

Language and Speech

<http://las.sagepub.com>

L1-Spanish Speakers' Acquisition of the English /i/ / Contrast II: Perception of Vowel Inherent Spectral Change¹

Geoffrey Stewart Morrison
Language and Speech 2009; 52; 437
DOI: 10.1177/0023830909336583

The online version of this article can be found at:
<http://las.sagepub.com/cgi/content/abstract/52/4/437>

Published by:



<http://www.sagepublications.com>

Additional services and information for *Language and Speech* can be found at:

Email Alerts: <http://las.sagepub.com/cgi/alerts>

Subscriptions: <http://las.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.co.uk/journalsPermissions.nav>

Citations <http://las.sagepub.com/cgi/content/refs/52/4/437>

L1-Spanish Speakers' Acquisition of the English /i/–/ɪ/ Contrast II: Perception of Vowel Inherent Spectral Change¹

Geoffrey Stewart Morrison

Australian National University²

Key words

English

L2 vowel perception

Spanish

Abstract

L1-Spanish learners of English have been reported to distinguish English /i/ and /ɪ/ on the basis of duration cues, whereas L1-English listeners primarily use spectral cues. Morrison (2008a) hypothesized that duration-based perception is a secondary developmental stage that emerges from an initial stage of multidimensional-category-goodness assimilation of tokens of English /i/ and /ɪ/ to Spanish /i/, with English vowel tokens perceived to be good examples of Spanish /i/ labeled as English /i/ and poor examples labeled as English /ɪ/.

- 1 The present article is a follow-up to “L1-Spanish speakers’ acquisition of the English /i/–/ɪ/ contrast: Duration-based perception is not the initial developmental stage” (Morrison, 2008a).
- 2 The bulk of the research and writing of this article took place while the author was affiliated with the Department of Linguistics, University of Alberta, and the Department of Cognitive & Neural Systems, Boston University.

Acknowledgments: This article is based in part on portions of the author’s doctoral dissertation (Morrison, 2006b) “L1 & L2 production and perception of English and Spanish vowels: A statistical modelling approach,” Department of Linguistics, University of Alberta, 2006, which was supported by a Social Sciences and Humanities Research Council of Canada (SSHRC) Doctoral Fellowship. Additional data collection and writing took place while the author was an SSHRC Postdoctoral Fellow affiliated with the Department of Cognitive & Neural Systems, Boston University. Portions of this article were presented as “Multidimensional category-goodness-difference assimilation of Canadian English /i/ and /ɪ/ to Spanish /i/” at the Fourth Joint Meeting of the Acoustical Society of America and the Acoustical Society of Japan, Honolulu, HI, November 2006. Thanks to everyone who assisted with data collection, to my PhD supervisor Terry Nearey for advice throughout the project, and to Paola Escudero, Susan Guion, Anders Højen, Erin Ingvalson, and Jim Scobbie for comments on earlier versions of this article.

Address for correspondence. School of Language Studies, Building 110, Australian National University, Canberra, ACT 0200, Australia; <geoff.morrison@anu.edu.au>

Language and Speech

© The Author, 2009. Reprints and permissions: www.sagepub.co.uk/journalsPermissions.nav

Language and Speech 0023-8309; Vol 52(4): 437–462; 336583; DOI:10.1177/0023830909336583
<http://las.sagepub.com>

The present article tests this hypothesis using a synthetic vowel continuum that includes vowel inherent spectral change. Perception results are reported for monolingual-Western-Canadian-English listeners, monolingual-Mexican-Spanish listeners, and L1-Spanish L2-English listeners. Given that Spanish /i/ is generally assumed to be a monophthong, stimuli with English-/ɪ/-like formant movement (converging F1 and F2) that are in the region of Spanish /i/ are predicted to be poorer matches for Spanish /i/ than otherwise similar stimuli without formant movement. The multidimensional-category-goodness-difference-assimilation hypothesis therefore predicts that stimuli with converging formants are more likely to be labeled as English /i/ by L1-Spanish L2-English listeners. Results are consistent with the hypothesis.

1 introduction

The present article investigates the perception of synthetic vowel stimuli by monolingual-Western-Canadian-English listeners, monolingual-Mexican-Spanish listeners, and first-language-Spanish second-language-English listeners (L1-Spanish L2-English listeners). The synthetic stimuli, presented in the context /b_pɑ/, cover an acoustic space that allows for the perception of English /i/, /ɪ/, /e/, /ɛ/, and Spanish /i/, /e/, and /ei/. In addition to variation in initial formant values and duration, the synthetic stimuli vary in final formant values. Therefore, whereas previous studies have examined L1-Spanish L2-English listeners' perception of synthetic vowels with static formants, the present study examines the effects of vowel inherent spectral change on L1-Spanish listeners' learning of the English /i/-/ɪ/ contrast. The results are consistent with the proposal in Morrison (2008a) that L1-Spanish learners initially perceive the English /i/-/ɪ/ contrast via a category-goodness-difference assimilation to Spanish /i/, and that their oft reported perceptual reliance on vowel duration reflects a later developmental stage.

L1-Spanish L2-English listeners often misidentify L1-English speakers' productions of /i/ as /ɪ/ and vice versa (Møller Glasbrenner, 2005). In some studies, using English dialects from the United States and England, L1-Spanish listeners have been found to assimilate most tokens of both the English /i/ category and the English /ɪ/ category to a single Spanish vowel category, Spanish /i/ (Álvarez González, 1980, ch. 5; Escudero, 2005, §1.2.2; Flege, 1991). In studies of English dialects from North America and England, L1-English listeners have been found to distinguish /i/ and /ɪ/ primarily on the basis of spectral cues, whereas L1-Spanish L2-English listeners show a tendency to primarily use duration cues (Escudero & Boersma, 2004; Flege, Bohn, & Jang, 1997; Morrison, 2002). Since Spanish does not have a phonemic vowel duration contrast, L1-Spanish L2-English listeners' use of duration is not attributable to L1-transfer,³ and several alternative explanations have been proposed.

One proposal to account for L1-Spanish L2-English listeners' preference for duration cues is that L2-English learners use duration because they are taught that /i/ is "long" and /ɪ/ is "short" (Flege et al., 1997; Morrison, 2008a; Wang & Munro, 1999). However, even if it is the case that L2-English learners' attention is explicitly directed towards the duration contrast, an explanation is still needed for why they do

3 Although Kondaurova and Francis (2008) hypothesized that long and short English vowels could undergo two-category assimilation to stressed and non-stressed Spanish vowels.

not immediately notice that there is also a spectral difference between the two English vowels. In the Desensitization Hypothesis, Bohn (1995) proposed that, because Spanish does not expose L1-Spanish listeners to phoneme-distinguishing spectral differences in the low-F1–high-F2 part of the vowel space, they are “linguistically desensitized” to the spectral differences that distinguish English /i/ and /ɪ/. Escudero and Boersma (2004) proposed that existing Spanish vowel categories impede L1-Spanish listeners’ ability to form L2 categories on the basis of the statistical distribution of spectral properties in L1-English speakers’ /i/ and /ɪ/ productions; however, because L1-Spanish listeners have no duration-based categories, there is no impediment to them learning L2 categories on the basis of the statistical distribution of duration in L1-English speakers’ /i/ and /ɪ/ productions. The explanations proposed by Bohn (1995) and Escudero and Boersma (2004) are both predicated on the assumption that duration cues are the first means by which L1-Spanish learners of English perceptually distinguish English /i/ and /ɪ/. It is then logical to propose a mechanism that selectively impairs the L2-learners’ ability to make use of spectral cues, or to propose a mechanism that selectively enhances their ability to use duration cues.⁴

Morrison (2008a) presented evidence that suggests that duration-based perception is not the first means by which L1-Spanish learners of English distinguish English /i/ and /ɪ/ (at least for Canadian English). In the results of a perception experiment, he found a group of listeners (5 out of 20 participants) who identified longer vowels with higher F1 and lower F2 as English /i/, and shorter vowels with lower F1 and higher F2 as English /ɪ/. Their use of duration cues was positively correlated with duration properties in L1-English speakers’ /i/ and /ɪ/ productions, but their use of spectral cues was negatively correlated. This perception pattern could not have emerged from an earlier duration-based perception pattern: All else being equal, longer vowels are more likely to be tokens of English /i/ and shorter vowels are more likely to be tokens of English /ɪ/. Low-F1–high-F2 is therefore positively correlated with long duration, and high-F1–low-F2 is positively correlated with shorter duration. If listeners begin by labeling long vowels as English /i/ and short vowels as English /ɪ/, exposure to L1-English speakers’ productions will present them with evidence that English /i/ also has low F1 and high F2, and English /ɪ/ also has high F1 and low F2. L1-Spanish L2-English listeners may therefore be able to use duration as a bootstrap for developing spectrally-based perception. Note, however, that this would result in both spectral and duration cues being used in a manner that is positively correlated with L1-English speakers’ productions. Positively correlated use of duration cues and negatively correlated use of spectral cues would not emerge.

The reversed-spectral pattern could not be explained via a labeling reversal, since such a reversal would result in both spectral and duration cues being negatively

4 In a study on Catalan listeners’ perception of English vowels (including /i/ and /ɪ/), Cebrian (2006) suggests a different explanation: If listeners hear an L2 contrast that is a combination of a feature that is used in their L1, for example, a spectral difference, and a feature that is not used in their L1, for example, a duration difference, then if they detect a difference they attribute it to the novel feature, for example, duration. Presumably the argument is that its novelty makes the duration difference more conspicuous.

correlated with L1-English speakers' productions. An explanation based on orthography, that the response keywords "beat" and "bead" were mistakenly interpreted as representing long /e/ vowels, was also discounted in Morrison (2008a) because the listeners made use of a "bed" response for long vowels with high F1 and low F2 (response options available in Morrison, 2008a, were: "beat," "bead," "bit," "bid," "bet," "bed").

The explanation proposed in Morrison (2008a) for the positively correlated use of duration and negatively correlated use of spectral properties, was that the listeners were performing a multidimensional-category-goodness-difference assimilation to Spanish /i/ (terminology from Best, 1995). Stimuli that were perceived as good matches for Spanish /i/ (those with shorter duration, lower F1 and higher F2) were labeled as English /ɪ/, and stimuli that were assimilated as poorer matches for Spanish /i/ (those with either longer duration, or higher F1 and lower F2, or both) were labeled as English /i/. Two hypotheses were suggested for why L1-Spanish listeners might choose this direction of labeling rather than the opposite. One hypothesis was that, in a multidimensional spectral and duration space, and considering only tokens of English vowels that are assimilated to Spanish /i/, tokens of English /i/ are on average further from the prototypical values for Spanish /i/ than are tokens of English /ɪ/. The other hypothesis was based on orthography: In English orthography 'i' frequently represents English /i/ but seldom represents English /ɪ/, and in Spanish orthography 'i' represents Spanish /i/. Bilingually literate L1-Spanish L2-English listeners may therefore make a link between English and Spanish words spelt with 'i', and associate English /ɪ/ with Spanish /i/.⁵

Morrison (2008a) argued that the positive-duration-negative-spectral pattern must logically be the first means by which L1-Spanish learners of English perceptually distinguish English /i/ and /ɪ/. He inserted an additional stage (Stage ½) into the developmental stages that had initially been hypothesized by Escudero (2000):⁶ Stage 0, L1-Spanish listeners do not distinguish English /i/ and /ɪ/. Stage ½, the two

5 It may even be that the orthographic effect is sufficiently strong that literate L2 listeners shift from perceiving most tokens of English /i/ as (possibly) poor examples of L1 /e/ to perceiving them as examples of L1 /i/. This appears to be the case for one L1-Spanish L2-English participant in a longitudinal study reported in Morrison (2006b). It may also account for the finding by Frieda and Nozawa (2007) that tokens of English /i/ and /ɪ/ were primarily two-category assimilated to Japanese /i:/ and /e:/ respectively by inexperienced L1-Japanese learners of English but were primarily single-category assimilated to Japanese /i:/ by experienced L1-Japanese learners of English. Arguably the experienced learners were more influenced by orthography and the inexperienced learners were more influenced by the acoustic similarity of English /i/ and Japanese /e:/ (context may account for the preponderance of long vowel responses). The orthographic effect may also account for why L1-Korean L2-English adults in Baker and Trofimovich (2005) merged both English /i/ and English /ɪ/ with Korean /i/, whereas L1-Korean L2-English children merged English /i/ with Korean /i/ but merged English /ɪ/ with Korean /e/. Arguably the adults were more influenced by orthography and the children were more influenced by the acoustic similarity of English /ɪ/ and Korean /e/.

6 For the purposes of qualitative description, it is convenient to talk in terms of stages; however, the use of the term "stage" is not meant to imply a discrete step. Development may be envisioned as gradual movement along a path.

English vowels are distinguished via a multidimensional-category-goodness-difference assimilation to Spanish /i/, with more-Spanish-/i/-like English vowels labeled as English /i/ and less-Spanish-/i/-like English vowels labeled as English /i/. Since perception is positively correlated with duration cues in L1-English speakers' productions, and spectral cues are negatively correlated, exposure to English provides positive feedback on the use of duration cues and negative feedback on the use of spectral cues. This leads to Stage 1, duration-based perception with long vowels labeled as English /i/ and short vowels labeled as English /i/. Since in L1-English speakers' /i/ and /i/ productions, spectral properties are partially correlated with duration properties, additional exposure to English may allow L1-Spanish listeners to use duration cues to bootstrap positively correlated use of spectral cues. This potentially leads to Stages 2 and 3, partially- and fully-L1-English-like perception of English /i/ and /i/. This will be referred to as the hypothesized *indirect developmental path*. Note that duration-based perception is hypothesized to emerge from an earlier stage in which L1-Spanish listeners attend to duration and spectral properties. A mechanism that impedes the use of spectral cues or enhances the use of duration cues is not part of this proposal. Also, duration-based perception and the indirect developmental path only emerge if good matches for Spanish /i/ are labeled as English /i/ and poor matches are labeled as English /i/. If the opposite direction of labeling were selected at Stage ½, spectral cues would be positively correlated and duration cues negatively correlated with L1-English speakers' /i/ and /i/ productions, and since spectral cues are primary for L1-English listeners, this would immediately result in relatively L1-English-like perception. Exposure to English could reinforce these listeners' use of spectral cues, but provide negative feedback on their use of duration cues and eventually lead them to reverse their use of the latter. This scenario represents a hypothesized *direct developmental path*.

In Escudero and Boersma (2004), Flege et al. (1997), and Morrison (2008a) the stimuli contained synthetic vowels in which the formant values were fixed throughout the duration of the vowel. However, vowel inherent spectral change (VISC) has been found to be an important factor in L1-English listeners' perception of English /i/ in several North American dialects (Andruski & Nearey, 1992; Assmann & Katz, 2005; Assmann, Nearey, & Hogan, 1982; Hillenbrand, Clark, & Nearey, 2001). Western-Canadian-English /i/ is produced with first and second formants (F1 and F2) that converge over the timecourse of the vowel, F1 rises and F2 falls, whereas Western-Canadian-English /i/ has negligible formant movement (Andruski & Nearey, 1992; Morrison, 2006b; Nearey & Assmann, 1986).⁷ Since Spanish /i/ is generally assumed to be a monophthong, Morrison (2008a) predicted that a stimulus with converging VISC would be a poorer match for Spanish /i/ than a stimulus with flat formants. L1-Spanish listeners at Stage ½ of the indirect developmental path were therefore predicted to be more likely to label converging-VISC stimuli as English /i/, even though the VISC pattern is appropriate for English /i/. The present study tests this hypothesis, and explores the influence of VISC on L1-Spanish L2-English listeners' vowel perception.

7 These patterns of formant movement were observed in the following contexts: isolated vowel, /b_b/, /p_p/, /b_pə/, and /b_p/. Pending additional data, these findings are assumed to generalize to other contexts.

Some tokens of English /i/ are assimilated to Spanish /e/ (Álvarez González, 1980, ch. 5; Escudero, 2005, §1.2.2; Flege, 1991) and confused with English /ɛ/ (Møller Glasbrenner, 2005).⁸ Tokens of English /e/ are also expected to be assimilated to Spanish /e/ (Højen & Flege, 2006), or to Spanish /ei/. The present study therefore explores a synthetic vowel space that allows for the perception of English /i/, /ɪ/, /e/, /ɛ/, and Spanish /i/, /e/, and /ei/. The same set of synthetic vowels is presented to L1-Spanish and L1-English listeners. The vowels are presented in the context /b_pa/, which results in possible but non-existent words in both Spanish and Canadian English. Data reported in the present article are for monolingual-Mexican-Spanish listeners' Spanish-vowel-category responses (/i/, /e/, /ei/), and for monolingual-Western-Canadian-English and L1-Spanish L2-English listeners' English-vowel-category responses (/i/, /ɪ/, /e/, /ɛ/).

2 Method

2.1

Participants

The present article reports on the perception of monolingual-English listeners, monolingual-Spanish listeners, and L1-Spanish L2-English listeners. Because English vowels can have substantial variation across dialects, the English dialect was carefully controlled. The monolingual-English participants were all speakers of Western-Canadian English, and the L1-Spanish L2-English participants were all living in Western Canada. Compared to English vowels, Spanish vowels appear to have little variation across most dialects of Spanish (see Morrison & Escudero, 2007, for an acoustical-statistical comparison of Lima and Madrid dialects). A common procedure is therefore to recruit a large number of L1-Spanish L2-English volunteers without controlling for dialect (Escudero & Boersma, 2004; Flege et al., 1997; Morrison, 2008a). In the present study approximately half the bilingual speakers were from Mexico, and baseline data were therefore also collected from monolingual-Mexican-Spanish listeners.

Nineteen monolingual-English speakers (8 male and 11 female) were recruited in Edmonton, Alberta, Canada (one was from Saskatchewan and all the others from Alberta). None reported an ability to speak any language other than English. They ranged in age from 18 to 54 with a median of 20.

Twenty monolingual-Mexican-Spanish speakers (10 male and 10 female) were recruited in Mexico City, Federal District, Mexico. They were all speakers of Mexico-City Spanish. Thirteen reported a limited ability to speak English or French, but reported being unable to participate in a conversation in these languages. They ranged in age from 18 to 31 with a median of 22.

Forty L1-Spanish L2-English participants (19 male and 21 female) were recruited in Edmonton. Twenty-one were from Mexico, four from Colombia, four from Peru,

8 The degree to which tokens of English /i/ are assimilated to Spanish /e/ as opposed to /ɪ/ is dependent on the English dialect: For example, compared to other English dialects, Scottish English has a relatively large spectral difference between /i/ and /ɪ/, and L1-Spanish listeners tend to assimilate tokens of Scottish-English /i/ and /ɪ/ to Spanish /i/ and /e/ respectively (Escudero, 2005, §1.2.2).

two from Argentina, two from Spain, two from Venezuela, and one each from Bolivia, Chile, El Salvador, Panama, and Puerto Rico. Thirteen reported a limited ability to speak one or more of French, Portuguese, Japanese, German, Italian, and Hebrew, but did not report being able to conduct a conversation in any of these languages (potential participants were excluded from the study if they had a higher degree of proficiency in any language other than Spanish or English). They ranged in age from 18 to 50 with a median of 26. Some had started learning English in a classroom as early as 4 years of age, but the median age at which they commenced learning English was 12, and none had had extensive exposure to English prior to their teenage years. Participants had spent varying amounts of time in Canada ranging from 2 weeks to 24 years (10 participants had also spent some time in the United States, and 2 of these had also spent some time in the United Kingdom, but their most recent exposure had been to Western-Canadian English).

Potential participants completed a brief language background questionnaire. One section of the questionnaire asked potential participants to list the languages they spoke and indicate their proficiency in those languages, the choices being: *a-little*, *some*, *well*, and *near-native* (or in Spanish *un-poco*, *algo*, *bien*, *casi-nativo*). Potential participants who responded well or near-native for any languages other than English or Spanish were not included in the study. Potential participants who had been exposed to other languages in early childhood were also excluded, as were potential participants who reported hearing or speech impediments. All the English listeners would have studied French at school, and all the Spanish listeners would have studied English. None of the monolingual-English participants indicated that they could speak any language other than English even at the *a-little* level. Some of the monolingual-Spanish participants indicated that they could speak *a little* English. In such cases, the author attempted to converse with the participant in English to ascertain whether they could conduct a conversation in English (potential bilingual participants who reported *a-little* or *some* ability to speak French or Japanese were also tested in this way). It was not practical to run any formal tests to confirm that participants were functionally monolingual; however, the author was satisfied that all the monolingual participants were indeed functionally monolingual and that all the L1-Spanish L2-English participants were no more than functionally bilingual (not trilingual or quadrilingual, etc.).

2.2

Stimuli

A version of the Klatt synthesizer (Klatt & Klatt, 1990) was used to create synthetic /bVp/ stimuli.⁹ A natural production of Spanish /pa/ was appended,¹⁰ and the results were appended to the natural Spanish and English carrier sentences “La próxima

9 One of the reviewers pointed out that the degree of prevoicing on the initial synthetic /b/ should have been varied to help cue the language spoken as English or Spanish. The initial synthetic /b/ was produced with a 4 ms lead VOT, and it is possible that some of the listeners, especially L1-Spanish listeners perceived this as a /p/.

10 The natural /pa/ was extracted from the carrier sentence originally recorded as “La próxima palabra es bipa.”

palabra es” and “The next word is” (both sentences have the same meaning). In pilot tests the unstressed utterance final Spanish /a/ was acceptable to L1-English listeners, that is, it was not perceived as non-English-like. In English-listening mode, the author would transcribe the sound as English schwa; its mean F1, F2, and F3 values were 696, 1357, and 2376 Hz. The natural portions of the stimuli were produced by a male bilingual speaker (the author), and care was taken to adjust synthesizer-parameter setting so as to produce synthetic speech which (in the opinion of the author) was a good match for the voice quality of the Spanish natural speech. The speaker's Spanish productions had a greater spectral tilt than his English productions, and the spectral tilt of the English carrier sentence was therefore increased so as to match the voice quality of the Spanish-based synthetic speech.

A large stimulus space (1464 stimuli) was initially constructed, and pilot studies were conducted in order to find a smaller set of stimuli that included stimuli that were acceptable as Spanish /i/, /e/, and /ei/ to L1-Spanish listeners, and stimuli that were acceptable as English /i/, /ɪ/, /e/, and /ɛ/ to L1-English listeners. Figure 1 provides plots of the smaller stimulus set. The supplementary material available on the publisher's website (<http://las.sagepub.com/cgi/content/full/52/4/437/DC1>) includes a version of Figure 1 in which clicking on a dot causes the related stimulus sound to be played. The 90 stimuli selected had 10 sets of initial formant values along a diagonal in the F1–F2 vowel space ranging from [F1, F2] of [283 Hz, 2090 Hz] to [580 Hz, 1730 Hz], in equal steps of [+33 Hz, –40 Hz]. At each start-point, stimuli were synthesized with three levels of VISC: F1 and F2 either diverged, did not change (were flat), or converged over the time-course of the vowel. Formant movements [$\Delta F1$, $\Delta F2$] from the beginning to the end of the vowel were [–99 Hz, +120 Hz], [0 Hz, 0 Hz], and [+99 Hz, –120 Hz] (minus three, zero, and plus three steps along the F1–F2 diagonal). Following Andruski and Nearey (1992), the formant trajectories described straight lines in a log-hertz F1–F2 space. Following Nearey (1989), third-formant (F3) values were set using a formula based on a linear regression of F1 and F2 values onto F3 values from the model speaker's vowel productions. Equation (1) provides the formula with F1, F2, and F3 values given in hertz.

$$F3 = 4235 - 2.427 \times F1 - 0.272 \times F2 \quad (1)$$

Each of the 30 initial- and final-target combinations was synthesized at three durations: 55, 70, and 85 ms (for each set of stimuli with the same initial- and final-target combinations, the shorter stimuli had steeper slopes than the longer stimuli). The synthetic stimuli also included a bilabial burst, bilabial onset and offset formant transitions, and a 90 ms long /p/ closure. The consonant transitions added an additional 25 ms to each vowel, resulting in total vowel durations of 80, 95, and 110 ms.

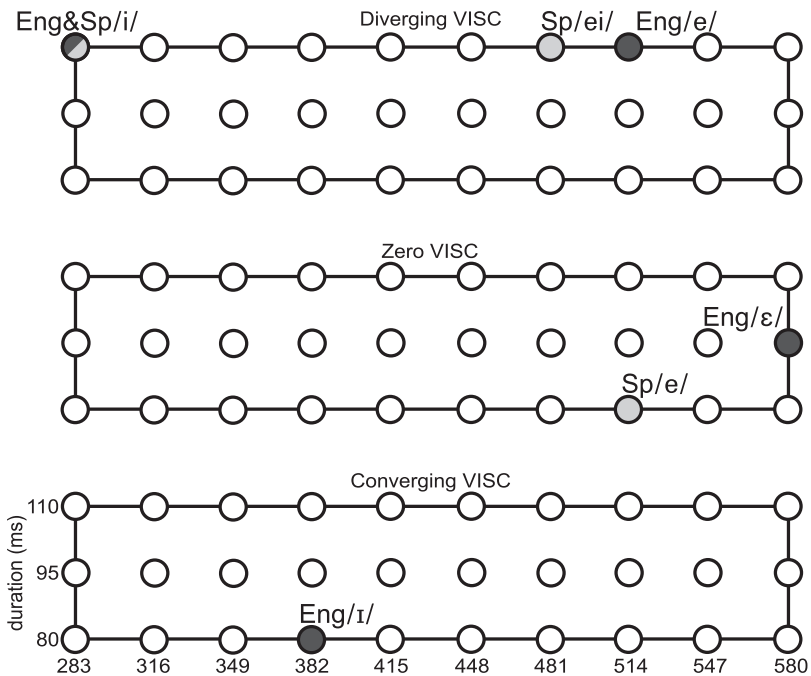
2.3

Procedures

Participants were tested one at a time using custom-written software. Monolingual-Mexican-Spanish participants were tested in the Phonic Studies Laboratory at El Colegio de México in Mexico City, and the remaining participants in the Centre for Comparative Psycholinguistics at the University of Alberta. In Canada testing took place in an IAC sound booth using a Roland ED UA-30 USB Audio Interface and

Figure 1

Properties of the synthetic stimuli. Labels are given for the duration of the vowels including consonant transitions. Labels are given for initial F1 values; F1 and F2 covaried, and the corresponding initial F2 values were: 2090, 2050, 2010, 1970, 1930, 1890, 1850, 1810, 1790, 1750, 1730 Hz. The top panel represents stimuli with diverging VISC, the final F1 value was 99 Hz less than the initial F1 value and the final F2 value was 120 Hz more than the initial F2 value. The middle panel represents stimuli with zero VISC, the formant values did not change over the timecourse of the vowel. The bottom panel represents stimuli with converging VISC, the final F1 value was 99 Hz more than the initial F1 value and the final F2 value was 120 Hz less than the initial F2 value. Filled circles represent the stimuli that were selected in pilot studies as the best examples of the four English and three Spanish vowels. An electronic version of this figure is provided in the supplementary materials available on the publisher's website (<http://las.sagepub.com/cgi/content/full/52/4/437/DC1>), in the latter version when one clicks on a dot the associated stimulus is played



Sennheiser HMD 280 PRO headphones. In Mexico testing took place in the quietest room available using an Edirol UA-25 Audio Interface and AKG K701 headphones.

All communications between the researcher and the listeners, all the instructions, and the stimulus carrier sentences were in the language being tested. This was designed to prime the bilingual listeners to perceive the stimuli in the appropriate language (Bohn & Flege, 1993; Escudero & Boersma, 2002), and was a necessity for the monolingual listeners. Listeners heard a stimulus sentence, and responded by clicking on the response button that corresponded to their identification of the synthetic vowel. A new stimulus was presented 500 ms after a response was given. In the Spanish

experiment the response buttons were labeled *BIPA*, *BEIPA*, and *BEPA* representing /bipa/, /beipa/, and /bepa/ respectively, and in the English experiment the response buttons were labeled *BEEPA*, *BIPPA*, *BAYPA*, and *BEPPA* representing /bipə/, /bɪpə/, /bepə/, and /bɛpə/ respectively. The spelling-to-phoneme relationship is transparent in Spanish, but less clear in English. Prior to the English experiment, listeners were therefore trained on the English spelling-to-phoneme relationship. Listeners saw written sets of real words illustrating the four English vowel categories, and each set was followed by the corresponding response word. Listeners read the real and response words out loud, and the researcher monitored to ensure that they pronounced the same vowel sounds in the response words as in the real words. Any mismatches between the real and response words were corrected by asking the listeners to read the response word with the same vowel as in the appropriate set of real words. The researcher pointed at the written forms of the words but did not pronounce the words or model the vowels in isolation. Training was restricted to making sure that participants produced the same vowel sound in real and response words. L2-English participants were not corrected or given training on how the phonemes would be pronounced by L1-English speakers. The training continued until the researcher was confident that the listeners understood the spelling-to-phoneme relationships.¹¹ The written sets of real and response words were also visible to the listeners during the experiment.

All 90 stimuli were presented in random order in two blocks, and in each of four subsequent randomized blocks an adaptive procedure selected 45 stimuli for presentation. In each of the last four blocks, category boundaries were estimated on the basis of the responses given in the earlier blocks, and stimuli in the vicinity of the category boundaries had the highest probability of being selected for presentation in the new block. This resulted in a total of 360 trials per listener, with each stimulus identified a minimum of twice and a maximum of six times. The procedure is described in detail in Morrison (2006a).

-
- 11 L1-Spanish L2-English listeners had little difficulty learning the English spelling-to-phoneme relationship, and this was also true for L1-English L2-Spanish listeners (data from the latter group of participants were reported in Morrison, 2006b), but a number of monolingual-English participants had considerable difficulty. Some of the author's colleagues posited that this could be because the English listeners were likely to have been taught at school using a so-called "whole language" approach to reading and may never have received instruction in orthography to sound correspondences. Once the problem became apparent (after the first four monolingual-English listeners had been tested), prior to the perception experiment monolingual-English participants underwent a screening test in which they identified the four synthetic stimuli that had been identified in pilot tests as the best examples of the four English vowel categories. They received immediate feedback as to whether their response was correct, and if incorrect heard the stimulus again and saw a visual indication of the correct written response. Training continued until listeners scored 100% correct on at least four repetitions of the training set. Listeners who failed to meet this criterion were not included in the study. This screening test is assumed not to have affected the results of the monolingual-listeners' perception experiments: The stimuli in the screening test were those that were most prototypical of each vowel category, and thus were expected to be unambiguously perceived. The aim of the training was to ensure that the listeners could consistently associate the correct orthographic labels with these unambiguous stimuli.

3 Results and discussion

Perception results were analyzed using logistic regression. For an explanation of the type of logistic regression modeling applied here and a guide to the interpretation of the results, it is highly recommended that the reader refer to Morrison (2007) and Morrison (2008a).

The logistic regression models estimated a set of coefficient values associated with each response category. Each set of coefficients consisted of a bias coefficient, a duration-tuned coefficient, an initial-formant-value-tuned coefficient, and a pair of VISC-tuned coefficients. For example the English model included the following:

bias coefficients:	$\alpha/i/, \alpha/l/, \alpha/e/, \alpha/\epsilon/$
initial-formant-tuned coefficients:	$\beta/i/\text{initialF}, \beta/l/\text{initialF}, \beta/e/\text{initialF}, \beta/\epsilon/\text{initialF}$
duration-tuned coefficients:	$\beta/i/\text{dur}, \beta/l/\text{dur}, \beta/e/\text{dur}, \beta/\epsilon/\text{dur}$
diverging-VISC-tuned coefficients:	$\beta/i/\text{div}, \beta/l/\text{div}, \beta/e/\text{div}, \beta/\epsilon/\text{div}$
converging-VISC-tuned coefficients:	$\beta/i/\text{conv}, \beta/l/\text{conv}, \beta/e/\text{conv}, \beta/\epsilon/\text{conv}$

Initial formant values and duration values were entered as continuous variables in just-noticeable-difference (JND) units. The JND scale for initial formant values was one-dimensional (F1 and F2 were 100% correlated in the synthetic stimuli) with its origin corresponding to the stimuli with the lowest F1 and highest F2 [283 Hz, 2090 Hz]. The JND used was 0.3 Bark (Kewley-Port, 2001). The conversion from hertz to the JND-formant scale (F_{JND}) was performed using Equation (2) (which includes the hertz-to-bark formula from Traunmüller, 1990):

$$F_{\text{JND}} = \sqrt{(\text{Bark}(F1) - \text{Bark}(283))^2 + (\text{Bark}(F2) - \text{Bark}(2090))^2} / 0.3 \quad (2)$$

$$\text{Bark}(F) = (26.81F / (1960 + F)) - 0.53$$

The origin of the JND scale for duration corresponded to the stimuli with the shortest duration (80 ms), and the JND used was 5 ms on a base value of 90 ms (Noteboom & Doodeman, 1980, similar to the Weber fraction of 0.05 used by Smits, Sereno, & Jongman, 2006). The conversion from milliseconds to the JND-duration scale (dur_{JND}) was performed using Equation (3):

$$\text{dur}_{\text{JND}} = \log_{1+(5/90)}(\text{dur}/90) - \log_{1+(5/90)}(80/90) \quad (3)$$

Use of JND-scales allows initial-formant and duration results to be compared on an equal footing.¹²

12 Note, however, that even when two perceptual dimensions are both tested using JND scales, listeners may exploit one more than the other (Holt & Lotto, 2006).

VISC was entered as three discrete levels, resulting in two dummy-coding coefficients [β_{div} β_{conv}]: [0 0] = zero VISC, [0 1] = diverging VISC, [1 0] = converging VISC. This encodes the onset + offset (or the onset + direction) hypothesis for the perceptually relevant aspects of VISC (Gottfried, Miller, & Meyer, 1993; Morrison, 2006b, Appendix 1; Morrison & Nearey, 2007; Nearey & Assmann, 1986; Pols, 1977).

3.1

Monolingual-English and monolingual-Spanish perception

Figures 2 and 3 provide population-average territorial maps and probability-surface plots based on logistic regression models fitted to monolingual-Western-Canadian-English listeners' response data, and monolingual-Mexican-Spanish listeners' response data respectively. The population-average plots were created by fitting a logistic regression model to each individual listener's response data, then taking the mean of the logistic regression coefficient estimates across all listeners within each group.¹³ The models impose logistic functions on the data which allow smooth interpolation of response values over the stimulus space.

Territorial maps indicate which category is the model's predicted modal response in each part of the stimulus space (see Nearey, 1990, 1997). In probability-surface plots, each response category is represented by a curved surface and the height of a particular surface indicates the modeled probability of its response category being chosen in each part of the stimulus space (see Morrison, 2007, 2008a). Whereas territorial maps allow one to visualize the most probable (modal) response only, probability-surface plots allow one to visualize the first, second, and third most probable responses. Probability-surface plots also provide a visualization of the rate of change from one response category to another as stimulus values change. A territorial map is equivalent to a probability-surface plot viewed from directly above the stimulus space, and allows an easier assessment of the orientation of the boundary between categories.

Examination of Figure 2 indicates that, for monolingual-Western-Canadian-English listeners, English /e/ is the modal response in approximately half the diverging-VISC portion of the stimulus space, consistent with its traditional description as a (diverging) phonetic diphthong. Western-Canadian-English /ɪ/ and /e/ are produced with converging VISC (Andruski & Nearey, 1992; Nearey & Assmann, 1986), and consistent with this, English /ɪ/ and /e/ were the modal responses over most of the converging-VISC portion of the stimulus space. Western-Canadian-English /i/ is produced as a monophthong, and consistent with this, English /i/ was the modal response in the low-F1 part of the zero-VISC portion of the stimulus space. Some parts of the stimulus space, for example, low-F1 converging-VISC, do not correspond to the production values of any English vowel categories, but listeners extrapolated one of the neighboring categories to give responses in these areas.

13 Territorial maps and probability-surface plots for each individual listener, along with model-to-data goodness-of-fit measures (the deviance statistic, G^2 , and the percentage sum of absolute errors in proportions, SAEP, see Morrison, 2006a) are provided in the supplementary material available on the publisher's website (<http://las.sagepub.com/cgi/content/full/52/4/437/DC1>).

Figure 2

(a) Territorial map based on a logistic regression model fitted to pooled monolingual-Western-Canadian-English listeners' perceptual response data. The territorial map indicates the model's modal predicted response category. The spectral and duration dimensions are represented as the x and y dimensions in each panel, and each panel represents one of the three levels of VISC. (b) Probability-surface plot based on the same data as in (a), the height of a surface indicates the model's predicted probability of a response category

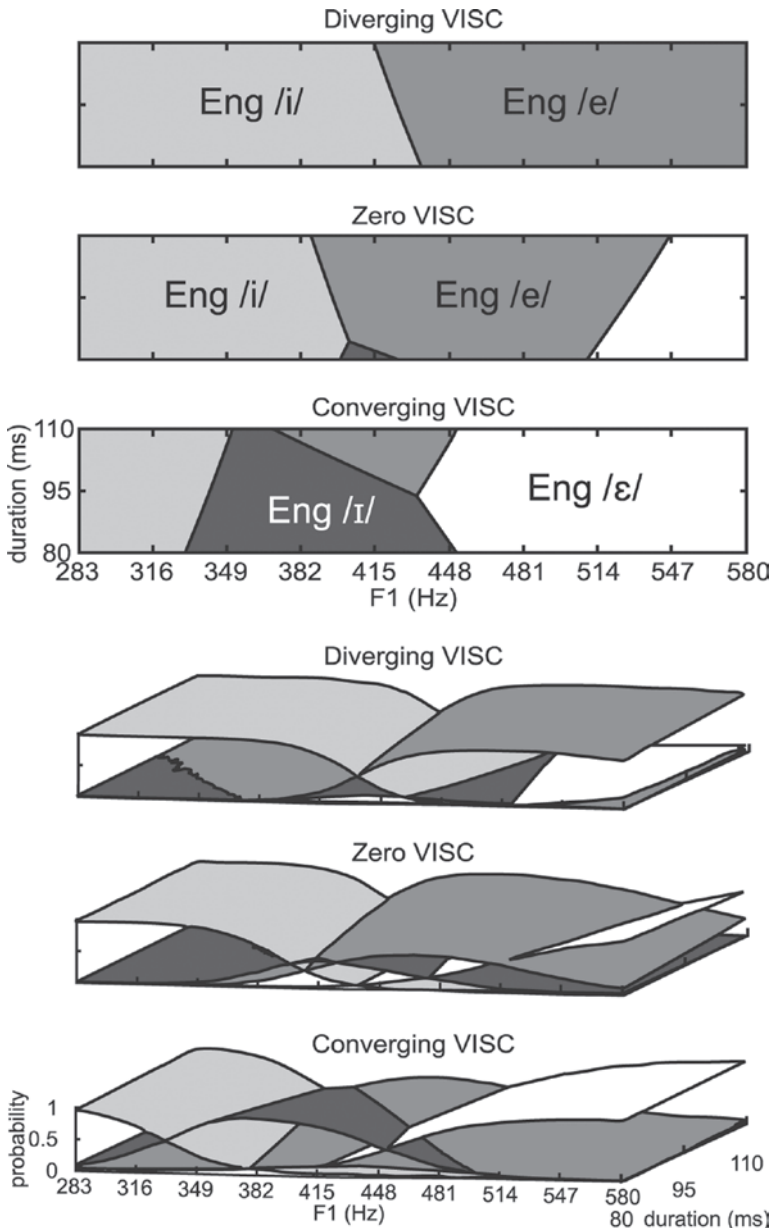
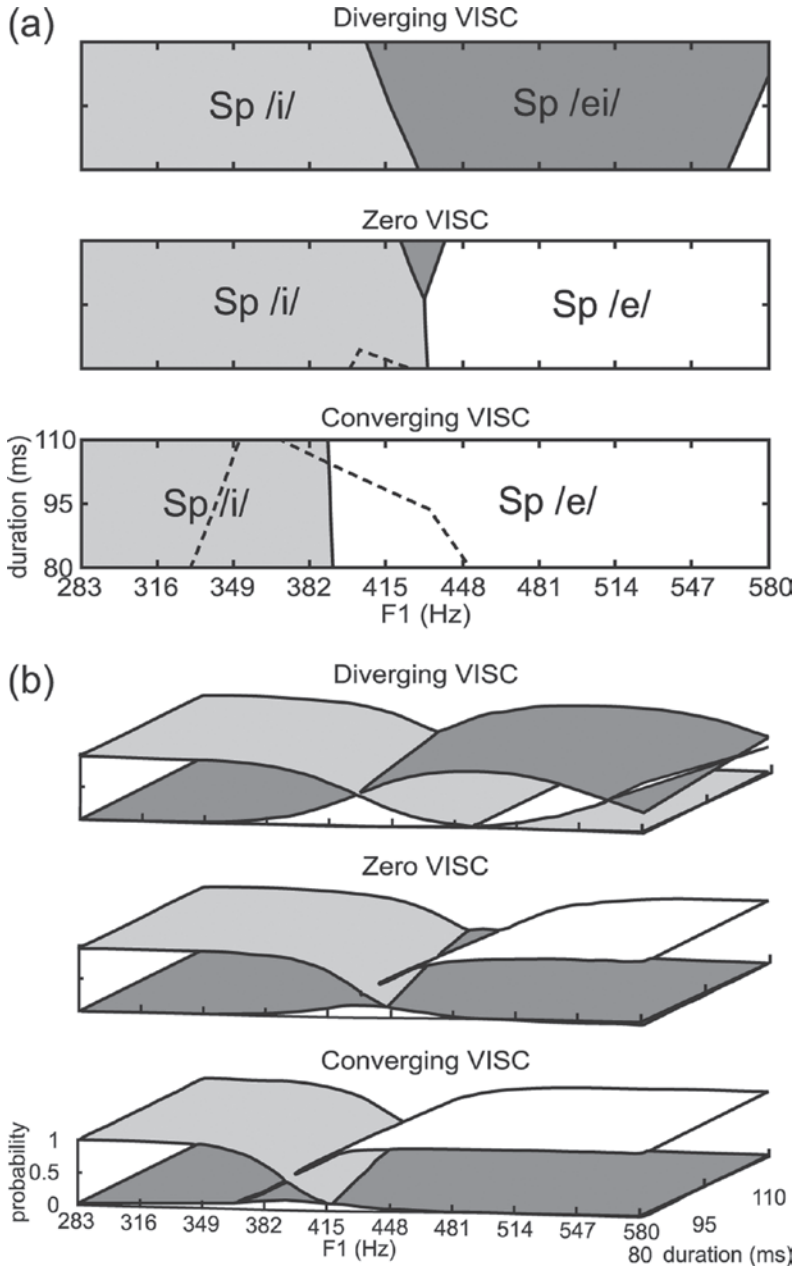


Figure 3

(a) Territorial map based on a logistic regression model fitted to pooled monolingual-Mexican-Spanish listeners' perceptual response data. Dashed lines correspond to the boundary of the modal English /ɪ/ response area in the monolingual-English listeners' territorial map. (b) Probability-surface plot based on the same data as in (a)



Examination of Figure 3 indicates that, for monolingual-Mexican-Spanish listeners, Spanish /ei/ is the modal response over about half of the diverging-VISC portion of the stimulus space. This is as expected given that Spanish /ei/ is a diverging diphthong. The zero-VISC stimulus space is divided between the Spanish /i/ and /e/. This is as expected assuming that these two vowels are monophthongs. Monolingual-Spanish listeners also had Spanish /i/ and /e/ as the modal response in the converging-VISC portion of the stimulus space. Note that Spanish does not have any vowels with acoustic properties similar to those in the converging-VISC portion of the stimulus space, but the results indicate that the monolingual-Spanish listeners perceived these stimuli as more similar to their Spanish /i/ and /e/ categories than to their Spanish /ei/ category. The boundaries between response categories were relatively close to parallel to the duration axis, suggesting that duration played little part in the monolingual-Spanish listeners' vowel perception.

3.2

Initial L2-perception predictions based on monolingual perception

Comparing the monolingual-Spanish and monolingual-English models, predictions can be made as to how L1-Spanish speakers just beginning to learn English would perceive the synthetic stimuli in terms of English categories. The area with English /i/ as the modal response in the monolingual-English listeners' territorial map (Figure 2(a)), corresponded almost exclusively to areas that had Spanish /i/ as the modal response in the monolingual-Spanish listeners' territorial maps (Figure 3(a)). Figure 3(a) includes an overlay of the English /i/ modal response area from Figure 2a; approximately two-thirds of the English /i/ modal response area corresponded to the Spanish /i/ modal response area in the monolingual-Spanish listeners' territorial map. This suggests that L1-Mexican-Spanish listeners will assimilate tokens of Western-Canadian English /i/ and /ɪ/ primarily via a category-goodness-difference assimilation to Spanish /i/, and may have difficulty distinguishing the two English vowels. The predictions for L1-Mexican-Spanish learners of English are consistent with the results of earlier studies of L1-Spanish listeners' perception of Canadian English /i/ and /ɪ/, which suggested substantial confusion between /i/ and /ɪ/ (Morrison, 2002, 2008a).

3.3

L1-Spanish L2-English perception

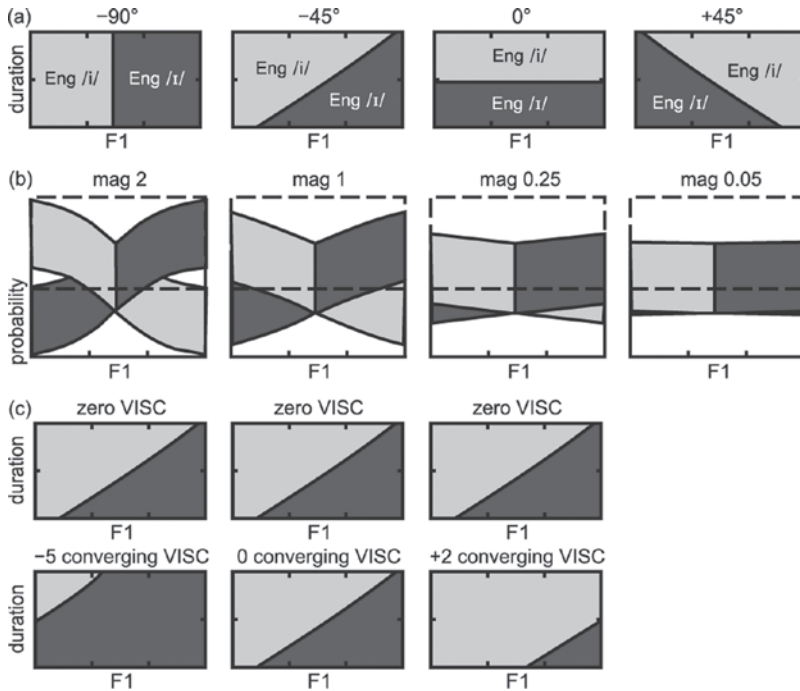
3.3.1 Quantification of the /i/-/ɪ/ boundary

A logistic regression model was fitted to each individual L1-Spanish L2-English listener's data set, and to each individual monolingual-English listener's data set. Since the focus of the present article is the English /i/-/ɪ/ contrast, numerical values characterizing the /i/-/ɪ/ boundary were extracted. For each listener's model, the difference between the estimated values of the stimulus-tuned coefficients for /i/ and /ɪ/ were calculated, for example, $\beta_{i/conv} - \beta_{\text{ɪ}/conv}$. These *contrast coefficients* will be reported directly for the difference in response between zero-VISC and converging-VISC stimuli, and will be reported as polar coordinates for the effect of initial-formant values and duration.¹⁴

14 polar angle = $\arctan((\beta_{i/\text{initial}F} - \beta_{\text{ɪ}/\text{initial}F}) / (\beta_{i/\text{dur}} - \beta_{\text{ɪ}/\text{dur}}))$, polar magnitude = $\sqrt{(\beta_{i/\text{initial}F} - \beta_{\text{ɪ}/\text{initial}F})^2 + (\beta_{i/\text{dur}} - \beta_{\text{ɪ}/\text{dur}})^2}$

Figure 4

(a) Example territorial maps illustrating different angles for the English /i/-/ɪ/ boundary. (b) Example probability-surface plots illustrating different magnitudes for the English /i/-/ɪ/ boundary (the boundary angle is fixed at -90°). (c) Example territorial maps illustrating different converging-VISC values for the English /i/-/ɪ/ boundary (the boundary angle is fixed at -45°)



The use of polar coordinates provides orthogonal measures of the orientation and crispness of the /i/-/ɪ/ boundary in the F_{JND} - dur_{JND} space (Morrison, 2007).

Example territorial maps illustrating boundary orientations are given in Figure 4(a). Orientations were calculated such that an angle of 0° represents the situation where the boundary is perpendicular to the duration dimension and there is a greater predicted probability of /i/ for durations greater than the boundary duration. For angles greater than 0° there is a greater predicted probability of /i/ for F1 values that are greater than the F1 value at the boundary. Because both the spectral and duration contrast coefficients were scaled in JND units, an angle of $\pm 45^\circ$ (gradient ± 1) indicates that spectral and duration differences are equally perceptually important, an angle of $\pm 67.5^\circ$ (gradient ± 2) indicates that spectral cues are twice as important as duration cues, etc.

Example probability-surface plots illustrating boundary crispness are given in Figure 4(b). Large magnitudes represent crisp boundaries as might be expected between well-established categories. Small magnitudes represent fuzzy boundaries such as

might be expected if listeners were asked to give dichotomous labels to stimuli that they perceived as representing gradient within-category differences. Relatively fuzzy boundaries might also be expected if the listeners do not have well-established criteria for distinguishing categories. If the boundary magnitude is very small, the listener may not actually distinguish the two categories and the boundary angle may simply reflect noise in the data (noise or random guessing would also result in poor goodness-of-fit of the model to the data). All four panels in Figure 4(b) have the same boundary angle and the same territorial map, but it would be erroneous to interpret the boundary angle of the right-most panel with the same confidence as that of the left-most panel.

Example territorial maps illustrating the converging-VISC contrast coefficient are given in Figure 4(c). Since English /i/ is a monophthong and English /ɪ/ is produced with converging VISC (Andruski & Nearey, 1992; Morrison, 2006b; Nearey & Assmann, 1986), the converging-VISC contrast coefficient but not the diverging-VISC contrast coefficient is of interest for characterizing the English /i/-/ɪ/ contrast. A positive value for the converging-VISC contrast coefficient indicates a higher probability of /i/ responses in the converging-VISC portion of the stimulus space relative to the probability of /i/ responses in the zero-VISC portion of the stimulus space.

A plot of individual listeners' English /i/-/ɪ/ contrast coefficient values is provided in Figure 5, sample territorial maps illustrating different hypothesized paths and stages of development are provided in Figure 6, and the same data is represented in probability surface plots in Figure 7 (the paths and stages are discussed below).

3.3.2

Hypothesized indirect developmental path

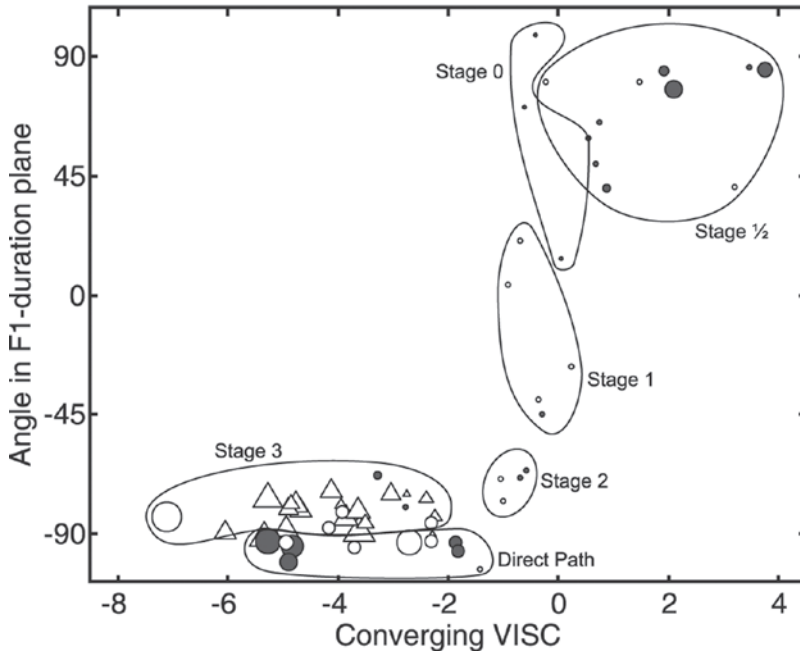
On the basis of a post hoc examination of their English /i/-/ɪ/ contrast coefficient values, 30 of the 40 L1-Spanish L2-English listeners had vowel identification patterns that were consistent with the hypothesized indirect developmental path. Groups of contrast coefficient values that were consistent with each stage of development are indicated in Figure 5.

Sample plots for a listener assigned to Stage 0 are provided in Figures 6(a) and 7(a). The four listeners assigned to Stage 0 had /i/-/ɪ/ boundary magnitudes of less than 0.18. This is consistent with them having little or no ability to distinguish the two English vowels (the range of /i/-/ɪ/ boundary magnitudes for monolingual-English listeners was 0.75–2.55). Note that with such a small magnitude, the orientation of the boundary may be due to noise and is therefore unlikely to be meaningful.

Sample plots for a listener assigned to Stage ½ are provided in Figures 6(b) and 7(b). The 11 listeners assigned to Stage ½ had positive /i/-/ɪ/ boundary angles and positive converging-VISC contrast coefficient values. This is consistent with them labeling stimuli with higher F1, longer duration, or converging VISC (or some combination of these three) as English /i/, and labeling stimuli with lower F1, shorter duration, or non-converging VISC (or some combination of these three) as English /ɪ/. The results were therefore consistent with these listeners distinguishing English /i/ and /ɪ/ via a multidimensional-category-goodness-difference assimilation, and labeling English vowels assimilated as good examples of Spanish /i/ as English /i/ and labeling poor examples as English /ɪ/.

Figure 5

Plot of the F1-duration angle and magnitude (magnitude indicated by marker size), and the converging-VISC contrast for the English /i/-/ɪ/ boundaries from logistic regression models fitted to each individual listener's perceptual response data. Filled circles: L1-Mexican-Spanish listeners. Unfilled circles: L1-non-Mexican-Spanish listeners. Triangles: monolingual-Western-Canadian-English listeners. Groupings indicate sets of L1-Spanish listeners assigned to different stages of the hypothesized indirect developmental path, or to the hypothesized direct developmental path

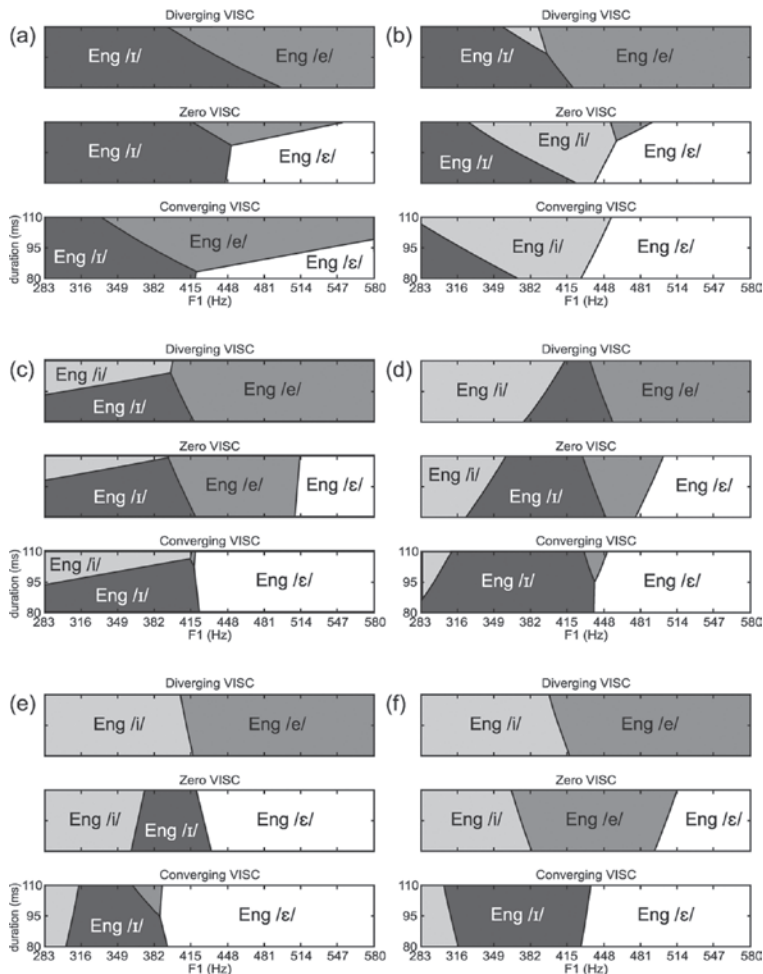


These results replicate the duration and steady-state spectral results reported in Morrison (2008a), and the VISC results are consistent with the predictions regarding the extension of hypothesized multidimensional-category-goodness-difference assimilation to a previously unexamined dimension. The results therefore support the hypothesis that the first means by which L1-Spanish listeners distinguish the English /i/-/ɪ/ contrast is multidimensional, using all available acoustic cues, and not unidimensional, using only duration cues (see the summary of Morrison, 2008a, in the introduction to the present article for arguments as to why the multidimensional-category-goodness-difference assimilation pattern must come before the duration-based perception pattern).

Sample plots for a listener assigned to Stage 1 are provided in Figures 6(c) and 7(c). The five listeners assigned to Stage 1 had /i/-/ɪ/ boundary angles that were relatively close to zero, and also had small converging-VISC contrast coefficient values. This is consistent with these listeners distinguishing the two English vowels almost exclusively on the basis of duration cues.

Figure 6

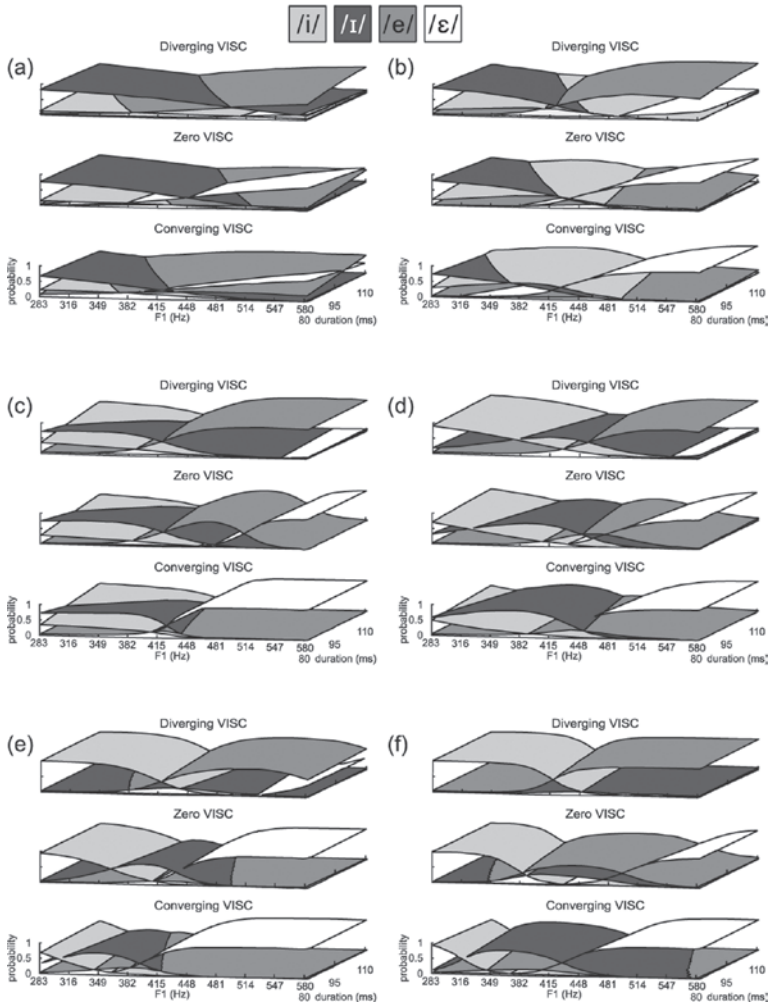
Example territorial maps illustrating the perception of listeners assigned to different stages of the hypothesized indirect developmental path (a–e) and to the hypothesized direct developmental path (f–e). Each plot is based on a logistic regression model fitted to one listener's response data. (a) Listener bs071 assigned to Stage 0, */i/-/i/* angle $+59^\circ$, magnitude 0.09, converging VISC $+0.52$. (b) Listener bs075 assigned to Stage $\frac{1}{2}$, */i/-/i/* angle $+59^\circ$, magnitude 0.35, converging VISC 0.55. (c) Listener bs058 assigned to Stage 1, */i/-/i/* angle -27° , magnitude 0.32, converging VISC 0.24. (d) Listener bs056 assigned to Stage 2, */i/-/i/* angle -77° , magnitude 0.58, converging VISC -1.01 . (e) Listener bs067 assigned to Stage 3, */i/-/i/* angle -85.7° , magnitude 1.10, converging VISC -2.31 . (f) Listener bs068 assigned to the direct developmental path, */i/-/i/* angle -94.7° , magnitude 1.93, converging VISC -4.83 .



Sample plots for a listener assigned to Stage 2 are provided in Figures 6(d) and 7(d). The four listeners assigned to Stage 2 had negative */i/-/i/* boundary angles, and small negative converging-VISC contrast coefficient values. This is consistent with

Figure 7

Example probability-surface plots of the same data as presented in Figure 6



these listeners distinguishing the two English vowels on the basis of both spectral and duration cues, using both sets of cues in the same direction as L1-English listeners. The magnitudes of their contrast coefficient values were, however, generally less than those of L1-English listeners.

Sample plots for a listener assigned to Stage 3 are provided in Figures 6(e) and 7(e). The six listeners assigned to Stage 3 had negative /i/-/ɪ/ boundary angles, and negative converging-VISC contrast coefficient values. The magnitudes of their contrast coefficient values were within the range for L1-English listeners. This was consistent

with these listeners distinguishing the English vowel contrast in an L1-English-like manner.

3.3.3

Hypothesized direct developmental path

Ten of the 40 listeners had a perception pattern that was not consistent with the hypothesized indirect developmental path, but was consistent with the hypothesized direct developmental path. Sample plots for a listener assigned to the hypothesized direct developmental path are provided in Figures 6(f) and 7(f). The listeners assigned to the hypothesized direct developmental path had converging VISC contrast coefficients within the L1-English range, but negative */i/-/ɪ/* boundary angles of magnitude greater than 90°.

This pattern of perception is consistent with several possibilities:

1. A multidimensional-category-goodness-difference assimilation of English */i/* and */ɪ/* to Spanish */i/* with the opposite direction of labeling to that of Stage ½ of the hypothesized indirect developmental path, that is, English vowels perceived as good examples of Spanish */i/* labeled as English */i/* and poor examples labeled as English */ɪ/*.
2. A two-category assimilation of English */i/* and */ɪ/* to Spanish */i/* and */e/*. The use of duration can be explained if Spanish */e/* is longer than Spanish */i/*. (Production results from Morrison, 2006b, indicated that Spanish */e/* is longer than Spanish */i/*.)
3. Given that their */i/-/ɪ/* boundary angles are only slightly beyond -90°, these listeners may have actually been at Stage 3 of the hypothesized indirect developmental path.

These listeners had perception patterns that were relatively similar to those of L1-English listeners, but their use of duration cues was reversed. Exposure to English may provide them with evidence that will allow them to correct their use of duration cues and obtain L1-English-like perception. Listeners assigned to Stage 3 may therefore have reached this stage via either the hypothesized indirect or direct developmental path.

4 Discussion and conclusion

Monolingual-Western-Canadian-English listeners and monolingual-Mexican-Spanish listeners were tested on their perception of synthetic vowel stimuli. These stimuli varied in initial formant values, final formant values, and duration so as to cover the acoustic space occupied by English */i/*, */ɪ/*, */e/*, */ɛ/*, and Spanish */i/*, */e/*, and */ei/*. Results suggested that Mexican-Spanish listeners just beginning to learn Western-Canadian English would be likely to perceive English */i/* and */ɪ/* tokens primarily via a single-category or category-goodness-difference assimilation to Spanish */i/*.

L1-Spanish L2-English listeners' perception was tested on the same synthetic vowel stimuli as had been used with the monolingual listeners. The L1-Spanish L2-English listeners had perception patterns consistent with them initially confusing

English /i/ and /ɪ/, as per the predictions made on the basis of monolingual listeners' perception. Three-quarters (30 of 40) of the L1-Spanish L2-English listeners had perception patterns consistent with the hypothesized indirect developmental path, and of these a third (11) had perception patterns consistent with Stage ½: Multidimensional-category-goodness-difference assimilation to Spanish /i/, with stimuli that were less Spanish-/i/-like in terms of duration or spectral properties (including vowel inherent spectral change) more likely to be labeled as English /i/, and more Spanish-/i/-like stimuli more likely to be labeled as English /ɪ/. The present study therefore replicates the results of Morrison (2008a) for the initial formant values and duration. In addition, the VISC results are consistent with the prediction that initial multidimensional-category-goodness-difference assimilation applies to all available dimensions without substantially privileging one over the others, and therefore that listeners would treat a previously unexamined acoustic dimension in the same manner as they treat the previously examined dimensions. Hence the results support the hypothesis that L1-Spanish L2-English listeners initially distinguish English /i/ and /ɪ/ using all acoustic dimensions available rather than only duration.

The hypothesized indirect developmental path provides an account of the previously observed tendency for L1-Spanish learners of English to use duration cues to distinguish English /i/ and /ɪ/. Earlier proposals by Bohn (1995) and Escudero and Boersma (2004) assumed that duration-based perception was the first means by which L1-Spanish learners of English attempted to distinguish the English /i/-/ɪ/ contrast. If this assumption is shown to be false, the earlier proposals become moot. In Morrison (2008a) and in the present article evidence has been presented of the existence of a perceptual pattern that is argued not to be consistent with duration-based perception being the initial stage of L2-English perception. It has been argued that this pattern of multidimensional-category-goodness-difference assimilation to Spanish /i/ must logically be the first means by which L1-Spanish learners of English attempted to distinguish the English /i/-/ɪ/ contrast (see the introduction to this article). Under the hypothesized indirect developmental path duration-based perception is a secondary stage that naturally emerges from initial multidimensional-category-goodness-difference assimilation. Unlike Bohn (1995) and Escudero and Boersma (2004), the present proposal does not posit any special treatment of either duration or spectral cues, and in fact it is claimed that all the available acoustic dimensions in the present study were used in multidimensional-category-goodness-difference assimilation, which implies that there was not substantial degradation or enhancement of one or other of these types of cue.

A quarter (10 of 40) of the L1-Spanish L2-English listeners had perception patterns consistent with the first stage of the hypothesized direct developmental path. One possible explanation for this stage is that listeners distinguish English /i/ and /ɪ/ via a multidimensional-category-goodness-difference assimilation to Spanish /i/, as in Stage ½ of the hypothesized indirect developmental path, but that they choose the opposite direction of labeling. Upon choosing this direction of labeling, the L1-Spanish listeners' use of spectral properties is positively correlated with L1-English speakers' production of English /i/ and /ɪ/, and since spectral properties are much more important than duration properties for L1-English listeners, the L1-Spanish listeners immediately

have relatively L1-English-like perception. Their use of duration cues is negatively correlated with L1-English speakers' production of English /i/ and /ɪ/, hence exposure to English will potentially reinforce their use of spectral cues and lead them to reverse their use of duration cues. L1-Spanish learners of English on the hypothesized direct developmental path could therefore arrive at L1-English-like perception of English /i/ and /ɪ/, and participants assigned to Stage 3 may have arrived there via either the direct or indirect paths.

In order to demonstrate whether these hypothesized developmental paths truly exist, it will be necessary to conduct longitudinal studies that are able to track changes in individuals' perception over time. A small-scale longitudinal study was reported in Morrison (2006b). There was some evidence in support of the indirect hypothesized developmental path, but the results were not conclusive. Morrison (2006b) also reported on the relationship between the length of residence in Canada of listeners and the stages of the hypothesized indirect developmental path to which they were assigned. Assignment to stages was made solely with reference to listeners' perception patterns, and length-of-residence data was only examined later. He argued that the results were generally consistent with the hypothesized indirect developmental path.

Orthographic and distribution-based explanations for the choice of the direction of labeling were given in Morrison (2008a) and summarized in the introduction to the present article. Since vowels with converging VISC will be less Spanish-/i/ like than vowels without zero VISC, natural tokens of Western-Canadian English /ɪ/ (and synthetic tokens with converging VISC) are less likely to be perceived as good examples of Spanish /i/ than synthetic tokens with zero VISC. In light of this, it seems less likely that good examples of Spanish /i/ are labeled as English /ɪ/ because of a posited greater average similarity between Spanish /i/ and English /ɪ/ tokens than between Spanish /i/ and English /i/ tokens. The orthographic explanation therefore seems more likely than the distribution-based explanation. English /ɪ/ is frequently represented in English orthography by the letter 'i', which represents Spanish /i/ in Spanish orthography. Orthography may therefore bias L1-Spanish L2-English listeners to associate English /ɪ/ with Spanish /i/, to label English vowels perceived as good matches for Spanish /i/ as English /ɪ/, and to label English vowels perceived as poorer matches for Spanish /i/ as English /i/. Whether individual learners follow the indirect or direct developmental path may depend on the extent to which they are influenced by orthography, which could depend on factors such as individual preferences for auditory versus visual learning, or the extent of formal reading-and-writing-based classroom education versus immersion in an L1-English-speaking environment at the early stages of L2 learning.

Although the present study has only examined L2 learners of Western-Canadian English, I would argue that multidimensional-category-goodness-difference assimilation being the first developmental stage can potentially be generalized to other English dialects in which /i/ and /ɪ/ are produced with substantial differences in both spectral and duration properties (and with Spanish dialects where both English vowels are primarily assimilated to Spanish /i/). However, across English dialects differences in the exact mixture of spectral and duration cues (including differences in the degree of VISC) may lead to differences in the details of developmental patterns.

L1-dialect did not appear to be related to whether a listener was assigned to one path or the other: Half the participants (five of ten) assigned to the first stage of the direct path were Mexican, as were approximately half the participants (16 of 30) assigned to the indirect path. However, Morrison (2008b) measured monolingual-North-Central-Peninsular-Spanish listeners' perception of the same synthetic stimuli as used in the present study, and results suggested that they may be more likely than monolingual-Mexican-Spanish listeners to assimilate tokens of English /i/ and /ɪ/ via two category assimilation to Spanish /i/ and /e/ respectively (similar to Peruvian-Spanish listeners' perception of Scottish-English /i/ and /ɪ/, Escudero, 2005, §1.2.2). Under such a scenario, L1-North-Central-Peninsular-Spanish learners of English would not be expected to have any difficulty perceiving the difference between Western-Canadian-English /i/ and /ɪ/. (Note, however, that the two Peninsular-Spanish listeners in the present study were assigned to Stages ½ and 1 of the hypothesized indirect developmental path.)

In order to determine the effects of L1-Spanish and L2-English dialect on the learning of English /i/ and /ɪ/ in different dialects of English, additional research would be necessary that includes systematic comparison of monolingual-Spanish listeners' perception of vowels from their own dialect of Spanish and from different dialects of English, as well as systematic comparison of the perception of English vowels by groups of L1-Spanish-L2-English listeners with different combinations of L1 and L2 dialect.

References

- ÁLVAREZ GONZÁLEZ, J. A. (1980). *Vocalismo español y vocalismo inglés* [Spanish and English vowels]. Unpublished Doctoral dissertation, Universidad Complutense de Madrid.
- ANDRUSKI, J. E., & NEAREY, T. M. (1992). On the sufficiency of compound target specification of isolated vowels in /bVb/ syllables. *Journal of the Acoustical Society of America*, **91**, 390–410.
- ASSMANN, P. F., & KATZ, W. F. (2005). Synthesis fidelity and time-varying spectral change in vowels. *Journal of the Acoustical Society of America*, **117**, 886–895.
- ASSMANN, P. F., NEAREY, T. M., & HOGAN, J. T. (1982). Vowel identification: Orthographic, perceptual, and acoustic aspects. *Journal of the Acoustical Society of America*, **71**, 975–989.
- BAKER, W., & TROFIMOVICH, P. (2005). Interaction of native- and second-language vowel system(s) in early and late bilinguals. *Language & Speech*, **48**, 1–27.
- BEST, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp.171–204). Timonium, MD: York Press.
- BOHN, O.-S. (1995). Cross-language speech perception in adults: First language transfer doesn't tell it all. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp.279–304). Timonium, MD: York Press.
- BOHN, O.-S., & FLEGE, J. E. (1993). Perceptual switching in Spanish/English bilinguals. *Journal of Phonetics*, **21**, 267–290.
- CEBRIAN, J. (2006). Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics*, **34**, 372–387.
- ESCUDERO, P. (2000). *Developmental patterns in the adult L2 acquisition of new contrasts: The acoustic cue weighting in the perception of Scottish tensellax vowels by Spanish speakers*. Unpublished Master's thesis, University of Edinburgh, Scotland, UK.

- ESCUADERO, P. (2005). *Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization*. Doctoral dissertation, University of Utrecht, The Netherlands. LOT.
- ESCUADERO P., & BOERSMA, P. (2002). The subset problem in L2 perceptual development: Multiple-category assimilation by Dutch learners of Spanish. In B. Skarabela, S. Fish, & A. H.-J. Do (Eds.), *Proceedings of the 26th Annual Boston University Conference on Language Development* (pp.208–219). Somerville, MA: Cascadilla.
- ESCUADERO, P., & BOERSMA, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, **26**, 551–585.
- FLEGE, J. E. (1991). The interlingual identification of Spanish and English vowels: Orthographic evidence. *Quarterly Journal of Experimental Psychology*, **43**, 701–731.
- FLEGE, J. E., BOHN, O.-S., & JANG, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, **25**, 437–470.
- FRIEDA, E. M., & NOZAWA, T. (2007). You are what you eat phonetically: The effect of linguistic experience on the perception of foreign vowels. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp.79–96). Amsterdam: John Benjamins.
- GOTTFRIED, M., MILLER, J. D., & MEYER, D. J. (1993). Three approaches to the classification of American English diphthongs. *Journal of Phonetics*, **21**, 205–229.
- HILLENBRAND, J. M., CLARK, M. J., & NEAREY, T. N. (2001). Effect of consonant environment on vowel formant patterns. *Journal of the Acoustical Society of America*, **109**, 748–763.
- HØJEN, A., & FLEGE, J. E. (2006). Early learners' discrimination of second-language vowels. *Journal of the Acoustical Society of America*, **119**, 3072–3084.
- HOLT, L. L., & LOTTO, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *Journal of the Acoustical Society of America*, **119**, 3059–3071.
- KEWLEY-PORT, D. (2001). Vowel formant discrimination II: Effects of stimulus uncertainty, consonantal context, and training. *Journal of the Acoustical Society of America*, **110**, 2141–2155.
- KLATT, D. H., & KLATT, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America*, **87**, 820–857.
- KONDAUROVA, M. V., & FRANCIS, A. L. (2008). The relationship between native allophonic experience with vowel duration and perception of the English tense/lax vowel contrast by Spanish and Russian listeners. *Journal of the Acoustical Society of America*, **124**, 3935–3971.
- MØLLER GLASBRENNER, M. (2005). *Vowel identification by monolingual and bilingual listeners: Use of spectral change and duration cues*. Unpublished Doctoral dissertation, University of South Florida.
- MORRISON, G. S. (2002). Perception of English /i/ and /ɪ/ by Japanese and Spanish listeners: Longitudinal results. In G. S. Morrison & L. Zsoldos (Eds.), *Proceedings of the North West Linguistics Conference 2002* (pp.29–48). Burnaby, BC, Canada: Simon Fraser University Linguistics Graduate Student Association.
- MORRISON, G. S. (2006a). An adaptive sampling procedure for speech perception experiments. In *Proceedings of the Ninth International Conference on Spoken Language Processing* (pp.857–860). Bonn, Germany: ISCA.
- MORRISON, G. S. (2006b). *L1 & L2 production and perception of English and Spanish vowels: A statistical modelling approach*. Unpublished Doctoral dissertation, University of Alberta, Edmonton, Alberta, Canada.

- MORRISON, G. S. (2007). Logistic regression modelling for first- and second-language perception data. In M. J. Solé, P. Prieto, & J. Mascaró (Eds.), *Segmental and prosodic issues in Romance phonology* (pp.219–236). Amsterdam: John Benjamins.
- MORRISON, G. S. (2008a). L1-Spanish speakers' acquisition of the English /i/-/ɪ/ contrast: Duration-based perception is not the initial developmental stage. *Language & Speech*, **51**, 285–315.
- MORRISON, G. S. (2008b). Perception of synthetic vowels by monolingual Canadian-English, Mexican-Spanish, and Peninsular-Spanish listeners. *Canadian Acoustics*, **36**(4), 17–23. [Typesetting errata published in 37(1), 34.]
- MORRISON, G. S., & ESCUDERO, P. (2007). A cross-dialect comparison of Peninsula- and Peruvian-Spanish vowels. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences: Saarbrücken 2007* (pp.1505–1508). Saarbrücken, Germany: Universität des Saarlandes.
- MORRISON, G. S., & NEAREY, T. M. (2007). Testing theories of vowel inherent spectral change. *Journal of the Acoustical Society of America*, **122**, EL15–EL22.
- NEAREY, T. M. (1989). Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustical Society of America*, **85**, 2088–2113.
- NEAREY, T. M. (1990). The segment as a unit of speech perception. *Journal of Phonetics*, **18**, 347–373.
- NEAREY, T. M. (1997). Speech perception as pattern recognition. *Journal of the Acoustical Society of America*, **101**, 3241–3254.
- NEAREY, T. M., & ASSMANN, P. F. (1986). Modeling the role of vowel inherent spectral change in vowel identification. *Journal of the Acoustical Society of America*, **80**, 1297–1308.
- NOTEBOOM, S. G., & DOODEMAN, G. J. N. (1980). Production and perception of vowel length in spoken sentences. *Journal of the Acoustical Society of America*, **67**, 276–287.
- POLS, L. C. W. (1977). *Spectral analysis and identification of Dutch vowels in monosyllabic words*. Doctoral dissertation, University of Amsterdam.
- SMITS, R., SERANO, J., & JONGMAN, A. (2006). Categorization of sounds. *Journal of Experimental Psychology: Human Perception and Performance*, **32**(3), 733–754.
- TRAUNMÜLLER, H. (1990). Analytical expressions for the tonotopic sensory scale. *Journal of the Acoustical Society of America*, **88**, 97–100.
- WANG, X., & MUNRO, M. J. (1999). The perception of English tense-lax vowel pairs by native Mandarin speakers: The effect of training on attention to temporal and spectral cues. In J. J. Ohala et al. (Eds.), *Proceedings of the 14th International Congress of Phonetic Sciences: San Francisco 1999* (pp.125–128). Berkeley: University of California Berkeley.