TONGUE MOVEMENT IN A SECOND LANGUAGE: THE CASE OF SPANISH /ei/-/e/ FOR ENGLISH LEARNERS OF SPANISH

Isabelle Darcy and Joan C. Mora

Indiana University, USA & Universitat de Barcelona, Spain
idarcy@indiana.edu, mora@ub.edu

ABSTRACT

This study examines the amount of tongue movement in the productions of native Spanish speakers and native English learners of Spanish for the Spanish diphthong/monophthong contrast /ei/-/e/. We hypothesized that English learners would use their native English category /e^{I}/ for both Spanish vowels. However, results show that against our prediction, for both Spanish vowels, learners produced *less* tongue movement than was expected if they used their L1 category. Instead, they produced both vowels as the monophthong /e/, effectively neutralizing the contrast in terms of tongue movement.

Keywords: L2 Spanish; diphthong; monophthong; production; tongue movement

1. INTRODUCTION

Vowel systems in Spanish and English are very different. Most dialects of Spanish have five monophthongs /i, e, a, o, u/, as well as a number of diphthongs including /ai, ia, au, oi, ei, ie, eu, ue/ [17]; see also [18], for a detailed overview of their characteristics. The American English vowel system is usually described as having three diphthongs /aɪ/, /aʊ/, and /ɔɪ/), two "phonetic diphthongs" (/e^{I}/ and /o^{w}/) as well as nine monophthongs [6], and thus unlike Spanish, no contrast between /e/ and /e^{I}/. Diphthongs are characterized by substantial vowel inherent spectral changes (VISC) in formant structure, reflecting tongue movement. By contrast, monophthongs, while not completely lacking movement, display much smaller spectral changes [19].

The Spanish vowel system is usually thought to pose relatively few problems for English native speakers who learn Spanish as a second language (L2). An exception to this are the diphthongs, yet surprisingly little research has examined this area. A study that examined the acquisition of various diphthongs within and between words suggests that Spanish diphthongs that have no counterpart in English tend to be initially realized as a hiatus by American English learners of Spanish. However, no quantitative measurements of the vowels' acoustic properties in learners' speech were undertaken [16].

Generally speaking, few studies have examined the acquisition of tongue movement in L2 learners, and most of them are directed at learners of English who need to acquire the phonetic diphthongs /e^{I}/ and /o^{u}/. In this paper, we focus on learners of Spanish, and on the contrast between the Spanish diphthong /ei/ and the Spanish monophthong /e/. This contrast is of specific interest because, unlike the other Spanish diphthongs, /ei/ might be confused with the English vowel /e^{I}/, hence, possibly creating perception and production difficulties. We examine the production of 26 American English learners of Spanish and 9 Spanish native speakers, who produced both vowels in a delayed sentence repetition task. Their productions were acoustically measured to extract tongue movement scores, and compared to L2 productions of the Spanish contrast.

2. BACKGROUND

Studies of the acquisition of accurate tongue movement in L2 are rare. A representative one is Flege, Schirru and MacKay [9], who examined the production of English /e^{I}/ (e.g. in 'face' [fe^{I}s]) by Italian-English bilinguals, showing that late bilinguals produced less tongue movement than native speakers, likely because they were initially using their Italian monophthongal category /e/; by contrast, early bilinguals produced more tongue movement for English /e^{I}/ compared to native English speakers.

Several perception studies have noted the similarity of the English /e^{I}/ to diphthongs, such as the Italian /ei/ [10] as well as the Spanish /ei/ [12, 18]. In [10] for example, just one of four groups of Italian–English bilinguals examined (early bilinguals who seldom used Italian) were able to discriminate Italian /ei/ from English /e^{I}/ tokens at a significantly above-chance rate (A′ = 0.69). The native English listeners who did not know Italian discriminated the contrast with a mean A′ score of 0.67. This relatively low sensitivity level, despite being above chance, is the lowest A′ score obtained by the native English speakers out of all contrasts tested, where it is usually around 0.90.

Acoustically, the Spanish diphthong /ei/ is relatively similar to the English vowel /e^{I}/, which is usually realized as a phonetic diphthong with

substantial tongue movement across English varieties [13, 19], including the Midland variety (Indiana/Ohio) spoken by the participants in this study (see [6]). However, compared to English /eᴵ/, Spanish /ei/ is longer and has greater VISC [18].

A study using multi-dimensional scaling [12] evaluated the perceptual similarity of Spanish and English vowels, and observed that for less proficient Spanish learners of English, both the Spanish monophthong /e/ and the English /eᴵ/ were perceived as relatively similar. In addition, perceptual assimilation data from a related language, Catalan, have shown that across English varieties, English /eᴵ/ is consistently mapped onto Catalan /ei/, and less often also onto Catalan /e/ [4]. In [5], Canadian English (CE) listeners identified Catalan vowels. Whereas Catalan /ei/ was consistently identified as CE /eᴵ/ (95%), Catalan /e/ was identified as CE /eᴵ/ 16% of the time, with a moderate goodness rating (3.4 out of 7). A similar pattern might therefore apply to English learners of Spanish, who might perceive both Spanish /e/ and /ei/ as relatively similar to their native English vowel /eᴵ/, but /ei/ more so than /e/. This in turn might also result in difficulty producing Spanish /e/ vs. /ei/ (e.g., *pena* [pena] "shame", vs. *peina* [peina] "(he) combs"), perhaps using /eᴵ/ to realize both vowels.

To our knowledge, only one study [18] has examined how English learners of Spanish (n = 26) perceive this contrast (through perceptual assimilation) and how they produce it. Their production accuracy of L2-Spanish vowels was assessed via classification of individual vowels tokens by a canonical discriminant function analysis (CDFA) model trained on L1-Spanish speakers' L1-Spanish productions. Based on the SLM [8] and taking into account the acoustic similarity between English /eᴵ/ and Spanish /ei/ in terms of F1, F2, VISC and duration, Morrison [18] predicted that Spanish /ei/ would initially be perceived (and produced) using their English vowel category /eᴵ/. The perceptual confusion matrix (p. 75) showed that Spanish /ei/ was indeed mapped onto English /eᴵ/ (76.9%), as well as onto /i/ (23.1%); Spanish /e/ was more ambiguous, being mapped onto English /ɪ/, /ɛ/ and /eᴵ/ (13.3%, 42.2% and 44.4% respectively). The learners' productions were then categorized by the CDFA model and resulted in some mis-classifications: The most common one involved classifying Spanish /ei/ as Spanish /e/. In addition, for 4 speakers overall, there were instances of misclassification of Spanish /e/ as /ei/. These results suggest perceptual difficulties with this contrast and are strongly indicative of difficulties in learners' ability to produce the appropriate tongue movement in distinguishing the diphthong/monophthong contrast.

However, no statistical comparison of the specific amount of tongue movement was conducted for the diphthong vs. the monophthong, even though the results indicate that L2 learners might be producing little tongue movement for the diphthong. To start filling this gap, we examined tongue movement in 26 L2-Spanish learners' productions of the /ei/-/e/ contrast.

2.1. Predictions

For L2-Spanish learners, the perception and production of the non-native /ei/-/e/ contrast was hypothesized to be difficult based on perceptual mapping data [18]. We follow this initial hypothesis that L2 learners of Spanish would mainly use their English vowel category /eᴵ/ for realizing both /ei/ and /e/ in Spanish (see also [24], for a similar hypothesis).

Specifically, we predicted that both the monophthong /e/ and the diphthong /ei/ would be produced similarly, more specifically like a long, diphthongized vowel similar to English /eᴵ/ (i.e. both [pena] "shame", and *peina* [peina] "(he) combs" would be produced as [peᴵna]). That is, compared to native speakers', we expected that L2 learners' monophthongs would display more tongue movement and be longer. Conversely, their diphthongs were expected to display less tongue movement, and be shorter.

2.2. Vowel inherent spectral change measurements

To estimate spectral change in vowels, sampling of formants at two locations near vowel onset and offset (i.e., 20%–80% or 20%–70%) or three locations including the vowel midpoint (the 50% point) has been commonly used in acoustic studies (e.g. [7, 9, 12, 14, 15, 18, 19]). Yet, a denser multiple sampling at four [11], five [13], nine [1], or 16 equidistant points [23], or also in 25 ms intervals (number depending on vowel duration, [21]) has also been done to estimate VISC in order to obtain more (dialect-specific) information about formant trajectory and timing.

Since the specific rate of change or the formant trajectory variability are not of direct relevance to the question at hand, we used three measurement points to estimate formant frequencies (F1 and F2) in the present study. This is expected to provide a sufficient estimation of the amount of tongue movement.

3. METHOD

3.1. Participants

A group of L1-English learners of Spanish (n = 26; 22 females; tested in Bloomington, Indiana, USA; mean age = 19.8, *SD* = 1.0) and nine L1-Spanish native speakers as controls (4 females; from Seville, Spain; mean age = 26.4, *SD* = 9.8) participated in exchange for a small payment. Table 1 presents the descriptive statistics for the main background variables about the L2 learners.

Table 1: Summary of demographic variables collected for L2 learners.

Measure	N	Mean	*SD*	Min.	Max.
Motivation	23	7.29	.95	5.33	8.78
Current L2 use (max. 36)	26	8.00	6.90	0	28
Self-evaluation	23	3.97	.61	3.0	5.0
LoR abroad (in weeks)	26	9.34	24.55	0	110
Years of study	26	7.81	3.92	0	15
Age of First Exposure to L2	23	8.48	4.14	1	14
Age of First Use of L2	23	10.70	3.78	3	15

Note. LoR = Length of residence in L2-speaking country. Variable Ns are due to missing data from questionnaires

A motivation score was obtained by averaging each participant's ratings on nine items about their desire to learn Spanish (rated on a 9-point Likert scale: 1=strongly agree; 9=strongly disagree). Current L2 use was a score (between 0 and 36) obtained by adding up participants' selected level of intensity of L2 use (0=0%; 1=1-25%; 2=26-50%; 3=51-75%; 4=76-100%) on nine contexts of language use (e.g., with friends, at home/work, media). L2 self-evaluation was a self-reported estimation of participants' ability to speak spontaneously, understand, read and write the L2, using the following descriptions (recoded as numeric score): very poorly (= 1); poorly (= 2); passably (= 3); well (= 4); very well (= 5). A mean self-evaluation score was obtained by averaging the four scores of each participant.

Participants were tested as part of a larger project that included cognitive measures not reported here[1], and all passed a pure-tone audiometry test at octave frequencies between 500 and 8000 Hz at 20 dB HL [20]. The order of tasks was the same for all participants with slight adjustments (e.g., only the L2 learners participated in L2 vocabulary tasks).

3.2. Stimuli and procedure

Each participant took part in a delayed sentence repetition task [22], either in L2 in the case of learners, or in L1 in the case of the controls. The participants sat in a sound-isolated recording booth equipped with headphones and a computer screen. They heard a question (prompt, voice 1), followed after 250 ms by an answer (response, voice 2). After a 500 ms delay, the prompt was presented again, and the participants had to repeat aloud the response heard previously. The written sentences appeared on the screen together with the first auditory presentation of the prompt/response pair, and disappeared for the second presentation of the prompt, and the recording of the answer. All L2 learners received instructions in English, and completed a warm-up prompt in English before moving on to Spanish. Native speakers completed the set in Spanish. The /ei/-/e/ contrast was examined together with a consonantal contrast (/d/-/ɾ/) not reported here; both contrasts were also examined in perception. There were four pairs of words to implement this contrast (Table 2). The task took about five minutes to complete.

Table 2: Sentences (responses) used to elicit the contrast, with English gloss.

Sí, vale la **pena** ir.	*Yes, it's worth going.*
Cuando se **peina** el pelo sí.	*When he combs his hair, yes.*
Creo que el **reno** es más grande.	*I think the reindeer is bigger.*
Se llama "el **reino** de los cielos".	*It's called "the kingdom of the skies."*
Si quieres **vente** conmigo.	*If you want, come with me.*
Tengo **veinte** años y dos meses.	*I'm twenty years and two months old.*
Es la **maceta** que se ha roto.	*It's the flowerpot that fell and broke.*
Un buen **aceite** de oliva.	*A good olive oil.*

3.3. Acoustic analysis

Three measurement points (MP) were placed 20%, 50% and 80% into the vowels, and the mean values for F1, F2, and F0 were extracted from a 10ms window centered at the three MPs. These frequency measures were first converted to Bark, and then a Bark-distance metric was applied subtracting B0 from B1 (B1-B0) for tongue height and B1 from B2 (B2-B1) for degree of tongue fronting, where B stands for Bark-converted frequency (Hz) values [2, 3]. We measured the amount of tongue movement in the vowel by computing the Euclidean distance between the 20% and the 50% MPs and between the 50% and the 80% MPs. Then we added up the two Euclidean distances and used this spectral distance

score as a measure of tongue movement, as represented on the Bark-normalized vowel space. We also assessed whether the duration ratio of monophthongs to diphthongs were comparable across speaker groups.

4. RESULTS

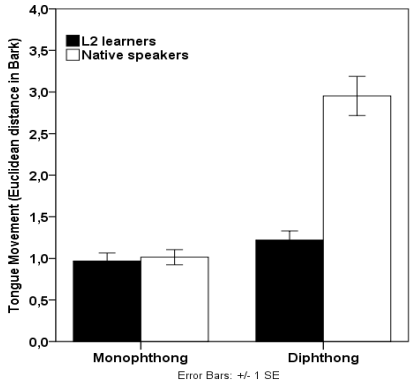
In terms of vowel duration, diphthongs were produced with significantly longer durations than monophthongs by both L2-Spanish learners ($t(25) = 6.27$, $p < 0.001$) and native controls ($t(8) = 8.21$, $p < 0.001$). The duration ratio (diphthong/monophthong) was significantly larger for native speakers than it was for L2 learners ($t(33) = -3.04$, $p = 0.005$; see Table 3).

Table 3: Overview of measurements for vowels and consonants for L2-Spanish learners and native controls (*SD* in parentheses).

L1	Vowel	Duration (ms)	Duration ratio (ms)	Tongue movement (Bark)
English	/ei/	116 (15)	1.19 (0.15)	1.22 (0.55)
(*n*=26)	/e/	99 (16)		0.97 (0.49)
Spanish	/ei/	113 (13)	1.36 (0.14)	2.95 (0.71)
(*n*=9)	/e/	83 (7)		1.01 (0.27)

The two groups also differed in the amount of tongue movement during vowel production. L2-Spanish learners were found to produce the Spanish diphthong /ei/ with much less tongue movement than native controls did, as shown in Figure 1. However, contrary to our predictions, L2-Spanish learners did not produce the Spanish monophthong /e/ as an English-like diphthongized vowel. Instead, the L2 learners produced both vowels as a monophthongal /e/, but distinguished them through duration.

Figure 1: Amount of tongue movement in Spanish /e/ (monophthong) and /ei/ (diphthong).



An ANOVA with *vowel* (monophthong, diphthong) as a within-subjects factor and *group* (L2 learner, native speaker) as a between-subjects factor revealed main effects of *vowel* ($F(1,33) = 56.91$, $p < 0.001$, $\eta^2 = .63$) and *group* ($F(1,33) = 40.83$, $p < 0.001$, $\eta^2 = .55$), and a significant *vowel x group* interaction ($F(1,33) = 39.11$, $p < 0.001$, $\eta^2 = .54$). The interaction indicates that while the groups did not differ in the amount of tongue movement for /e/, native controls produced significantly larger tongue movement for /ei/ ($M = 2.95$, min. = 2.14, max. = 4.13) than L2-Spanish learners did ($M = 1.22$, min. = 0.24, max. = 2.36), $t(33) = -7.56$, $p < 0.001$.

A two-tailed paired t-test confirmed that the learners did not produce a significant difference in tongue movement between the /e/ (M = .97) and /ei/ (M = 1.22; $t(25) = 1.76$, $p > 0.05$).

5. DISCUSSION AND CONCLUSIONS

We found that the L1-English learners of Spanish did not make a clear difference in the amount of tongue movement to distinguish between monophthong /e/ and the diphthong /ei/ in Spanish. Against our prediction however, they did not maintain an intermediate amount of tongue movement: rather, they realized both vowels with relatively little movement which rendered them very similar to Spanish monophthongs, thus resulting in the partial neutralization of this contrast, since only duration differentiates both vowels in learners' speech.

Hence, the hypothesis (see also [18]) that English learners of Spanish use their English /e^{I}/ category to realize the Spanish /ei/ (and /e/) was not supported. It appears that learners have merged the contrast in production and use the Spanish monophthong category for both. Of course, this finding needs to be compared with L1 tongue movement data for their English /e^{I}/ before being able to conclude that they are not using their L1 /e^{I}/ category but rather the L2 /e/ category with a duration distinction.

Several reasons might explain the use of the L2 category. First, the frequency of the contrast and its functional load might play a role: words containing /ei/ are less frequent than those containing /e/. Hence, perceptual learning might favor the use of monophthongal categories first, while allowing learners to differentiate the two using duration. A second possibility could be a perceptual problem. Data not reported here suggest that for L2 learners of Spanish, distinguishing the contrast is difficult in the first place. We found an error rate of 17.6% on this contrast, which is significantly higher than for the native controls (10.1%). In the current experiment, participants had to rely mainly on an accurate perception of the vowel in the word, in order to reproduce it correctly – since orthography was not used while they were producing it.

Finally, an important implication of these findings is that predictions based on acoustic similarity and perceptual assimilation data may not be fully able to predict the extent to which an L1 category will be used in L2 speech. More research is needed to understand the additional factors that underlie the present findings.

6. ACKNOWLEDGEMENTS

We are deeply grateful to Ryan Lidster, Danielle Daidone, and Franziska Krüger for their valuable feedback on previous versions of this paper.

7. REFERENCES

[1] Adank, P., van Hout, R., and Smits, R. (2004). An acoustic description of the vowels of Northern and Southern Standard Dutch. *J. Acoust. Soc. Am.* 116, 1729–1738.

[2] Baker, W., & Trofimovich, P. (2005). Interaction of native- and second-language vowel system(s) in early and late bilinguals. *Lang. Speech* 48, 1–27.

[3] Bohn, O.-S., & Flege, J. E. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Appl. Psycholing.* 11, 303–328.

[4] Cebrian, J. (2006). Experience and the use of duration in the categorization of L2 vowels. *J. Phon*, 34, 372–387.

[5] Cebrian, J. (2009). Effects of native language and amount of experience on crosslinguistic perception. *J. Acoust. Soc. Am.* 125, 2775(A).

[6] Clopper, C. G., Pisoni, D. B., & De Jong, K. (2005). Acoustic characteristics of the vowel systems of six regional varieties of American English. *J. Acoust. Soc. Am.* 118, 1661–1676.

[7] Ferguson, S. H., and Kewley-Port, D. (2002). Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. *J. Acoust. Soc. Am.* 112, 259–271.

[8] Flege, J. E. (1995). Second language speech learning theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp 233–277). Timonium, MD: York Press.

[9] Flege, J. E., Schirru, C., & MacKay, I. R. (2003). Interaction between the native and second language phonetic subsystems. *Speech comm.* 40, 467–491.

[10] Flege, J.E., MacKay, I.R.A., Meador, D., (1999). Native Italian speakers' production and perception of English vowels. *J. Acoust. Soc. Am.* 106, 2973–2987.

[11] Fox, R. A. (1983). Perceptual structure of monophthongs and diphthongs in English. *Lang. Speech* 26, 21–60.

[12] Fox, R. A., Flege, J. E., & Munro, M. J. (1995). The perception of English and Spanish vowels by native English and Spanish listeners: A multidimensional scaling analysis. *J. Acoust. Soc. Am.* 97, 2540–2551.

[13] Fox, Robert Allen, & Jacewicz, Ewa. (2009). Cross-dialectal variation in formant dynamics of American English vowels. *J. Acoust. Soc. Am.* 126, 2603–2618

[14] Hillenbrand, J. M., Clark, M. J., and Nearey, T. M. (2001). Effects of consonantal environment on vowel formant patterns. *J. Acoust. Soc. Am.* 109, 748–763.

[15] Hillenbrand, J. M., Getty, L. A., Clark, M. J., and Wheeler, K. (1995). Acoustic characteristics of American English vowels. *J. Acoust. Soc. Am.* 97, 3099–3111.

[16] Lord, G. (2005). (How) Can we teach foreign language pronunciation? On the effects of a Spanish phonetics course. *Hispania*, 557–567.

[17] Martínez Celdrán, E. (1995). En torno a las vocales del español: Análisis y reconocimiento [Concerning the vowels of Spanish: Analysis and recognition]. *Est. de Fonética Exper.* 7, 195–218.

[18] Morrison, G. S. (2006). L1 & L2 production and perception of English and Spanish vowels: A statistical modelling approach. *Ph.D. dissertation*, University of Alberta, Canada.

[19] Nearey, T. M., and Assmann, P. F. (1986). Modeling the role of inherent spectral change in vowel identification. *J. Acoust. Soc. Am.* 80, 1297–1308.

[20] Reilly, J., Troiani, V., Grossman, M., & Wingfield, R. (2007). An introduction to hearing loss and screening procedures for behavioral research. *Beh. Res. Meth.* 39, 667–672.

[21] Scobbie, J. M., Turk, A., & Hewlett, N. (1999). Morphemes, Phonetics and Lexical Items: The Case of the Scottish Vowel Length Rule. In *Proc. 14th ICPhS* San Francisco. (Vol. 2, pp. 1617–1620). International Congress of Phonetic Sciences.

[22] Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *SSLA*, 28, 1-30.

[23] Van Son, R. J. J. H., and Pols, L. C. W. (1992). Formant movements of Dutch vowels in a text, read at normal and fast rate. *J. Acoust. Soc. Am.* 92, 121–127.

[24] White, Laurence, & Mattys, Sven L. (2007). Calibrating rhythm: First language and second language studies. *J. Phon*, 35, 501–522.

[1] Participants took part in an audiometry test, an inhibitory control task, an attention control task, an ABX categorization task, a production task, and a working memory task. Finally, they took a vocabulary test, and filled out a background questionnaire.