Differences in the association between segment and language: Early bilinguals pattern with monolinguals and are less accurate than late bilinguals

1   **Cynthia P. Blanco[1]*, Colin Bannard[2], Rajka Smiljanic[1]**

2

3   [1]Department of Linguistics, University of Texas at Austin, Austin, TX, USA

4

5   [2]Department of Psychological Sciences, University of Liverpool, Liverpool, UK

6

7   *** Correspondence:** Cynthia P. Blanco, Department of Linguistics, University of Texas at Austin, 1
8   University Station B5100, Austin, TX, 78712, USA.

9

10   cynthiapblanco@gmail.com

13

14   Number of words: 11175

15

16   Number of figures and tables: 13

17   **Abstract**
18   Early bilinguals often show as much sensitivity to L2-specific contrasts as monolingual speakers of
19   the L2, but most work on cross-language speech perception has focused on isolated syllables, and
20   typically only on neighboring vowels or stop contrasts. In tasks that include sounds in context,
21   listeners' success is more variable, so segment discrimination in isolation may not adequately
22   represent the phonetic detail in stored representations. The current study explores the relationship
23   between language experience and sensitivity to segmental cues in context by comparing the
24   categorization patterns of monolingual English listeners and early and late Spanish-English
25   bilinguals. Participants categorized nonce words containing different classes of English- and Spanish-
26   specific sounds as being more English-like or more Spanish-like; target segments included phonemic
27   cues, cues for which there is no analogous sound in the other language, or phonetic cues, cues for
28   which English and Spanish share the category but for which each language varies in its phonetic
29   implementation. Listeners' language categorization accuracy and reaction times were analyzed.

30

31   Our results reveal a largely uniform categorization pattern across listener groups: Spanish cues were
32   categorized more accurately than English cues, and phonemic cues were easier for listeners to
33   categorize than phonetic cues. There were no differences in the sensitivity of monolinguals and early
34   bilinguals to language-specific cues, suggesting that the early bilinguals' exposure to Spanish did not
35   fundamentally change their representations of English phonology. However, neither did the early
36   bilinguals show more sensitivity than the monolinguals to Spanish sounds. The late bilinguals
37   however, were significantly more accurate than either of the other groups. These findings indicate
38   that listeners with varying exposure to English and Spanish are able to use language-specific cues in
39   a nonce-word language categorization task. Differences in how, and not only when, a language was
40   acquired may influence listener sensitivity to more difficult cues, and the advantage for phonemic
41   cues may reflect the greater salience of categories unique to each language. Implications for foreign-

42 accent categorization and cross-language speech perception are discussed, and future directions are
43 outlined to better understand how salience varies across language-specific phonemic and phonetic
44 cues.

45 **1   Introduction**
46
47 Listeners make judgments about talkers and their speech after only brief exposure. Considerable
48 work has investigated the suprasegmental and segmental acoustic cues most important for listeners in
49 their decisions about talker-specific characteristics like region of origin, age, and gender (Clopper &
50 Pisoni, 2004, 2007; Harnsberger et al., 1997; Klatt & Klatt, 1990; Strand, 1999; Strand & Johnson,
51 1996; Tracy et al., 2015). Other cues may indicate that a talker grew up using a language other than
52 the one being spoken, yielding a foreign accent (e.g., Flege 1991; Flege et al., 1997a, 1997b; Flege &
53 Munro, 1994). At times it may even be necessary for listeners to identify which language a talker is
54 using, for example, so that a bilingual can map a new word to the appropriate language or to facilitate
55 a bilingual's access of a known word in one of their languages (Flege, 2007). However, unlike the
56 work investigating associations of acoustic properties with indexical information like region of
57 origin, cross-language speech perception tasks typically test only isolated vowels without a larger
58 phonological context or consonants in a single CV syllable (although some work also presents stop
59 bursts without context, e.g., Flege, 1984). These segments are often very limited in range (e.g.,
60 comparing neighboring vowels only). It is therefore unclear which segmental cues are most useful to
61 listeners in making distinctions between their languages or whether listeners attend to all language-
62 specific acoustic cues equally. The current project seeks to test listener sensitivity to a range of
63 language-specific segments in nonce word contexts and considers how a listener's language
64 background influences their use of these cues in a cross-language speech perception task.
65
66 Previous work has examined how listeners' language experience shapes their ability to categorize or
67 discriminate isolated, or nearly-isolated, segments and subsegmental cues in cross-language speech
68 perception. In these studies, bilingual listeners categorize or discriminate between pairs or triplets of
69 sounds ranging along a continuum, most often the VOT continuum (e.g. between /t/ and /d/) or
70 formant continua between neighboring vowels in the L2 (e.g. /i/ and /ɪ/). These studies have shown
71 that monolingual English listeners and early bilinguals make similar distinctions among English
72 categories (e.g., Flege et al., 1999a; Mack, 1989), and that this is especially true for bilinguals who
73 have lower rates of continued use of or exposure to their L1 (Flege & MacKay, 2004). In some vowel
74 discrimination tasks, even late bilinguals pattern like English monolinguals (Flege et al., 1994).
75 However, listeners use a host of cues when perceiving speech beyond isolated segments or syllables,
76 and in fact, differentiating native and non-native stop bursts may not require accessing linguistic
77 representations at all, as is the case when listeners make parallel judgments between continua of non-
78 speech sounds (Diehl & Walsh, 1989; Pisoni, 1977). It is possible that listeners use different, even
79 non-linguistic and general auditory, strategies to make decisions about the isolated segments and
80 syllables and acoustic cues used in these identification and discrimination tasks (Flege, 1987).
81 Furthermore, these studies typically only evaluate listener sensitivity to cues in the L2, most often
82 English, so very little is known about how they process segments particular to their first language.[1]
83
84 A few studies have attempted to extend the findings on the perception of segments in isolation or in
85 syllables to the perception of language-specific speech and accented productions in longer stimuli. In
86 a series of experiments, Flege (1984) found that listeners could distinguish native and non-native

---

[1] See Carlson et al. (2015) for recent work on early bilinguals' use of L1 phonotactics in speech perception.

87   talkers of English after hearing CV syllables, single words, and three-word phrases. Even more
88   remarkably, native English listeners could use input as brief as 30ms of a stop burst to differentiate
89   productions from native- and French-accented talkers. However, it is not clear that the strategies
90   listeners used are the same across these varying materials despite the fact that listeners mostly
91   accurately categorized stimuli from across this range of input. For the longer utterances, listeners
92   may not have necessarily made use of stop burst differences at all, even though they can identify
93   these differences in other tasks. Instead, listeners may pay more attention to other segmental and
94   suprasegmental cues present in the longer stretches of speech. That is, the presence of a usable
95   language-specific cue like a stop burst does not necessarily mean that this will be the most useful cue
96   when other cues are present, and other cues may in fact be more salient to listeners than VOT. For
97   example, evidence from a perceptual-similarity task using phrase-length stimuli from 17 languages
98   suggests that marked back consonants and front vowel rounding might be particularly salient
99   dimensions for non-native listeners (Bradlow et al., 2010). However, there remains some question
100   about the interpretation of at least the vowel dimension in the perceptual-similarity study, so the
101   number of cues present in even short phrases makes it difficult to identify the most influential
102   acoustic factors.
103
104   Flege and Munro (1994) tested listener sensitivity to the multiple cues available in word-length
105   stimuli by asking monolingual English listeners to categorize productions of *taco* as having been
106   produced in English or in Spanish. The length of VOT associated with the initial /t/ explained more
107   variance in listeners' responses than any other acoustic cue, but this language-specific difference is
108   confounded with having occurred so early in the word – listeners may not have attended to the whole
109   word if they could confidently make a decision based on the first segment or syllable. Since all four
110   segments were Spanish-like or English-like in any production of *taco*, the results also do not reveal
111   which cue(s) listeners would rely on, in the absence of the other cues. The VOT of /t/ was the
112   strongest cue, but it is unclear if the other cues would have been sufficient for listeners to categorize
113   productions accurately. The sensitivity of monolingual listeners to language-specific stops in Flege
114   (1984) and Flege and Munro (1994) suggests that listeners can compare the VOT of the stimulus to
115   their stored representations of what is an acceptable or atypical VOT for English stops. It remains to
116   be seen whether bilinguals would show the same sensitivity to these cues in more naturalistic, word-
117   length contexts. By manipulating a single cue in a stimulus word, and holding constant the remaining
118   segments, we can begin to understand whether listeners from different language backgrounds can
119   make use of a given cue when evaluating their lexical representations.
120
121   Work from mispronunciation studies indicates that bilingual listeners who can easily discriminate
122   segments or syllables in isolation might be less able to identify those same differences in word-length
123   stimuli, and this disparity across tasks is true even for early, highly-proficient bilinguals. Listeners in
124   these studies complete identification and discrimination tasks, and then identify whether a stimulus is
125   the typical pronunciation of the word or if it is mispronounced. For the segment identification tasks
126   contrasting neighboring vowels in Catalan (e.g. /ɛ/~/e/), there are conflicting results: highly-
127   proficient Spanish-dominant Spanish-Catalan bilinguals in Barcelona were unable to reliably
128   distinguish the Catalan mid-vowels is isolation (Sebastián-Gallés & Soto-Faraco, 1999), while their
129   peers in Majorca were successful (Amengual, 2015). However, Spanish-dominant bilinguals in both
130   locales responded similarly poorly in the mispronunciation tasks, in which they heard a word's actual
131   mid-vowel replaced with the neighboring vowel (e.g., /ɛ/ replaced with /e/, as in /ərɛl/ 'root'
132   pronounced as */ərel/). Sebastián-Gallés and colleagues (1999, 2005) attribute the lack of detail in
133   Spanish-dominant bilinguals' representations of Catalan to their exposure to Spanish in the first years
134   of life, before acquiring Catalan. However, Amengual's results indicate that early Spanish exposure

135 itself is not the cause of early bilinguals' decreased discrimination abilities in the mispronunciation
136 task, since listeners in Majorca could reliably perceive differences when the segments were presented
137 in isolation. This suggests that, in both regions, the Spanish-dominant bilinguals' lexical
138 representations of Catalan contain less phonetic detail for Catalan-specific contrasts, despite the
139 ability of some listeners to discriminate the segments in other tasks. This difference in the detail of
140 bilinguals' lexical representations reflects the kinds of variation to which listeners are exposed, and
141 the construction of representations is likely more complex than would be suggested by a listener's
142 ability to discriminate isolated sounds or syllables. It is therefore important that investigations into
143 the nature of bilinguals' representations of their languages use tasks that force listeners to respond to
144 more complex input as language to better understand the level of detail encoded in lexical
145 representations and to more closely approximate the challenge of processing naturalistic speech.
146
147 In fact, lexical representations incorporate not only phonological variation but social information
148 associated with that variation as well. These indexical features, such as speaker and contextual
149 characteristics, are encoded in the lexical representations, and they may be incorporated even after
150 only brief exposure in the lab (e.g. Allen & Miller, 2004; Kraljic & Samuel, 2006, 2007; Nygaard &
151 Pisoni, 1998). If the Spanish-Catalan bilinguals heard more variable input in the productions of real
152 words, their representations of Catalan may have included both productions as possible, explaining
153 their difficulty identifying mispronunciations, whereas the monolinguals in Flege (1984) and Flege &
154 Munro (1994) may have been exposed to less variation in English and so were more sensitive to
155 deviations from typical productions. There is also evidence demonstrating that listeners with
156 exposure to specific accents, even in absence of knowing the L2, show improved processing and
157 categorization of those accents (Clopper & Pisoni, 2004, 2007; Vieru et al., 2011; Witteman et al.,
158 2013), so language and a talker's language proficiency must also be linked to specific productions.
159
160 These associations of indexical information with productions, and the incorporation of acoustic
161 variation in lexical representations, are in line with exemplar theories of speech perception (Johnson,
162 1997; Pierrehumbert, 2002). Listeners use stored exemplars – those from an exposure period in a lab
163 or from hearing productions in normal life – to inform their expectations about unheard productions
164 and word forms. Thus, listeners can generalize over a number of stored exemplars about what kinds
165 of stops, for example, occur in English or in the productions of a particular talker of English.
166 Listeners like bilinguals who have experience with a sound category in both languages must associate
167 productions with each language in order to make the appropriate conclusions about the phonological
168 categories in each language (as in the related BLINCS model in Shook & Marian, 2013). For
169 example, a Spanish-English bilingual who hears a word produced with a /t/ will store with this
170 exemplar whether the sound was produced in English or Spanish, and information about how it was
171 produced (e.g., the VOT of the stop) will be added to the listener's representation for the production
172 of /t/ in the language. Spanish-English bilinguals will therefore have developed detailed phonological
173 representations for English and Spanish, and their sensitivity to the distribution of sounds particular
174 to each language might be expected to be greater than that of English monolinguals, who have only
175 English productions on which to base their language representations. While English monolinguals
176 may have some, or even significant, exposure to Spanish-accented English, their knowledge of
177 Spanish phonology will be less than that of bilinguals who have acquired Spanish since birth. In fact,
178 due to existence of multiple (language-specific) categories in the same phonological space, Spanish-
179 English bilinguals' representations might also be unlike English monolinguals' in other ways:
180 bilinguals might use categories more extreme than monolinguals to maximize differences between
181 languages (cf. Flege, 1995), or bilinguals' categories may show evidence of cross-linguistic transfer

182 and be less like the monolinguals', especially for later-acquired sounds and for later learners (Flege,
183 2007).
184
185 The present study tests the effect of language experience on listener sensitivity to language-specific
186 segments to better understand how language-specific sounds are represented and related in the
187 bilingual lexicon. We use a novel task in which listeners are told they are hearing snippets of
188 continuous speech (either in Spanish or English) and are asked to associate the nonce words
189 containing a Spanish- or English-specific sound with the appropriate language. Accuracy and
190 reaction times are compared across listener groups for each of the classes of segment. The use of
191 nonce words has two advantages. First, presenting word-length stimuli forces listeners to process the
192 sounds linguistically and not just auditorily, and there is evidence that listeners in previous studies
193 may have perceived segments without linguistic context differently than when the same sounds were
194 processed as words. Second, unlike real words, nonce-word stimuli avoid inducing lexical effects
195 related to listeners' actual exposure to the phonological variations of real words. Finally, the use of
196 word-length nonce stimuli, purportedly taken from naturally produced speech, forces listens to
197 generalize over the phonological properties of their languages and decide in which language a given
198 stimulus must have been produced. The present study also extends previous work, which mostly
199 tested contrasts from only one language (e.g., English in Flege's work and Catalan in the work of
200 Sebastián-Gallés and Amengual), by including cues from both English and Spanish to more fully
201 investigate how listeners' language backgrounds influence perception in both languages.
202
203 The nonce words tested here include segmental categories that are unique to English or Spanish
204 ("phonemic" cues) and segments that vary in how they are implemented phonetically along a
205 continuum between the Spanish variant and the English variant ("phonetic" cues). Similar
206 distinctions among segments have been made for the perception of non-native sounds that vary in
207 similarity to native categories (Best, 1991) and for the acquisition of second language sounds, in the
208 Speech Learning Model (Flege, 1987, 1995). Evidence suggests that sound categories that are "new"
209 to an L2 and have no counterpart in the L1, like the phonemic cues proposed here, are easier to
210 perceive as a distinct category and to produce authentically than "similar" L2 phones that differ along
211 some particular acoustic-articulatory dimension from the L1 variant, like the phonetic cues described
212 here. One study (Flege & Munro, 1994) has specifically examined phonetic cues in context and
213 found that listeners could use these cues to varying degrees depending on the language background
214 of the talker, but no work has directly compared phonemic and phonetic cues. Following Flege and
215 Munro (1994) and the predictions outlined in the Speech Learning Model for new and similar
216 phones, both classes of cues are expected to be successfully associated with their respective
217 languages but phonemic cues are expected to be stronger indicators of language than phonetic cues in
218 a language categorization task.[2]
219
220 Finally, this study also systematically compares the sensitivity of monolingual English listeners and
221 early and late Spanish-English bilinguals. Previous work in cross-language speech perception
222 indicates similarities between English monolinguals and early Spanish-English bilinguals in the
223 categorization of English sounds, but evidence regarding how late bilinguals compare to these groups
224 is more limited. It is expected that the bilingual groups will show greater sensitivity to language-
225 specific cues from both languages than the monolinguals, since the bilinguals' considerable exposure

---

[2] While the Speech Learning Model distinguishes between new and similar phones in a second language, this binary may not be sufficient to include all relationships between the sounds of one's native language and the categories in a second language. For example, it is unclear how to classify a shared phone with different statuses in each language, e.g., both Spanish and English use the tap [ɾ], but this sound is phonemic in Spanish and allophonic in English.

226   to both English and Spanish productions should foster more reliable associations between language
227   and the phonetic detail in stored representations.

228   **2   Methods**

229   **2.1   Materials**

230   **2.1.1 Language-specific target segments**

231   Three language-specific phonemic cues were chosen for the categorization task: the English-specific
232   segments /θ/ and /ɹ/, and the Spanish-specific trill /r/. We limited the selection of phonemic cues to
233   those sounds that form categories not present in the other language and that do not form a continuum.
234   For example, the English voiced alveolar approximant /ɹ/ and the Spanish voiced alveolar trill /r/ are
235   not different extremes of a continuum between /ɹ/ and /r/, in the way that English and Spanish voiced
236   and voiceless stops vary along a single dimension (VOT). That is, there is not a single dimension or
237   acoustic correlate that distinguishes /ɹ/ and /r/ that could be increased or decreased to derive one from
238   another, since the two sounds are produced with fundamentally different manners of articulation (/ɹ/
239   as an approximant and /r/ as a trill). One additional English-specific cue was identified for inclusion
240   as a phonemic cue, /θ/. Although /θ/ is a phoneme in Peninsular Spanish (it is produced as /s/ in Latin
241   America), it was included as an English-specific phoneme since exposure to Peninsular Spanish
242   among our listeners was expected to be very limited, and native speakers of Peninsular Spanish were
243   excluded from the study. Early Spanish-English bilingual listeners living in Central Texas, where this
244   study was conducted, may have some exposure to Peninsular Spanish, for example through movies,
245   but are most familiar with Latin American dialects of Spanish. The late bilingual participants likely
246   have more exposure to Peninsular Spanish than early bilinguals, but it is not expected that this
247   exposure would be more influential on L1 representations than native dialect phonology. In fact,
248   many monolingual English listeners probably have exposure to the trill /r/ in Scottish English, also
249   through media, but it would be surprising if their language-segment associations reflected occasional
250   exposure to the trill /r/ in English.[3] Vowels were excluded as phonemic cues for this language pair for
251   two reasons. First, all five Spanish vowel categories exist in English, minimally in English
252   diphthongs, so there were no Spanish-specific vowels to consider for phonemic cues. Second,
253   English-specific vowels (e.g. /ɪ/) can be differentiated from the nearest shared vowels (e.g. /i/) by
254   both spectral cues and duration differences; while native listeners attend to the spectral differences in
255   these English-specific vowels, non-native listeners may rely on vowel duration to distinguish these
256   categories (Escudero, 2006; Flege et al., 1997a; Kondaurova & Francis, 2008). In this case, non-
257   native listeners would be able to use the duration continuum between the short /ɪ/ and the long /i/.
258   Instead, we wanted to ensure as much as possible that all listener groups included in this study were
259   attending to the same acoustic property of the target segment.[4]
260
261   In addition to the phonemic cues, we also tested phonetic cues, which vary along a continuum. These
262   sound categories exist in both languages but their articulation in each language is characterized by

---

[3] In fact, our results suggest that late bilingual listeners were even more sensitive than the other listener groups to the association of /θ/ with English. See the discussion for additional analysis of how the different listener groups categorized stimuli with /θ/.

[4] While vowels can be described as differing from one another along (minimally) three continuous dimensions (F1, F2, and duration), there can in fact be phonemic or "new" categories across languages. This would be the case, for example, for English listeners perceiving French /y/, which does not exist as a category in English, even though it may initially be confused with English /u/ or French /u/ (Flege, 1987); English listeners treat French /y/ as a language-specific category sooner than they recognize French /u/ as a category unique from English /u/. This, however, is not the case for any Spanish-specific vowel, which are in line with the French /u/-English /u/ relationship.

263 sub-phonemic differences in place of articulation. Two language-specific phonetic segments were
264 chosen for the task, the lateral approximant /l/ and the high back vowel /u/. The lateral approximant is
265 produced as a 'light' [l] at the alveolar ridge in Spanish, while in American English the segment is
266 realized as the 'darker' [ɫ], with an additional closure near the velum, particularly in closed syllables
267 (Recasens, 2004, 2012). The back vowel differs along F2 in English and Spanish: it is fronted to [ʉ]
268 for many speakers of American English and is produced further back, as [u], in Spanish (Bradlow,
269 1995; Clopper et al., 2005; Mendez, 1982).
270

271 **2.1.2 Nonce words**

272 Nonce words were created to test the contributions of specific sounds to listeners' conceptualizations
273 of Spanish and English. All nonce words were disyllabic trochees with either two open syllables (i.e.
274 CVCV) or /l/ in coda position of the first syllable (i.e. CV/l/CV). The CV/l/CV structure was
275 included in the nonce words to provide two phonological contexts for /l/ stimuli that were both
276 permissible in Spanish and in which /l/ was most likely to be velarized to [ɫ] in American English
277 (Recasens, 2012). The inclusion of disyllabic words with stress on the first syllable meant that the
278 second English vowel would be reduced to schwa, resulting in an additional vowel-quality cue
279 beyond the language-specific target segment. However, this strategy was preferred to the
280 development of monosyllabic words for several reasons. Spanish has relatively few monosyllabic
281 words compared to English (cf. Costa & Caramazza, 1999) so monosyllables may be biased towards
282 English responses. The set of possible word-final consonants in Spanish is very small: /ð, s, n, l, ɾ/.
283 Some of these are subject to lenition (/ð/) or aspiration (/s/), or are already included as a language-
284 specific target segment (/l/). Words ending in /ɾ/ are associated with infinitive morphemes, and /ɾ/ is
285 also in free variation with /r/ word-finally. The inclusion of a second syllable and vowel reduction
286 was therefore preferred. Vowel reduction and its potential influence on listeners' language decisions
287 are addressed in the discussion (see Section 4).
288

289 Each nonce word included one language-specific segment that served as a cue to language
290 categorization. The remaining segments in the nonce words exist in both English and Spanish (at
291 least phonemically, as in the case of the English unstressed schwa) and are not expected to differ
292 between the two languages, so that listeners would be obligated to use the target segment for the
293 language categorization decision. The segments identified as common to both English and Spanish
294 were the fricatives /m,f,s,h/[5] and the affricate /tʃ/, which do not differ between the languages in point
295 of articulation or in voicing, and the vowels /i,a/. While /i,a/ are realized somewhat differently in
296 English and Spanish, with the English variants sometimes transcribed as /ij/ and /ɑ/, respectively,
297 these vowels were preferable over others. Mid-vowels are diphthongized in American English, and
298 /u/ was included as a target segment due to the variation in its articulation in English and Spanish.
299 The symbol /i/ is used here to indicate the vowel in Spanish *mi* 'my' /mi/ and English *me*, and /a/ is
300 used to represent Spanish *la* /la/ 'the' and the vowel in English *cot*. Although /a/ is more variable
301 than /i/ across the languages (Bradlow, 1995), it was included to increase the number of possible
302 stimuli.
303

304 For each target segment, eight nonce CVCV and CV/l/CV words were constructed from the set of
305 segments overlapping in English and Spanish. Each nonce word was a possible, but non-existent,
306 word in both English and Spanish, and all words ended with /a/, which was reduced to [ə] in the
307 English stimuli. See Table 1 for the set of stimuli containing language-specific phonemes and Table 2

---

[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; see Section 2.1.3.

308   for the set of stimuli containing language-specific phonetic segments. One phonemic stimulus, *racha*,
309   was identified as a real Spanish word meaning 'gust of wind' after the study had been completed, so
310   it was excluded from the following analyses. The English nonce word /ɹatʃə/ was also removed due
311   to its similarity to the Spanish *racha* /ratʃa/, since bilingual listeners may have interpreted this
312   stimulus as the Spanish word *racha* produced with an English accent and not as a uniquely English
313   word.
314

315   2.1.3 Stimuli recordings and speaker

316   A single speaker was chosen to record both English and Spanish stimuli, and this was crucial to the
317   experimental task. A single speaker was preferred over recording two monolinguals to avoid voice
318   being a cue to language, and using natural productions of the stimuli ensured there were no acoustic
319   artefacts from splicing or otherwise manipulating segments within a word frame. Using natural
320   productions from a single talker also permitted the selection of the desired segments as target
321   segments, regardless of difficulties isolating them (e.g. with the English /ɹ/).
322

323   Since it was also important for the stimuli to lack any language-specific cues, or accent, beyond the
324   controlled target segment, care was taken to recruit a balanced Spanish-English bilingual who
325   produced both languages as natively as possible. The chosen talker was a 37-year-old Spanish-
326   English bilingual who was born and raised in Colombia until the age of 7 at which point he moved to
327   the state of New York with his family. He continued to speak Spanish at home in New York, and as
328   an adult he moved to Texas for graduate school, during part of which he lived in Guatemala and
329   Spain to conduct research. While most of his current daily interactions were in English, he also used
330   Spanish on a daily basis with his family and frequently for translating and interpreting professionally
331   at work. An accentedness rating study was conducted to ensure that the talker's English and Spanish
332   productions sounded native-like to native English and native Spanish speakers, respectively. In both
333   languages, the talker was rated as native-like as other talkers who grew up as monolingual speakers
334   of each language. See the appendix for a complete description of the accentedness ratings.
335

336   The English and Spanish nonce words were recorded in separate sessions to further ensure minimal
337   cross-linguistic transfer. The recordings took place in a sound-attenuated booth using a MOTU
338   UltraLite-mk3 Hybrid recorder at a sampling frequency of 44.1 kHz (16 bit). The talker repeated
339   each nonce word three times so that the clearest repetition could be chosen. The words were written
340   in English and Spanish orthography (e.g. English *leefuh* for [ɬifə] and Spanish *chirra* for /tʃira/) and
341   not in the International Phonetic Alphabet (IPA), so for some items the talker was coached to arrive
342   at the intended pronunciation. The pitch contours were manipulated to match a naturally-produced
343   token with a falling contour using Praat (Boersma & Weenink, 2012). The beginning and end points
344   of the F0 contours were set to 170Hz and 124Hz to match the values of model token. The intervening
345   pitch points were interpolated between the two end points.

346   2.2   Participants

347   Participants (n=53) were recruited through the Department of Linguistics subject pool and received
348   course credit for their participation. To supplement the subject pool participants with the listeners
349   who had the needed language backgrounds, the remaining Spanish-English bilinguals, both early and
350   late (n=27) were recruited through the University of Texas Events Calendar. These participants were
351   paid $10/hour for their time.
352

353 Listeners completed a language history questionnaire (Chan, 2014) that included questions about
354 participants' biographical information, the places they had lived and for how long, their language
355 exposure and proficiency, and their language(s) of education. Based on their responses to the
356 questionnaire, participants were divided into three groups: monolingual English speakers with
357 minimal or no exposure to Spanish (Monolingual), Spanish-English bilinguals from the U.S. who
358 acquired both languages in early childhood (Early Bilinguals), and Spanish-English bilinguals from
359 Spanish-speaking countries who acquired English as adults (Late Bilinguals). Participants who did
360 not fit into one of these groups were not included in the final sample (n=24). See Table 3 for a
361 summary of participant characteristics.

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

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

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

390 **2.3 Procedure**

391 Participants completed the nonce-word categorization experiment in the UT Sound Lab in the
392 Department of Linguistics at the University of Texas at Austin. The study was approved by the
393 Institutional Review Board at UT Austin, and the experimenter obtained written informed consent
394 from the participant before beginning the study, in accordance with the IRB's recommendations.
395 Listeners answered an online language history questionnaire and were tested for normal hearing,
396 followed by the categorization experiment.

398 Listeners performed the language categorization task in a sound-attenuated booth on a PC running E-
399 Prime 2.0 (Psychology Software Tools, 2010). Listeners wore Sennheiser XX headphones and were

9

400 oriented to the serial response button box (Psychology Software Tools, 2003). Participants were
401 instructed to place the index and middle fingers of their dominant hand on the two leftmost buttons,
402 which were labeled with "ENG" and "SPAN," the order of which was counterbalanced across
403 participants. The language that corresponded to each button was also presented on the computer
404 screen, e.g. "ENGLISH" appeared on the left side of the screen for the group of participants who
405 used the left button to indicate English words. Listeners began with a practice block in which they
406 read instructions presented on-screen and decided if each word sounded more like English or more
407 like Spanish. The practice block included 20 real words (10 English, 10 Spanish).
408
409 After the practice block, the test portion began. At test, listeners were told they would hear "snippets
410 of speech that were taken out of longer recordings while the speaker was talking in either English or
411 Spanish, " and they were asked to decide if what they heard sounded more like it came from the
412 English recording or the Spanish recording. This wording and context was provided after piloting
413 indicated that some listeners had the impression that they were hearing accented productions instead
414 of words from two languages. To avoid this confusion between accent and language, the
415 categorization task was rephrased to ask about the language being used to produce the word.[6]
416 Listeners categorized the 56 nonce words (listed in Tables 1 and 2) eight times, and stimuli were
417 randomized within each of the eight blocks, for a total of 448 trials. There was a one second pause
418 between a listener's response and the onset of the audio for the next stimulus. Reaction time (RT)
419 was calculated from the onset of the audio file, and categorization decision and RT were recorded for
420 each trial.

3   Results

423 Categorization decision (Spanish or English) and reaction time (RT) were recorded for each trial.
424 Decisions were coded as accurate if words containing the English-specific phoneme /ɹ/ or /θ/ or the
425 English variants [ɫ] or [ʉ] were classified as English and if words with the Spanish-specific phoneme
426 /r/ or the Spanish variants [l] or [u] were classified as Spanish. Trials with the Spanish stimulus *racha*
427 /ratʃa/ and the English stimulus /ɹatʃə/ were excluded from the analyses (cf. Section 2.1.2). RTs were
428 calculated by subtracting the length of the stimulus .wav file from the time calculated by E-Prime
429 between trial onset and button press. This ensured that the RTs analyzed here reflected the length of
430 time for the listener to make a categorization decision, after hearing the end of the stimulus word.
431 Trials with RTs less than 200ms (n=665; 1.9%) were discarded as spurious responses. RTs were log-
432 transformed from milliseconds to normalize the distribution of responses for the regression analyses.
433 Less than 0.5% of responses exceeded 5000ms and the distance of these from the mean was reduced
434 in the log transformation. Trials more than three standard deviations above or below a participant's
435 log-transformed mean were excluded as outliers (n=228; 0.7%). The spurious responses and outliers
436 accounted for 2.6% of all trials (n=893), after *racha* and the English /ɹatʃə/ were removed. The
437 following analyses include the remaining 33667 trials (Monolinguals: n=16800; Early Bilinguals:
438 n=7441; Late Bilinguals: n=9426). Accuracy (correct, incorrect) and log-transformed RT were
439 submitted to separate regression analyses, which were analyzed using Bayesian inference with the
440 *glmer2stan* package (v0.995) in R (v3.2.2) to interface with Stan via RStan (v2.8.2).

---

[6] This phrasing invites the possibility that listeners may have looked for other patterns in the stimuli to make their categorization decisions, such as the appearance of language-specific morphemes in the nonce words. All nonce words did end in /a/, which is the Spanish morpheme for feminine adjectives (e.g. *rojo* /roho/ 'red-MASC' vs. *roja* /roha/ 'red-FEM') and is also one of the morphemes for third-person singular (e.g. *habla* /abla/ 'speaks-3SG'). However, since all nonce words uniformly ended in /a/, it is not a feature that distinguishes some stimuli from others. See Section 4 for discussion of other potential language-specific properties of the nonce words.

441   **3.1   Acoustic analyses**

442   Segmental properties of each stimulus were measured using Praat to ensure that the Spanish and
443   English productions differed in the expected dimensions. The duration and first three formants of
444   both vowels of each stimulus were measured, and the same measures were taken for the /l/ variant in
445   the stimuli containing an English or Spanish /l/. Formant measurements were taken at the vowel
446   midpoint and at 25% and 75% through the vowel. Recall that the vowels /i,a/ were used in the first
447   vowel position of the disyllabic nonce words to create a sufficient number of non-word stimuli, and
448   the second vowel (V_2) of each nonce word was realized as the full-vowel [a] in Spanish words and as
449   the reduced [ə] in English words. The Spanish [u] and English [ʉ] segments were target vowels
450   representative of phonetic cues. The acoustic properties of the segments are reported in Table 4: in
451   (A) are reported the mean duration and formant values for the English and Spanish productions of the
452   non-target vowels, and in (B) are the measurements of the language-specific variants of the target
453   segments /l,u/. Formant values are the mean of the measurements taken at the midpoint of each
454   vowel. Standard deviations are included in parentheses.

456   In order to test whether the English and Spanish variants were distinct from each other, the
457   concordance statistic (c-statistic) of a logistic regression model was analyzed. The c-statistic is the
458   proportion of outcomes that are correctly predicted by the fitted model. For each vowel, a logistic
459   regression model was constructed in R (RStudio 0.99.489) using the *rms* package (v4.2-1) with
460   language (English, Spanish) as the dependent variable and the duration and midpoint measures of F1
461   and F2 as fixed effects. Measurements were centered and scaled, and duration was removed from the
462   model where singularity remained. The model for English and Spanish /l/ additionally included the
463   midpoint measure of F3 as a fixed effect. Constructing such a model for the c-statistic was preferable
464   to testing for differences between each fixed effect separately since listeners hear the multiple
465   acoustic cues at once; that is, listeners may attend to differences in all three dimensions (F1, F2, and
466   duration), so all three should be considered together when determining if the sounds were distinct in
467   the two languages.

469   For the two target segments that were measured, /l/ and /u/, it was expected that the formants and the
470   duration of the segment would be sufficient to distinguish the English and Spanish variants. The
471   model with these three main effects as well as the midpoint of F3 made perfect discrimination
472   between the English [ɫ] and the Spanish [l] (C=1.000). For English [ʉ] and Spanish [u], the duration
473   variable was removed to avoid singularity, and the model with the midpoints of F1 and F2 was also
474   highly successful (C=0.969).

476   The other three segments were the two vowels /i,a/, which were used in the first syllables of the
477   nonce words, and the final vowel of the nonce words. The initial model for /i/, with duration and the
478   midpoint measurements for F1 and F2, produced a c-statistic of 0.681, which represents a moderately
479   good fit to the differences in /i/ in English and Spanish words, but which falls short of the clear
480   distinction between the phonetic variants described above. For /a/ in the position of nucleus of the
481   first syllable, the model was highly successful for discrimination (C=1.000). Finally, the model for
482   the second (unstressed) vowel in the nonce words fit well (C=0.853). The acoustic distance between
483   English and Spanish /a/ in stressed and unstressed positions, as well as those between the /i/ variants,
484   was expected (cf. Bradlow, 1995); see Section 4 for a discussion of how the accuracy and RT results
485   should be understood in light of these differences.

486   **3.2   Accuracy analysis**

487 The mean accuracy score of each group for each stimulus type is presented in Table 5. The accuracy
488 results were analyzed using a Bayesian mixed effects logistic regression model with listener language
489 group (Monolingual, Early Bilingual, Late Bilingual), stimulus language (English, Spanish), and
490 stimulus type (phonemic, phonetic) as fixed effects and participant and stimulus word as random
491 intercepts. The models were fitted via a Markov Chain Monte Carlo procedure using STAN (Gelman,
492 Lee & Guo, 2015). Model comparison was performed using the Deviance Information Criterion
493 (DIC; Spiegelhalter et al., 2002). A model with a three-way interaction among the fixed effects
494 provided an improved fit over models with two-way interactions or with only main effects (see Table
495 6 for the model summary). The reference group, reflected in the model intercept, represents the
496 accuracy of Monolinguals categorizing stimuli with an English phoneme. The fitted log odds of
497 accuracy for each stimulus language and listener language group are plotted in Figure 1, with the
498 phonemic cues in the left panel and the phonetic cues in the right panel. The error bars represent the
499 95% Bayesian credible intervals.
500

501 ### 3.2.1 Comparing Spanish and English phonemic and phonetic cues

502 Overall, listeners responded more accurately to Spanish cues than to English cues, and to phonemic
503 cues than to phonetic cues. The difference between the languages was greater for phonemic cues than
504 for phonetic cues. The Spanish phoneme was categorized more accurately than the English phonemes
505 (Monolinguals: β=2.242, posterior SD=0.459, $p<0.0001$; Early Bilinguals: β=2.019, posterior
506 SD=0.484, $p<0.0001$; Late Bilinguals: β=1.556, posterior SD=0.491, $p<0.001$), and the Spanish
507 phonetic cues were also categorized more accurately than the English phonetic cues (Monolinguals:
508 β=1.680, posterior SD=0.367, $p<0.0001$; Early Bilinguals: β=1.292, posterior SD=0.373, $p<0.001$;
509 Late Bilinguals: β=1.120, posterior SD=0.372, $p<0.001$). The Early Bilinguals trended towards
510 categorizing the English phonemic cues more accurately than the English phonetic cues (β=0.448,
511 posterior SD=0.358, $p=0.09$). The Late Bilinguals categorized English phonemic cues significantly
512 better than English phonetic cues (β=0.922, posterior SD=0.358, $p<0.01$). All groups categorized the
513 Spanish phonemic cue more accurately than the Spanish phonetic cue (Monolinguals: β=0.763,
514 posterior SD=0.451, $p<0.01$; Early Bilinguals: β=1.175, posterior SD=0.477, $p<0.0001$; Late
515 Bilinguals: β=1.359, posterior SD=0.480, $p<0.0001$).
516

517 ### 3.2.2 Comparing listener groups

518 The three listener groups responded very similarly within each segment type, with the exception of
519 the categorization of nonce words with an English phoneme. For the English phonemes,
520 Monolinguals and Early Bilinguals responded less accurately than the Late Bilinguals (vs.
521 Monolinguals: β=1.014, posterior SD=0.236, $p<0.0001$; vs. Early Bilinguals: β=0.715, posterior
522 SD=0.294, $p<0.05$). There were no group differences in the English phonetic cue conditions, and
523 there were also no significant group differences in response to the Spanish phonemic or the Spanish
524 phonetic cues.

525 ### 3.3 Reaction time analysis

526 The mean RTs (in milliseconds) of each group for correct responses to each stimulus type is
527 presented in Table 7. Log-transformed RTs were analyzed using a Bayesian mixed effects linear
528 regression model with listener language group (Monolingual, Early Bilingual, Late Bilingual),
529 stimulus language (English, Spanish), stimulus type (phonemic, phonetic), and accuracy (correct,
530 incorrect) as fixed effects. Participant and stimulus word were included as random intercepts. These
531 models were also fitted via a Markov Chain Monte Carlo procedure using STAN, as described above.

532   Testing for a significant effect of categorization accuracy evaluated the possibility that listeners' RTs
533   were unaffected by the accuracy of the categorization decision. A model with the same three fixed
534   effects as the accuracy model – listener group, stimulus language, and stimulus type – was
535   significantly improved by adding accuracy as a fixed effect. RTs thus significantly differed between
536   accurate and inaccurate trials, and subsequent models calculated separate betas for each type of trials.
537   The model with a four-way interaction among the fixed effects provided a better fit than models with
538   only main effects, with two-way interactions, or with three-way interactions. See Table 8 for the
539   model summary. The reference group, reflected in the model intercept, represents the log RT of
540   inaccurate responses by Monolinguals categorizing stimuli with an English phoneme. The fitted log
541   RT for correct responses to each target segment and listener language group are plotted in Figure 2.
542   The error bars represent 95% Bayesian credible intervals. The following sections report the results of
543   correct trials from the four-way interaction and the differences between correct and incorrect
544   responses.

3.3.1 Comparing Spanish and English phonemic and phonetic cues

547   For the four cue types, there were few significant differences in RTs. The only differences appeared
548   for the Spanish cues: the Early Bilinguals trended towards faster RTs for the Spanish phonemic cue
549   compared to the Spanish phonetic cues (β=0.144, posterior SD=0.073, p=0.08), and the Late
550   Bilinguals responded significantly faster to the Spanish phoneme than to the Spanish phonetic cues
551   (β=0.164, posterior SD=0.073, p<0.05). There was no difference between the Spanish categories for
552   Monolingual listeners. The differences in RT between the English phonemic cues and the English
553   phonetic cues did not reach significance for any listener group. There were also no differences in RTs
554   between the English and Spanish phonemic cues or between the English and Spanish phonetic cues.

3.3.2 Comparing listener groups

557   The pattern of differences in RTs among the listener groups was mostly constant across segments:
558   Monolinguals and Early Bilinguals responded with similar RTs, and both these groups were faster
559   than Late Bilinguals. For the Spanish phonemic cue, there was no difference between Monolinguals
560   and Early Bilinguals, and both groups were significantly faster than Late Bilinguals (vs.
561   Monolinguals: β=0.252, posterior SD=0.100, p<0.01; vs. Early Bilinguals: β=0.238, posterior
562   SD=0.124, p<0.05). For English phonemes, Monolinguals and Early Bilinguals also responded faster
563   than Late Bilinguals (vs. Monolinguals: β=0.227, posterior SD=0.100, p<0.01; vs. Early Bilinguals:
564   β=0.176, posterior SD=0.124, p<0.05), and there was again no difference between the Monolinguals
565   and Early Bilinguals. For trials with Spanish phonetic cues, Monolinguals and Early Bilinguals
566   responded faster than Late Bilinguals (vs. Monolinguals: β=0.320, posterior SD=0.099, p<0.0001; vs.
567   Early Bilinguals: β=0.258, posterior SD=0.123, p<0.01), and there was no differences in RTs for the
568   Monolinguals and Early Bilinguals. Finally, for nonce words with an English phonetic cue,
569   Monolinguals and Early Bilinguals were also significantly faster than Late Bilinguals (vs.
570   Monolinguals: β=0.294, posterior SD=0.100, p<0.0001; vs. Early Bilinguals: β=0.182, posterior
571   SD=0.123, p<0.05), and Monolinguals trended faster than Early Bilinguals (β=0.112, posterior
572   SD=0.109, p=0.06).

3.3.3 Comparing accurate and inaccurate trials

575   Overall, RTs for correct responses were faster than for incorrect responses. For Monolinguals, this
576   difference reached significance for all four types of nonce words (English phonemic: β=0.178,
577   posterior SD=0.25, p<0.01; Spanish phonemic: β=0.244, posterior SD=0.74, p<0.01; English

578   phonetic: β=0.187, posterior SD=0.023, *p*<0.01; Spanish phonetic: β=0.224, posterior SD=0.035,
579   *p*<0.01). For Early Bilinguals, correct trials were faster than incorrect trials for the Spanish cues
580   (phonemic: β=0.374, posterior SD=0.133, *p*<0.0001; phonetic: β=0.297, posterior SD=0.052,
581   *p*<0.001), but there was no difference for the English cues. For Late Bilinguals, the difference
582   between correct and incorrect trials was significant for both kinds of Spanish cues (phonemic:
583   β=0.157, posterior SD=0.131, *p*<0.05; phonetic: β=0.267, posterior SD=0.047, *p*<0.01) and for the
584   English phonemes (β=0.310, posterior SD=0.040, *p*<0.001), but not for the English phonetic cues.
585
586   The results of the accuracy and RT analyses are summarized in Tables 9 and 10. Table 9 summarizes
587   how Spanish and English stimuli were categorized by each listener group (A) and how the listeners
588   categorized the different stimuli classes (B). Table 10 summarizes how the listener groups compared
589   within each stimulus type. The "=" is used to illustrate differences that were not significant, and the
590   ">" and "<" indicate significant differences. The "»" and "«" represent differences that approached
591   significance.

592   **4   Discussion**

593
594   The current study tested the sensitivity of monolingual and early and late bilingual adults to
595   language-specific sounds in a nonce-word categorization task to determine which segments listeners
596   are most sensitive to and how language experience influences listeners' sensitivity. Overall, listeners
597   very accurately categorized phonemic cues and Spanish cues but struggled more with English cues
598   and phonetic cues. There was also a significant interaction between stimulus language and cue type,
599   with the difference between phonemic and phonetic cues greater for Spanish than for English. This
600   difference also significantly interacted with listener group, such that the difference between Spanish
601   and English phonemic cues and Spanish and English phonetic cues was smaller for Late Bilinguals
602   and greater for Early Bilinguals. The categorization accuracy of the Monolinguals, Early Bilinguals,
603   and Late Bilinguals was very similar overall, with the only significant difference between groups
604   occurring for the English phonemic cues, which Late Bilinguals categorized more accurately than the
605   other groups. The response times for Monolingual and Early Bilingual listeners were comparable,
606   and both of these groups responded more quickly than Late Bilinguals for all cue types. Based on
607   models of native and second-language speech perception (Best 1991; Flege 1987, 1995), we
608   predicted a greater sensitivity to phonemic properties of lexical and language representations than to
609   phonetic cues. The results here provide new evidence supporting these predictions in a language-
610   decision task with word-length stimuli: early and late bilinguals can use both kinds of segments for
611   categorization, but they were more sensitive to phonemic cues than phonetic cues. Unexpectedly, all
612   listeners were more sensitive to Spanish-specific cues than English-specific cues. Finally, language
613   background had only a limited effect on listeners' access to these representations.

614
615   Overall, there were no differences between the Monolingual and Early Bilingual listeners. The Late
616   Bilinguals were as sensitive to some cues as the other two listener groups, and there was limited
617   evidence that Late Bilinguals might even be more sensitive to some cues. The Late Bilinguals also
618   responded significantly more slowly than the other groups, so it is possible that there was a speed-
619   accuracy trade-off for these listeners; however, it only appeared for the Late Bilinguals'
620   categorization of English phonemic cues, for which they were significantly more accurate than
621   Monolinguals and Early Bilinguals but also significantly slower. The performance of the
622   Monolinguals and Early Bilinguals reveals that the language representations of the Early Bilinguals,
623   despite their having learned Spanish at home before English, do not differ in the phonemic categories
624   or the phonetic detail encoded in their language representations. This is not to say that our Early

625   Bilinguals would not have shown evidence of their Spanish exposure in other tests, such as
626   production or phoneme identification tasks. The current results do suggest that the ability of Early
627   Bilinguals to generalize about the properties of their native languages and associate phonological
628   properties in particular with each language is not distinct from Monolinguals' awareness of these
629   language-specific properties. This sets our early Spanish-English bilinguals apart from the early
630   Spanish-Catalan bilinguals in Sebastián-Gallés et al. (2005), whose sensitivity to Catalan-specific
631   contrasts was purportedly compromised by their early exposure to Spanish. Rather, the similarity
632   between our responses from Monolinguals and Early Bilinguals supports the language assessment
633   used by Amengual (2014, 2015), in which adults' current language exposure and use seem to
634   override the effect of non-simultaneous early exposure and contribute to their equivalent performance
635   (Gertken et al., 2014). The role of ongoing exposure in addition to and even superseding age of
636   acquisition is also supported by Flege and colleagues who found that among listeners with similar
637   ages of acquisition, greater exposure to, use of, and education in the L1 led to less native-like
638   perception and production (Flege, 1991; Flege et al., 1997b; Flege & MacKay, 2004) and
639   grammaticality judgments (Flege et al., 1999b) in the L2. It is important for future work on the
640   association of language and segments to consider dominance and exposure to each language as
641   factors influencing cross-linguistic speech perception in context.
642
643   While we only indirectly assessed the bilingual listeners' language dominance and exposure though
644   the language background questionnaire, the Monolingual and Early Bilingual groups did share some
645   commonalities. Examining those further may assist in understanding the similarities in their
646   categorization decisions and potentially why the Late Bilinguals outperformed these groups in the
647   English phoneme trials. Our Early Bilinguals live and study immersed in their (chronological) L2,
648   English, and as a result, they may have the same awareness of the generalizability of the
649   phonological properties of each of their languages as the monolingual speakers who know only
650   English. The difference between the two bilingual groups for the English phoneme category, on the
651   other hand, may reflect variation in dominance, exposure, or the method of English acquisition. Most
652   of the Early Bilinguals (11 of 18) learned English when they began kindergarten, and language
653   instruction at this age is likely to be much less explicit than the middle and high school foreign-
654   language classrooms in which the Late Bilinguals learned English. Even where there are parallels in
655   L2 teaching at these ages, the experience of English language learning is much more recent for the
656   Late Bilinguals than for the Early Bilinguals, and attending foreign language classes, practicing the
657   language, and laboring to master the rules of and achieve proficiency in the L2 may lead the Late
658   listeners to a greater metalinguistic awareness about properties of the language (Dąbrowska & Street,
659   2006), including increased sensitivity to language-segment associations. The study of phonological
660   and metalinguistic awareness in adults has been limited to literacy and disorders (e.g., Pennington et
661   al., 1990), although additional work with children has investigated bilingualism (Bialystok, 2001;
662   Bruck & Genesee, 1995) and literacy development (e.g. Anthony & Francis, 2005). It is therefore
663   unclear how metalinguistic awareness and cue sensitivity may affect cross-language speech
664   perception in adults. The current findings suggest that the listeners who acquired an L2 in early
665   childhood may lack the metalinguistic awareness evident in the Late Bilingual listeners, or that this
666   sensitivity may decline into adulthood. Over time and as English proficiency increases, young
667   bilingual listeners may lose their initial phonological sensitivity and may later categorize segments
668   no differently than Monolingual adults who acquired their only language in infancy.
669
670   Given the potential differences in language teaching and language learning in kindergarten and high
671   school, the Late Bilinguals may have increased sensitivity to some language-specific phonological
672   properties due to the circumstances of their bilingualism and not necessarily due to the age of

673 acquisition. In fact, this formal training may also explain why there were group differences for the
674 English phonemic cues but not for the English phonetic ones. Phonemic differences across languages
675 may get more attention in foreign-language classes than subsegmental differences between categories
676 shared by the two languages. Just as the phonetic cues were more difficult for listeners in general,
677 Late Bilinguals may not have had the same metalinguistic instruction about English phonetic
678 differences and so may have been less able to associate those cues with English, even though this was
679 possible for the phonemic cues. Future work on cue sensitivity should work to separate recency of
680 language acquisition from method of language acquisition to disentangle how these factors influence
681 phonological awareness and especially awareness of subsegmental differences. For example, Early
682 Bilinguals may be more sensitive to English phonemic cues during earlier stages of English
683 acquisition, and we might also expect listeners who acquire a language without formal classes (e.g.
684 from being immersed in a new community) to be less sensitive to language-specific cues, especially
685 phonemes, than listeners who study the language in a formal setting.
686
687 The consistency of categorization accuracy across the three listener groups suggests that language
688 experience was less important than cue salience in this task. Phonemic cues were more accurately
689 categorized than phonetic cues, for both English and Spanish, supporting the parallel distinction
690 made between new and similar phones in Flege (1987, 1995)'s Speech Learning Model (SLM). In
691 this model, second language learners create independent categories for sounds judged to be "new"
692 (unique to the L2 and not present in the L1), which facilitates the production and perception of such
693 sounds. Phones that are recognized as similar to existing L1 segments are discriminated less well if
694 no new category is established for them. The phonemes in the present task may be like the SLM's
695 new phones, even for the Monolinguals who have not acquired Spanish, and as such they are
696 immediately recognizable as language-specific sounds (Best, 1991), which leads to more accurate
697 categorization. In contrast, the phonetic cues pattern like the SLM's similar phones, a category for
698 which, according to Best (1991), the L2 or non-dominant language sounds would be mapped to the
699 L1 or dominant-language categories. This would cause more competition in deciding between
700 English or Spanish for the language identity of the word.
701
702 There may have also been an effect of the specific segments included in each category. Since there
703 was only one Spanish-specific phonemic cue included, the Spanish phoneme category in fact
704 represents listener responses to a single sound, the Spanish trill /r/, which was easily perceived and
705 strongly associated with Spanish phonology for all three listener groups. The English phoneme
706 category may have been very different in this sense, since it included the English rhotic /ɹ/ and the
707 interdental fricative /θ/. Fricatives and interdentals in particular are acquired late by English-learning
708 children (Clark, 2003; Dodd et al., 2003), and even native-English-speaking adults are susceptible to
709 mishearing /θ/ more than they mishear other segments (Cutler et al., 2004). That is, there may be
710 inherent differences in the perceptual salience of the two English phonemes, irrespective of the
711 strengths of associations between English and each segment. Since only a single Spanish phonemic
712 cue was available and given the asymmetry in salience of the English phonemic cues, future work
713 should more systematically compare a wider range of phonemes in other language pairs to consider
714 whether there may be variability within the phonemic category. However, despite the inherent
715 difficulty of at least the English /θ/, it is even more striking that the Late Bilinguals outperformed the
716 groups that had acquired the English phonemes in childhood. In fact, since the Late Bilinguals may
717 be aware of /θ/ being a phonemic sound in Peninsular Spanish, we might have expected this
718 awareness to cause confusion and thus fewer accurate responses in English phoneme trials for the
719 Late Bilinguals, but just the opposite was the case. This suggests that the absence of this phoneme in
720 the native language and dialects of the Late Bilinguals may have heightened their sensitivity to /θ/.

721 Instead, the difficulty all listeners had responding to the English phoneme category may be motivated
722 by perceptual salience more generally, and future work should further probe variation with each of
723 these cue types.
724
725 The difficulty listeners from all backgrounds experienced in accurately categorizing phonetic cues
726 also requires further investigation. The English [ɫ] is more velarized, i.e. produced with the tongue
727 further back in the oral cavity, than the Spanish [l], while the English [ʉ] is fronted, so the difference
728 between English and Spanish phonetic cues is unlikely to be due to a single property that sets English
729 apart from Spanish, since the English variants differ in opposite directions from the Spanish ones. It
730 may be that listeners hear more variation in English input between lighter or darker /l/ and more or
731 less fronted /u/ across dialects, speakers, and phonological contexts than exists for Spanish [l] and
732 [u]. However, it would be surprising if our monolingual English listeners were also sensitive to the
733 greater consistency of these segments in Spanish, given their lack of exposure to the language.[7]
734 Furthermore, if the variability present in the realization of these sounds in English motivated the
735 difference in accuracy between English and Spanish segments, we should expect a different
736 categorization pattern entirely. A light [l] or a backed [u] may be either from Spanish or English,
737 since these variants exist in many dialects of English, so the Spanish phonetic cues should have
738 received responses more mixed between the languages. It is the darker [ɫ] and fronted [ʉ] that should
739 be unambiguously associated with English, but in fact we find the English cues receive more of a mix
740 of Spanish and English categorization decisions while the Spanish cues are relatively consistently
741 identified as Spanish.
742
743 While every effort was made to create nonce words that were equally plausible in both languages,
744 except for the language-specific target segment, the naturally-produced stimuli used here inevitably
745 carried additional indicators of language. The phonotactic restrictions of Spanish may have meant
746 that the CVCV stimuli were simply more Spanish-like than English-like, even though this word
747 structure is permitted in English. The Spanish-ness of these stimuli is supported by the reactions of
748 participants in two pilot studies; in the first pilot, theoretically congruous stimuli that overlapped
749 English and Spanish in all segments, e.g. /tʃima/, were categorized as Spanish significantly more than
750 English, and in the second pilot (cf. Section 2.3), listeners reported confusion about whether words
751 were English or English-accented Spanish. In the present study, listeners from all three language
752 backgrounds were able to overcome this potential bias towards Spanish for English: the log odds of
753 responding correctly were significantly above 0 (chance performance) in all four cases, including for
754 the English segments. Therefore, listeners showed sensitivity to the English-ness of the English cues
755 even if the word structure is less common in English than it is in Spanish. Furthermore, Monolinguals
756 might not be expected to suffer from such a potential bias as much as the bilingual groups, since the
757 Monolinguals do not have representations of Spanish phonotactics against which to judge the nonce
758 word forms. Instead, their categorization patterns were in line with the bilingual groups'. Why, then,
759 might listeners have been less accurate in categorizing stimuli with English cues?
760
761 The difficulties that persisted for English cues are especially interesting given that the naturally
762 produced nonce words used here likely contained multiple phonetic cues to language. As was
763 mentioned in Section 2.1.2, the disyllabic nature of the nonce words meant that the unstressed vowel
764 /a/ in the second syllable was reduced to [ə] in the English words; therefore, all the English nonce
765 words contained both a language-specific target segment (e.g. /ɪ/) and the reduced vowel.

---

[7] We would additionally have to assume that exposure to Spanish-accented English is sufficient for the development of phonological categories that accurately reflect the properties of these categories as they are realized in Spanish.

766   Furthermore, the acoustic analyses of the /i/ and /a/ vowels in the first syllable of the nonce words
767   indicate that there were also language-specific differences in the productions of these non-target
768   segment (cf. Section 3.1). But again, despite these potential additional cues to language, listeners
769   categorized the English-specific segments less accurately than Spanish cues. Given the more accurate
770   performance of the Late Bilinguals than the other groups for English phonemes we might be tempted
771   to conclude that the Late Bilinguals were better able to use these supplementary language-specific
772   cues than their peers, but their accuracy did not significantly differ from the Monolinguals and Early
773   Bilinguals in the English phonetic condition. If the Late Bilinguals were more sensitive to the
774   English-ness of the nonce word filler vowels in the phonemic condition, where they outperformed
775   their peers, it is unclear why they wouldn't have been able to make use of the additional cues in the
776   English phonetic words.
777
778   Moving forward, it will continue to be important to consider the contributions of language-specific
779   segments in the context of a word, as discussed earlier, since listeners may use different processing
780   strategies and respond to the same sound categories differently when presented in isolation and in
781   context. To this end, it will be necessary to also involve language pairs for which there are more
782   language-specific contrasts and a wider variety of segments to be studied than those available for
783   English and Spanish. All phonemic cues used here were consonants, with a necessary but
784   confounding overreliance on the differences in rhotics across the languages. Similarly, the
785   mispronunciation studies in Spanish and Catalan by Sebastián-Gallés et al. (2005) and Amengual
786   (2014, 2015) were restricted in scope, and focused only on vowels. Contrasting a language pair that
787   differs more significantly in both consonants and vowels at the phonemic and phonetic levels would
788   provide the evidence needed to further test the conclusions drawn from the present results.
789
790   Finally, the current study speaks to other related speech perception phenomena, namely foreign-
791   accent detection. To date, our knowledge of the perception of foreign-accented speech has been
792   largely based on monolingual listeners, but the findings of the present study support the inclusion of
793   listeners actually proficient in, and not just familiar with, the L1 of the accented speech. Based on our
794   results, bilingual listeners might be expected to identify accented talkers as well as monolingual
795   listeners, and if the foreign accent contains non-native phonemic cues like those tested here, late
796   bilinguals might be more sensitive to accented speech than other listeners. Benefits of exposure to
797   accented speech have likewise been reported for categorizing sentences produced in regional
798   (Clopper & Pisoni, 2004, 2007) and foreign (Vieru et al., 2011) accents. High-exposure listeners also
799   processed foreign-accented words faster and more accurately than low-exposure listeners (Witteman
800   et al., 2013), so listeners with experience can attend to the relatively few cues available in a single
801   word. Even so, given the nature of the naturally-produced words and sentences used in these studies,
802   it is not clear what cues the listeners with greater exposure were using in their processing, or which
803   cues the less-experienced listeners were not able to capitalize on. We might expect foreign-accented
804   speech to contain more of the difficult phonetic cues that most challenged our Monolingual listeners,
805   and this could explain the performance of the low-familiarity listeners in Vieru et al. (2011) and
806   Witteman et al. (2013). The contribution of phonemic and phonetic cues to foreign-accented speech
807   detection could be tested by controlling these cues in real words, as was done in the present study
808   with nonce words, to determine if real foreign-accented words with deviant phonemic cues are in fact
809   categorized more easily than words with phonetic cues. Furthermore, the processing of foreign-
810   accented speech may also be influenced by the presence of phonemic and phonetic cues. Since
811   phonetic cues are less clearly linked to a specific language and listeners of all backgrounds are less
812   sensitive to deviations in phonetic cues, speech that contains only phonetic deviations (e.g., from

813   more proficient L2 speakers) may be easier to process than speech that also contains phonemic
814   deviations.
815
816   In summary, the results of the nonce-word categorization task indicate that listeners are better able to
817   use Spanish-specific cues than English-specific cues and that listeners categorize phonemic cues,
818   modeled on Flege's (1987, 1995) "new" sounds, better than phonetic cues. This distinction supports
819   similar divisions made between native and non-native sounds in speech perception literature more
820   generally and for second language acquisition in particular (Best, 1991; Flege, 1987, 1995). Our
821   findings also show similarities in categorization patterns across listener groups, in parallel with the
822   work of Flege et al. (1989) and Mack (1999) on early bilinguals' phoneme discrimination, and even
823   the late bilinguals categorized the nonce-word stimuli like early learners. The early bilinguals'
824   sensitivity to English-specific cues was not degraded by their early exposure to and proficiency in
825   Spanish, deviating from the conclusions of Sebastián-Gallés et al. (2005), but their knowledge of
826   Spanish also did not improve the accuracy of their language classification decisions for Spanish
827   nonce words, which might have been expected given the advantages for high-exposure listeners in
828   accent categorization tasks (e.g. Witteman et al., 2013). Such facilitation was observed for the late
829   bilinguals for words with English phonemic cues, although the late bilingual listeners responded
830   significantly more slowly than the other groups for all cues. The study of additional language pairs
831   will strengthen the conclusions we make here about differences in listener sensitivity to language-
832   specific phonemic and phonetic cues by providing additional segments and contrasts and allowing for
833   systematic comparisons, e.g. of consonantal and vowel contributions to each category. The finding
834   that listeners use phonemic cues more successfully than phonetic cues in word contexts should shape
835   future directions of work on the perception of foreign-accented speech and cross-language speech
836   perception.

5   Acknowledgments

6   References

848   Amengual, M. (2014). The perception and production of language-specific mid-vowel contrasts:
849   Shifting the focus to the bilingual individual in early language input conditions. *International Journal*
850   *of Bilingualism*. Advance online publication. doi: 10.1177/1367006914544988.
851
852   Amengual, M. (2015). The perception of language-specific phonetic categories does not guarantee
853   accurate phonological representations in the lexicon of early bilinguals. *Applied Psycholinguistics.*
854   Advance online publication. doi: 10.1017/S0142716415000557.
855
856   Anthony, J.L., & Francis, D.J. (2005). Development of Phonological Awareness. *Current Directions*
857   *in Psychological Science*, 14 (5), 255-259.
858

859 Baker, R.E., Bonnasse-Gahot, L., Kim, M., Van Engen, K.J., and Bradlow, A.R. (2011). Word
860 durations in non-native English. *Journal of Phonetics,* 39 (1), 1–17.
861
862 Best, C.T. (1991). The Emergence of Native-Language Phonological Influences in Infants: A
863 Perceptual Assimilation Model. *Haskins Laboratories Status Report on Speech Research*, SR-
864 107/108, 1-30.
865
866 Bialystok, E. (2001). 10. Metalinguistic aspects of bilingual processing. *Annual Review of Applied*
867 *Linguistics*, 21, 169-181.
868
869 Boersma, P., & Weenink, D. (2012). Praat: doing phonetics by computer [Computer program].
870 Version 5.3.32, retrieved 17 October 2012 from http://www.praat.org/.
871
872 Bradlow, A. (1995). A comparative acoustic study of English and Spanish vowels. *Journal of the*
873 *Acoustical Society of America*, 97 (3), 1916-1924.
874
875 Bradlow, A., Clopper, C., Smiljanic, R., & Walter, M.A. (2010). A perceptual phonetic similarity
876 space for languages: Evidence from five native language listener groups. *Speech Communication*, 52,
877 930-942.
878
879 Bruck, M., & Genesee, F. (1995). Phonological awareness in young second language learners.
880 *Journal of Child Language*, 22, 307-324.
881
882 Carlson, M.T., Goldrick, M., Blasingame, M., & Fink, A. (2015). Navigating conflicting phonotactic
883 constraints in bilingual speech perception. *Bilingualism: Language and Cognition*. Advance online
884 publication. doi: 10.1017/S1366728915000334.
885
886 Chan, C.L. (2014). NUsub-db: Northwestern University Subject Database [Web Application].
887 Department of Linguistics, Northwestern University. https://babel.ling.northwestern.edu/nusubdb2/.
888
889 Clark, E.V. (2003). *First language acquisition*. Cambridge: Cambridge University Press.
890
891 Clopper, C.G., & Pisoni, D.B. (2004). Homebodies and army brats: Some effects of early linguistic
892 experience and residential history on dialect categorization. *Language Variation and Change*, *16*,
893 31–48.
894
895 Clopper, C.G., & Pisoni, D.B. (2007). Free classification of regional dialects of American English.
896 *Journal of Phonetics*, *35*, 421–438.
897
898 Clopper C. G., Pisoni D. B., & de Jong K. (2005). Acoustic characteristics of the vowel systems of
899 six regional varieties of American English. *Journal of the Acoustical Society of America,* 118, 1661–
900 1676.
901
902 Costa, A., & Caramazza, A. (1999). Is lexical selection in bilingual speech production language-
903 specific? Further evidence from Spanish-English and English-Spanish bilinguals. *Bilingualism:*
904 *Language and Cognition*, 2 (3), 231-244.
905

906 Cutler, A., Weber, A., Smits, R., & Cooper, N. (2004). Patterns of English phoneme confusions by
907 native and non-native listeners. *Journal of the Acoustical Society of America*, 116, 3668–3678.
908
909 Dąbrowska, E., & Street, J.A. (2006). Individual differences in language attainment: Comprehension
910 of passive sentences by native and non-native English speakers. *Language Sciences*, 28, 604-615.
911
912 Diehl, R.L., & Walsh, M.A. (1989). An auditory basis for the stimulus-length effect in the perception
913 of stops and glides. *Journal of the Acoustical Society of America*, 85 (5), 2154-2164.
914
915 Dodd, B., Holm, A., Hua, Z., & Crosbie, S. (2003). Phonological development: a normative study of
916 British English-speaking children. *Clinical Linguistics & Phonetics*, 17 (8), 617-643.
917
918 Escudero, P. (2006). The phonological and phonetic development of new vowel contrasts in Spanish
919 learners of English. In *English With a Latin Beat: Studies in Portuguese/Spanish-English*
920 *Interphonology, Studies in Bilingualism*, Vol. 31, ed. B.O. Baptista, and M.A. Watkins (Amsterdam:
921 John Benjamins), 149–161.
922
923 Flege, J. E. (1984). The detection of French accent by American listeners. *Journal of the Acoustical*
924 *Society of America,* 76 (3), 692-707.
925
926 Flege, J. E. (1987). The production of "new" and "similar" phones in a foreign language: evidence
927 for the effect of equivalence classification. *Journal of Phonetics*, 15, 47-55.
928
929 Flege, J.E. (1991). Age of learning affects the authenticity of voice onset time (VOT) in stop
930 consonants produced in a second language. *Journal of the Acoustical Society of America,* 89, 395–
931 411.
932
933 Flege, J.E. (1995). Second Language Speech Learning: Theory, Findings, and Problems. In *Speech*
934 *Perception and Linguistic Experience: Issues in Cross-Language Research*, ed. W. Strange
935 (Timonium, MD: York Press). 233-277.
936
937 Flege, J.E. (2007). Language contact in bilingualism: Phonetic system interactions. In *Laboratory*
938 *Phonology 9*, eds. J. Cole & J. Hualde (Berlin: Mouton de Gruyter). 353-382.
939
940 Flege, J.E., Bohn, O.-S., & Jang, S. (1997a). Effects of experience on non-native speakers'
941 production and perception of English vowels. *Journal of Phonetics*, 25, 437-470.
942
943 Flege, J.E., Frieda, E.M., & Nozawa, T. (1997b). Amount of native-language (L1) use affects the
944 pronunciation of an L2. *Journal of Phonetics*, 25, 169-186.
945
946 Flege, J.E., & MacKay, I.R.A. (2004). Perceiving vowels in a second language. *Studies in Second*
947 *Language Acquisition*, 26, 1-34.
948
949 Flege, J.E., MacKay, I.R.A., & Meador, D. (1999a). Native Italian speakers' perception and
950 production of English vowels. *Journal of the Acoustical Society of America,* 106 (5), 2973-2987.
951
952 Flege, J.E., & Munro, M.J. (1994). The word unit in second language speech production and
953 perception. *Studies in Second Language Acquisition,* 16, 381-411.

954

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

957

958 Flege, J.E., Yeni-Komshian, G.H., & Liu, S. (1999b). Age Constraints on Second-Language
959 Acquisition. *Journal of Memory and Language*, 41, 78-104.

960

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

963

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

967

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

970

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

972

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

975

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

978

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

982

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

985

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

988

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

991

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

994

995 Mendez, A. (1982). Production of American English and Spanish vowels. *Language and Speech*, 25,
996 191-197.

997

998 Munro, M.J., & Derwing, T.M. (2001). Modeling Perceptions of the Accentedness and
999 Comprehensibility of L2 Speech: The Role of Speaking Rate. *Studies in Second Language*
1000 *Acquisition*, 23 (4), 451-468.

1001

1002 Munro, M.J., Derwing, T.M., & Burgess, C.S. (2010). Detection of nonnative speaker status from
1003 content-masked speech. *Speech Communication*, 52: 626-637.

1004

1005 Munro, M.J., Flege, J.E., & Mackay, I.R.A. (1996). The effects of second language learning on the
1006 production of English vowels. *Applied Psycholinguistics*, 17, 313-334.

1007

1008 Nygaard, L.C., & Pisoni, D.B. (1998). Talker-specific learning in speech perception. *Perception
1009 andPsychphysics*, 60, 355–376.

1010

1011 Pennington, B.F., van Orden, G.C., Smith, S.D., Green, P.A., & Haith, M.M. (1990). Phonological
1012 Processing Skills and Deficits in Adult Dyslexics. *Child Development,* 61 (6), 1753-1778.

1013

1014 Pierrehumbert, J.B. (2002). Word-specific phonetics. In *Laboratory phonology, Vol. VII*, eds. C.
1015 Gussenhoven, & N. Warner, (Berlin: Mouton de Gruyter). 101–139.

1016

1017 Pisoni, D.B. (1977). Identification and discrimination of the relative voice onset time of two
1018 component tones: Implication for voicing perception in stops. *Journal of the Acoustical Society of
1019 America,* 61 (5), 1352-1361.

1020

1021 Psychology Software Tools, Inc. E-Prime 2.0. (2010). Retrieved from http://www.pstnet.com.

1022

1023 Recasens, D. (2004). Darkness in [l] as a scalar phonetic property: implications for phonology and
1024 articulatory control. *Clinical Linguistics & Phonetics*, 18 (6-8), 593-603.

1025

1026 Recasens, D. (2012). A cross-language acoustic study of initial and final allophones of [l]. *Speech
1027 Communication,* 54, 368-383.

1028

1029 RStudio Team (2015). *RStudio: Integrated Development for R*. RStudio, Inc., Boston, MA. URL
1030 http://www.rstudio.com/.

1031

1032 Sebastián-Gallés, N., Echeverría, S., & Bosch, L. (2005). The influence of initial exposure on lexical
1033 representation: Comparing early and simultaneous bilinguals. *Journal of Memory and Language,* 52,
1034 240-255.

1035

1036 Sebastián-Gallés, N., & Soto-Faraco, S. (1999). Online processing of native and non-native
1037 phonemic contrasts in early bilinguals. *Cognition,* 72, 111-123.

1038

1039 Shook, A., & Marian, V. (2013). The Bilingual Language Interaction Network for Comprehension of
1040 Speech. *Bilingualism: Language and Cognition,* 16 (2), 304-324.
1041 Spiegelhalter, D.J., Best, N.G., Carlin, B.P., & van der Linde, A. (2002). Bayesian measures of
1042 model complexity and fit (with discussion). *Journal of the Royal Statistical Society*, Series B, 64 (4),
1043 583–639.

1044

1045 Strand, E.A. (1999). Uncovering the role of gender stereotypes in speech perception. *Journal of
1046 Language and Social Psychology*, 18 (1), 86-99.

1047

1048 Strand, E.A., & Johnson, K. (1996). Gradient and visual speaker normalization in the perception of
1049 fricatives. In *Natural language processing and speech technology: Results of the 3rd KONVENS*
1050 *conference, Bielefeld, October 1996,* ed. D. Gibbon (Berlin: Mouton). 14-26.
1051
1052 Tajima, K., Port, R., & Dalby, J. (1997). Effects of temporal correction on intelligibility of foreign-
1053 accented English. *Journal of Phonetics*, 25 (1), 1-24.
1054
1055 Tracy, E.C., Bainter, S.A., & Satariano, N.P. (2015). Judgments of self-identified gay and
1056 heterosexual male speakers: Which phonemes are the most salient in determining sexual orientation?
1057 *Journal of Phonetics*, 52, 13-25.
1058
1059 Vieru, B., Boula de Mareüil, P., & Adda-Decker, M. (2011). Characterisation and identification of
1060 non-native French accents. *Speech Communication*, 53, 292-310.
1061
1062 Weber, A., and Cutler, A. (2006). First-language phonotactics in second-language listening. *Journal*
1063 *of the Acoustical Society of America*, 119 (1), 597–607.
1064
1065 Witteman, M. J., Weber, A., & McQueen, J. M. (2013). Foreign accent strength and listener
1066 familiarity with an accent codetermine speed of perceptual adaptation. *Attention, Perception, and*
1067 *Psychophysics*, 75, 537-556.

1068 **7   Appendix**
1069 To ensure that the stimuli talker's productions were native-like in both languages, an accentedness
1070 rating study was completed. Native English and native Spanish listeners rated the nativeness of the
1071 productions of eight talkers, including the stimuli talker. All talkers recorded Æsop's *The North Wind*
1072 *and the Sun* in Spanish and English, and the final set of talkers included one male and one female
1073 from each of the following four groups: monolingual English talkers, L1 English talkers who learned
1074 Spanish late and had completed college and graduate coursework in Spanish, L1 Spanish talkers from
1075 Latin America who learned English late and had moved to the U.S. to attend college, and early
1076 Spanish-English bilinguals (including the stimuli talker). The recordings from these eight talkers
1077 were divided into seven phrases, yielding 56 sound files of the talkers' English and 56 sound files of
1078 their Spanish.
1079
1080 The raters included ten monolingual English listeners and 10 L1 Spanish listeners from Latin
1081 America who learned English after age 14. None participated in the main study. Raters heard
1082 productions in their native language and decided how native- or foreign-sounding each production
1083 was by using the mouse to click on a horizontal line. The line appeared on the screen after the audio
1084 presentation of each sentence and represented a continuum between "Perfectly native sounding"
1085 (labeled as such at the left extreme) and "Very foreign sounding" (so labeled at the right extreme).
1086 The Spanish translations "Suena totalmente nativo" and "No suena nada nativo" were used in the
1087 Spanish version with the native Spanish listeners and the talkers' Spanish productions. The
1088 accentedness rating was recorded as the x-intercept of the mouse at the click. The 56 sentences were
1089 randomized for each listener.
1090
1091 Accentedness ratings were converted to *z*-scores to account for listeners using the continua
1092 differently, and the *z*-transformed accentedness ratings for English and Spanish productions were
1093 submitted to separate mixed-effects linear regression models using the *lme4* (v1.1-7) and *lmerTest*
1094 (v2.0-20) packages in R (RStudio 0.99.489). Listener was included as a random intercept, and testing

1095 talker as a fixed effect significantly improved the fit of a model with the random intercept alone, for
1096 both the English model (χ^2=1317.3, df=7, p<0.001) and the Spanish model (χ^2=948.25, df=7,
1097 p<0.001). See Table 11 for the model summaries. The stimuli talker (early bilingual male) was
1098 designated as the referent class for the talker variable. The intercept for the stimuli talker was
1099 significantly less than zero (p<0.001) in both the English and Spanish models and was thus
1100 significantly closer to the "Perfectly native sounding" extreme than to the center for both languages.
1101 The stimuli talker's English was not rated as significantly different from the monolingual English
1102 male (p=0.29) or the L1 English male (p=0.12), and he was rated as significantly more native
1103 sounding than all other talkers (at least p<0.01) except the monolingual English female (p<0.05).[8]
1104 The stimuli talker's Spanish was also rated as significantly more native sounding than all the other
1105 talkers (p<0.001), except for the L1 Spanish male and female, with whom there was no significant
1106 difference in rating (for L1 Spanish male, p=0.80; for L1 Spanish female, p=0.29).
1107
1108 Tables and Figures
1109
1110 Table 1: Nonce words with language-specific phonemes /θ,ɹ,r/.
1111

English phoneme /θ/	English phoneme /ɹ/	Spanish phoneme /r/
/tʃiθə/	/tʃaɹə/	/tʃira/
/fiθə/	/fiɹə/	/fara/
/hiθə/	/hiɹə/	/fira/
/maθə/	/maɹə/	/mara/
/saθə/	/ɹatʃə/	/mira/[9]
/siθə/	/ɹitʃə/	/ratʃa/
/θitʃə/	/ɹimə/	/ritʃa/
/θisə/	/siɹə/	/sira/

1112
1113 Table 2: Nonce words with language-specific phonetic variants of /l,u/.
1114

/l/		/u/	
English	Spanish	English	Spanish
[tʃałsə]	[tʃaltʃa]	[tʃʉtʃə]	[tʃuma]
[fałmə]	[filfa]	[fʉtʃə]	[fufa]
[hiłfə]	[lafa]	[fʉfə]]	[fusa]
[łitʃə]	[litʃa]	[fʉsə]	[mufa]
[łifə]	[lifa]	[hʉtʃə]	[muma]
[małfə]	[malfa]	[hʉsə]	[sutʃa]
[sałfə]	[silma]	[mʉmə]	[hutʃa]
[siłtʃə]	[halfa]	[sʉfə]	[husa]

1115

---

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

[9] Note that the Spanish nonce-word /mira/, which would be written *mirra*, is distinct from the real Spanish word *mira* /miɾa/ 'look,' which is produced with the tap /ɾ/. Such minimal pairs contrasting /r/ and /ɾ/ exist elsewhere in Spanish; consider *carro* /karo/ 'car' vs. *caro* /kaɾo/ 'expensive' and *perro* /pero/ 'dog' vs. *pero* /peɾo/ 'but.'

1116 Table 3. Demographic information and language background of participants.

1117

	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

1118

1119 Table 4. Acoustic properties of segments.

1120

1121 **(A)** Non-target vowels

1122

	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)

1123

1124 **(B)** Target segments

1125

	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)		

1126

1127 Table 5. Mean accuracy of each listener group for each stimulus type. Standard deviations are
1128 presented in parentheses.

1129

		Monolinguals	Early Bilinguals	Late Bilinguals
English Cues	Phonemic	72.7 (44.5)	78.8 (40.9)	86.1 (34.6)
	Phonetic	70.5 (45.6)	73.2 (44.3)	76.0 (42.7)
Spanish Cues	Phonemic	95.6 (20.5)	96.9 (17.3)	97.6 (15.4)
	Phonetic	91.0 (28.7)	90.4 (29.5)	90.6 (29.1)

1130

1131 Table 6. Summary of mixed effects logistic regression model fitting accuracy results.

1132

Predictor	*Mean*	*Posterior SD*	*95% CI*	*p value*
Intercept (Monolingual, English phonemes)	1.391	0.299	(0.763, 1.983)	<0.0001
Phonetic cues	-0.200	0.356	(-0.916, 0.489)	NS
Early Bilinguals	0.299	0.273	(-0.244, 0.818)	NS
Late Bilinguals	1.014	0.236	(0.546, 1.468)	<0.0001

Spanish cues	2.242	0.459	(1.402, 3.190)	<0.0001
Phonetic * Early	-0.247	0.097	(-0.442, -0.059)	NS
Phonetic * Late	-0.722	0.097	(-0.911, -0.533)	NS
Phonetic * Spanish	-0.562	0.559	(-1.629, 0.521)	<0.0001
Early * Spanish	-0.223	0.230	(-0.659, 0.233)	<0.0001
Late * Spanish	-0.686	0.231	(-1.135, -0.226)	<0.0001
Phonetic * Early * Spanish	-0.165	0.256	(-0.674, 0.322)	<0.0001
Phonetic * Late * Spanish	0.126	0.253	(-0.365, 0.619)	<0.0001

Random effects	*Variance*
Listener	0.892
Stimulus Word	0.970

1133
1134 Table 7. Mean RT (in milliseconds) for correct trials for each listener group and stimulus type.
1135

		Monolinguals	Early Bilinguals	Late Bilinguals
English Cues	Phonemic	542.0 (594.1)	629.8 (727.8)	662.7 (640.7)
	Phonetic	592.3 (742.9)	715.5 (833.4)	770.8 (791.5)
Spanish Cues	Phonemic	538.0 (591.8)	530.4 (545.1)	639.8 (675.8)
	Phonetic	595.4 (636.6)	641.2 (711.6)	777.1 (792.7)

1136
1137 Table 8. Summary of mixed effects linear regression model fitting log-transformed RT results.
1138

Predictor	*Mean*	*Posterior SD*	*95% CI*	*p value*
Intercept (Monolingual, English phonemes)	6.191	0.074	(6.046, 6.333)	<0.0001
Phonetic cues	0.040	0.059	(-0.074, 0.158)	NS
Early Bilinguals	-0.011	0.114	(-0.236, 0.216)	NS
Late Bilinguals	0.359	0.107	(0.146, 0.557)	<0.0001
Spanish cues	0.026	0.102	(-0.179, 0.226)	NS
Correct response	-0.178	0.025	(-0.224, -0.128)	<0.01
Phonetic * Early	-0.055	0.053	(-0.155, -0.051)	NS
Phonetic * Late	-0.192	0.052	(-0.296, -0.090)	<0.01
Phonetic * Spanish	0.037	0.121	(-0.208, 0.273)	<0.10
Early * Spanish	0.155	0.152	(-0.145, 0.449)	<0.05
Late * Spanish	-0.194	0.153	(-0.492, 0.106)	<0.01
Phonetic * Correct	-0.009	0.033	(-0.073, 0.054)	NS
Early * Correct	0.062	0.045	(-0.025, 0.150)	<0.10
Late * Correct	-0.132	0.046	(-0.223, -0.041)	NS
Spanish * Correct	-0.066	0.078	(-0.221, 0.091)	<0.01
Phonetic * Early * Spanish	0.045	0.164	(-0.265, 0.371)	<0.001
Phonetic * Late * Spanish	0.389	0.166	(0.058, 0.706)	<0.0001
Phonetic * Early * Correct	0.116	0.061	(-0.003, 0.237)	NS
Phonetic * Late * Correct	0.259	0.059	(0.144, 0.378)	<0.05
Phonetic * Spanish * Correct	0.029	0.088	(-0.150, 0.200)	<0.05
Early * Spanish * Correct	-0.192	0.156	(-0.499, 0.115)	<0.05
Late * Spanish * Correct	0.219	0.157	(-0.093, 0.520)	NS

Phonetic * Early * Spanish * Correct	-0.059	0.170	(-0.397, 0.267)	NS
Phonetic * Late * Spanish * Correct	-0.389	0.172	(-0.717, -0.046)	<0.01

Random effects	*Variance*
Listener	0.366
Stimulus Word	0.151

1139

1140 Table 9. Summary of results from stimuli comparisons.

1141

1142 **(A)** Cross-language comparisons

1143

	Accuracy	Reaction Times
Monolinguals	Spanish phonemic > English phonemic	Spanish phonemic = English phonemic
Early Bilinguals	Spanish phonetic > English phonetic	Spanish phonetic = English phonetic
Late Bilinguals		

1144

1145 **(B)** Cross-class comparisons

1146

	Accuracy	Reaction Times
Monolinguals	Spanish phonemic > Spanish phonetic English phonemic = English phonetic	Spanish phonemic = Spanish phonetic English phonemic = English phonetic
Early Bilinguals	Spanish phonemic > Spanish phonetic English phonemic » English phonetic	Spanish phonemic « Spanish phonetic English phonemic = English phonetic
Late Bilinguals	Spanish phonemic > Spanish phonetic English phonemic > English phonetic	Spanish phoneme < Spanish phonetic English phonemic = English phonetic

1147

1148 Table 10. Summary of results from listener group comparisons.

1149

	Accuracy	Reaction Times
Spanish phonemes	Monolinguals = Early = Late	
English phonemes	Monolinguals = Early < Late	Monolinguals = Early < Late
Spanish phonetic	Monolinguals = Early = Late	
English phonetic		Monolinguals « Early < Late

1150

1151 Table 11. Model summaries for mixed-effects linear regression models predicting accentedness
1152 ratings.

1153

1154 **(A)** English productions

1155

Predictor	*Estimate*	*Standard Error*	*t value*	*p value*
Intercept (Stimuli talker)	-0.632	0.037	-17.186	<0.001
Monolingual male	0.0554	0.052	1.065	0.29
Monolingual female	-0.131	0.052	-2.516	<0.05
L1 English male	-0.082	0.052	-1.575	0.12
L1 English female	0.163	0.052	3.127	<0.01
Early bilingual female	0.613	0.052	11.785	<0.001

| L1 Spanish male | 2.123 | 0.052 | 40.797 | <0.001 |
| L1 Spanish female | 2.318 | 0.052 | 44.537 | <0.01 |

Random effects	*Variance*
Listener	<0.001
Residual	0.095

1156

1157 **(B)** Spanish productions

1158

Predictor	*Estimate*	*Standard Error*	*t value*	*p value*
Intercept (Stimuli talker)	-0.873	0.051	-17.062	<0.001
Monolingual male	2.272	0.072	31.404	<0.001
Monolingual female	2.241	0.072	30.970	<0.001
L1 English male	1.292	0.072	17.861	<0.001
L1 English female	0.661	0.072	9.144	<0.001
Early bilingual female	0.458	0.072	6.323	<0.001
L1 Spanish male	-0.018	0.072	-0.255	0.80
L1 Spanish female	0.077	0.072	1.070	0.29

Random effects	*Variance*
Listener	<0.001
Residual	0.183

1159

1160 Figure 1. Predicted log odds of accuracy for phonemic and phonetic cues.

1161

1162 Figure 2. Model log reaction time for phonemic and phonetic cues in accurate trials.