

Flege, J.E. (2007). Language contact in bilingualism: Phonetic system interactions. In *Laboratory Phonology 9*, eds. J. Cole & J. Hualde (Berlin: Mouton de Gruyter). 353-382.

Flege, J.E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, 25, 437-470.

Flege, J.E., MacKay, I.R.A., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America,* 106 (5), 2973-2987.

Flege, J.E., Munro, M.J., & Fox, R.A. (1994). Auditory and categorical effects on cross-language vowel perception. *Journal of the Acoustical Society of America*, 95 (6), 3623-3641.

Floccia, C., Butler, J., Girard, F., & Goslin, J. (2009). Categorization of regional and foreign accent in 5- to 7-year-old British children. International Journal of Behavioral Development 33 (4), 366-375.

Fogerty, D., & Kewley-Port, D. (2009). Perceptual contributions of the consonant-vowel boundary to sentence intelligibility. *Journal of the Acoustical Society of America*, 126 (2), 847-857.

Francis, A.L., & Nusbaum, H.C. (2002). Selective Attention and the Acquisition of New Phonetic Categories. *Journal of Experimental Psychology: Human Perception and Performance*, 28 (2), 349-366.

Gelman, A., Lee, D., & Guo, J. (2015). A Probabilistic Programming Language for Bayesian Inference and Optimization. *Journal of Educational and Behavioral Statistics*, 40 (5), 530-543.

Genesee, F., Boivin, I., & Nicoladis, E. (1995). Talking with Strangers: A Study of Bilingual Children's Communicative Competence. *Applied Psycholinguistics,* 17: 427-442.

Gertken, L. M., Amengual, M., & Birdsong, D. (2014). Assessing language dominance with the Bilingual Language Profile. In *Measuring L2 proficiency: Perspectives from SLA*, eds. P. Leclercq, A. Edmonds, & H. Hilton (Bristol: Multilingual Matters). 208–225.

Girard, F., Floccia, C., & Goslin, J. (2008). Perception and awareness of accents in young children. *British Journal of Developmental Psychology*, *26*(3), 409–433.

Goldrick, M., Runnqvist, E., & Costa, A. (2014). Language Switching Makes Pronunciation Less Nativelike. *Psychological Science*, 25 (4), 1031-1036.

Gollan, T.H., Montoya, R.I., Fennema-Notestine, C., & Morris, S.K. (2005). Bilingualism affects picture naming but not picture classification. *Memory & Cognition*, 33 (7), 1220-1234.

Gollan, T.H., Slattery, T.J., Goldenberg, D., Van Assche, E., & Duyck, W. (2011). Frequency Drives Lexical Access in Reading but Not in Speaking: The Frequency-Lag Hypothesis. *Journal of Experimental Psychology: General*, 140 (2), 186-209.

Grosjean, F. (1989). Neurolinguists, Beware! The Bilingual is Not Two Monolinguals in One Person. *Brian and Language*, 36, 3-15.

Harnsberger, J.D., Shrivastav, R., Brown, Jr., W.S., Rothman, H., & Hollien, H. (1997). Speaking rate and fundamental frequency as speech cues to perceived age. *Journal of Voice*, 22 (1), 58-69.

Harrell, Jr., F.E. (2014). rms: Regression Modeling Strategies. R package version 4.2-1. Retrieved from http://CRAN.R-project.org/package=rms.

Hualde, J. I. (2005). *The Sounds of Spanish*. New York: Cambridge University Press.

Iverson, P., Kuhl, P.K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, 87, B47-B57.

Jenkins, J.J., Strange, W., & Polka, L. (1995). Not everyone can tell a "rock" from a "lock": Assessing individual differences in speech perception. In *Assessing individual differences in human behavior: New concepts, methods, and findings*, eds. D.J. Lubinski & R.V. Dawis (Palo Alto, CA: Davies-Black Publishing). 297-325.

Johnson, K. (1997). Speech perception without speaker normalization. In *Talker Variability in Speech Processing*, eds. K. Johnson & J.W. Mullennix (New York: Academic Press). 145-166.

Johnson, C. E., & Wilson, I. L. (2002). Phonetic evidence for early language differentiation: Research issues and some preliminary data. *International Journal of Bilingualism* 6 (3), 271-289.

Kewley-Port, D., Burkle, T.Z., & Lee, J.H. (2007). Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners. *Journal of the Acoustical Society of America*, 122 (4), 2365-2375.

Kinzler, K. D., Shutts, K., DeJesus, J., & Spelke, E. S. (2009). Accent trumps race in guiding children's social preferences. *Social Cognition* 27 (4), 623-634.

Kinzler, K. D., Corriveau, K. H., & Harris, P. L. (2011). Children's selective trust in native-accented speakers. *Developmental Science* 14 (1), 106-111.

Klatt, D.H., & Klatt, L.C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America*, 87 (2), 820-857.

Kluender, K.R., Diehl, R.L., & Killeen, P.R. (1987). Japanese Quail Can Learn Phonetic Categories. *Science*, 237 (4819), 1195-1197.

Kondaurova, M.V., & Francis, A.L. (2008). The relationship between native allophonic experience with vowel duration and perception of the English tense/lax vowel contrast by Spanish and Russian listeners. *Journal of the Acoustical Society of America*, 124 (6), 3959-3971.

Kovelman, I., Baker, S.A., & Petitto, L.A. (2008). Bilingual and Monolingual Brains Compared: A Functional Magnetic Resonance Imagining Investigation of Syntactic Processing and a Possible "Neural Signal" of Bilingualism. *Journal of Cognitive Neuroscience*, 20 (1), 153-169.

Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Cognitive Psychonomic Bulletin & Review*, 13 (2), 262–268.

Kraljic, T., & Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language*, 56, 1-15.

Kroll, J.F., & Stewart, E. (1994). Category Interference in Translation and Picture Naming: Evidence for Asymmetric Connections Between Bilingual Memory Representations. *Journal of Memory and Language*, 33 (2), 149-174.

Kuhl, P.K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, 50 (2), 93-107.

Kuhl, P.K., & Miller. J.D. (1978). Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. *Journal of the Acoustical Society of America*, 63 (3), 905-917.

Kuhl, P. K., Tsao, F.-M., & Liu, H.-M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *PNAS*, 100 (15): 9096-9101.

Kuznetsova, A., Brockhoff, P.B., & Christensen, R.H.B. (2014). lmerTest: Tests in Linear Mixed Effects Models. R package version 2.0-20. Retrieved from http://CRAN.R-project.org/package=lmerTest.

Labov, W. (1971).The study of language in its social context. In *Advances in the Sociology of Language*, ed. J.A. Fishman (The Hague, The Netherlands: Mouton). 152-216.

Lasky, R.E., Syrdal-Lasky, A., & Klein, R.E. (1975). VOT Discrimination by Four to Six and a Half Month Old Infants from Spanish Environments. *Journal of Experimental Child Psychology*, 20, 215-225.

Lisker, L., & Abramson, A.S. (1970). The voicing dimension: Some experiments in comparative phonetics. In *Proceedings of the 6th International Conference of Phonetic Sciences*, (Prague: Academia). 563-567.

Lively, S.E., & Pisoni, D.B. (1997). On Prototypes and Phonetic Categories: A Critical Assessment of the Perceptual Magnet Effect in Speech Perception. *Journal of Experimental Psychology*, 23 (6), 1665-1679.

Lotto, A.J., Kluender, K.R., & Holt, L.L. (1998). Depolarizing the perceptual magnet effect. *Journal of the Acoustical Society of America*, 103 (6), 3648-3655.

Mack, M. (1989). Consonant and vowel perception and production: Early English-French bilinguals and English monolinguals. *Perception and Psychophysics*, 46 (2), 187-200.

Mani, N., & Plunkett, K. (2007). Phonological specificity of vowels and consonants in early lexical representations. *Journal of Memory and Language*, *57*(2), 252–272.

Maye, J., Werker, J.F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82, B101-B111.

McElreath, R. (2013). glmer2stan: RStan models defined by glmer formulas. R package version 0.995.

McElree, B., Murphy, G.L., & Ochoa, T. (2006). Time course of receiving perceptual information: A speed-accuracy trade-off study. *Psychonomic Bulletin & Review*, 13 (5), 848-853.

Mendez, A. (1982). Production of American English and Spanish vowels. *Language and Speech*, 25, 191-197.

Miller, J.L. (1997). Internal Structure of Phonetic Categories. *Language and Cognitive Processes*, 12 (5-6), 865-870.

Miller, J.L., & Volaitis, L.E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics*, 46 (6), 505-512.

Munro, M.J., Derwing, T.M., & Burgess, C.S. (2010). Detection of nonnative speaker status from content-masked speech. *Speech Communication*, 52: 626-637.

Munro, M.J., Flege, J.E., & Mackay, I.R.A. (1996). The effects of second language learning on the production of English vowels. *Applied Psycholinguistics*, 17, 313-334.

Nathan, L., Wells, B., & Donlan, C. (1998). Children's comprehension of unfamiliar regional accents: a preliminary investigation. *Journal of Child Language,* 25, 343-365.

Nazzi, T. (2005). Use of phonetic specificity during the acquisition of new words: Differences between consonants and vowels. *Cognition*, *98*, 13–30.

Nicoladis, E. (1998). First clues to the existence of two input languages: Pragmatic and lexical differentiation in a bilingual child. *Bilingualism: Language and Cognition* 1, 105-116.

Nicoladis, E., & Genesee, F. (1996). A Longitudinal Study of Pragmatic Differentiation in Young Bilingual Children. *Language Learning* 46 (3), 439-464.

Nygaard, L.C., & Pisoni, D.B. (1998). Talker-specific learning in speech perception. *Perception and Psychphysics*, 60, 355–376.

Olson, D.J. (2013). Bilingual language switching and selection at the phonetic level: Asymmetrical transfer in VOT production. *Journal of Phonetics*, 41 (6), 407-420.

Owren, M.J., & Cardillo, G.C. (2006). The relative roles of vowels and consonants in discriminating talker identify versus word meaning. *Journal of the Acoustical Society of America*, 119, 1727-1739.

Paradis, J., & Nicoladis, E. (2007). The Influence of Dominance and Sociolinguistic Context on Bilingual Preschoolers' Language Choice. *International Journal of Bilingual Education and Bilingualism* 10 (3), 277-297.

Pennington, B.F., van Orden, G.C., Smith, S.D., Green, P.A., & Haith, M.M. (1990). Phonological Processing Skills and Deficits in Adult Dyslexics. *Child Development,* 61 (6), 1753-1778.

Perani, D., Abutalebi, J., Paulesu, E., Brambati, S., Scifo, P., Cappa, S.F., & Fazio, F. (2003). The Role of Age of Acquisition and Language Usage in Early, High-Proficient Bilinguals: An fMRI Study During Verbal Fluency. *Human Brain Mapping*, 19, 170-182.

Pierrehumbert, J.B. (2002). Word-specific phonetics. In *Laboratory phonology, Vol. VII*, eds. C. Gussenhoven, & N. Warner, (Berlin: Mouton de Gruyter). 101–139.

Pisoni, D.B. (1977). Identification and discrimination of the relative voice onset time of two component tones: Implication for voicing perception in stops. *Journal of the Acoustical Society of America,* 61 (5), 1352-1361.

Pisoni, D.B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics,* 15 (2), 285-290.

Psychology Software Tools, Inc. E-Prime 2.0. (2010). Retrieved from http://www.pstnet.com.

R Core Team. (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.R-project.org/.

Ransdell, S.E., & Fischler, I. (1987). Memory in a Monolingual Mode: When are Bilinguals at a Disadvantage? *Journal of Memory and Language*, 26, 392-405.

Recasens, D. (2004). Darkness in [l] as a scalar phonetic property: implications for phonology and articulatory control. *Clinical Linguistics & Phonetics*, 18 (6-8), 593-603.

Recasens, D. (2012). A cross-language acoustic study of initial and final allophones of [l]. *Speech Communication,* 54, 368-383.

Repp, B.H. (1982). Phonetic Trading Relations and Context Effects: New Experimental Evidence for a Speech Mode of Perception. *Psychological Bulletin*, 92 (1), 81-110.

Repp, B.H., & Lin, H.-b. (1991). Effects of preceding context on the voice-onset-time-category boundary. *Journal of Experimental Psychology: Human Perception and Performance*, 17 (1), 289-302.

RStudio Team (2015). *RStudio: Integrated Development for R*. RStudio, Inc., Boston, MA. URL http://www.rstudio.com/.

Schmale, R., Hollich, G. J., & Seidl, A. (2011). Contending with Foreign Accent in Early Word Learning. *Journal of Child Language,* 38, 1096-1108.

Sebastián-Gallés, N., Echeverría, S., & Bosch, L. (2005). The influence of initial exposure on lexical representation: Comparing early and simultaneous bilinguals. *Journal of Memory and Language,* 52, 240-255.

Sebastián-Gallés, N., & Soto-Faraco, S. (1999). Online processing of native and non-native phonemic contrasts in early bilinguals. *Cognition,* 72, 111-123.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society*, Series B, 64 (4), 583–639.

Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature* 388, 381-382.

Strand, E.A. (1999). Uncovering the role of gender stereotypes in speech perception. *Journal of Language and Social Psychology*, 18 (1), 86-99.

Strand, E.A., & Johnson, K. (1996). Gradient and visual speaker normalization in the perception of fricatives. In *Natural language processing and speech technology: Results of the 3rd KONVENS conference, Bielefeld, October 1996,* ed. D. Gibbon (Berlin: Mouton). 14-26.

Tajima, K., Port, R., & Dalby, J. (1997). Effects of temporal correction on intelligibility of foreign-accented English. *Journal of Phonetics*, 25 (1), 1-24.

Tracy, E.C., Bainter, S.A., & Satariano, N.P. (2015). Judgments of self-identified gay and heterosexual male speakers: Which phonemes are the most salient in determining sexual orientation? *Journal of Phonetics*, 52, 13-25.

Vieru, B., Boula de Mareüil, P., & Adda-Decker, M. (2011). Characterisation and identification of non-native French accents. *Speech Communication*, 53, 292-310.

Wagner, L., Clopper, C. G., & Pate, J. K. (2013). Children's perception of dialect variation. *Journal of Child Language*.

Weber, A., and Cutler, A. (2006). First-language phonotactics in second-language listening. *Journal of the Acoustical Society of America*, 119 (1), 597–607.

Werker, J. F., & Tees, R. C. (1984). Cross-Language Speech Perception: Evidence for Perceptual Reorganization During the First Year of Life. *Infant Behavior and Development* 7, 49-63.

Werker, J. F., & Tees, R. C. (1999). Influences on Infant Speech Processing: Toward a New Synthesis. *Annual Review of Psychology* 50, 509-535.

Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York.

Witteman, M. J., Weber, A., & McQueen, J. M. (2013). Foreign accent strength and listener familiarity with an accent codetermine speed of perceptual adaptation. *Attention, Perception, and Psychophysics*, 75, 537-556.