

# Perception and Production of Spanish Vowels by English Speakers

Geoffrey Stewart Morrison

University of Ottawa, Department of Linguistics, PO Box 450 Stn A, Ottawa, Ontario, Canada  
gsm@alumni.sfu.ca

## ABSTRACT

Native Spanish speakers and Canadian English learners of Spanish produced the five Spanish vowels in utterance final /sV/ syllables. The syllables were presented in random order to Spanish and English listeners for identification of vowels in terms of Spanish vowel categories. The English listeners also identified the vowels in terms of English vowel categories. Initial analysis of perceptual results was consistent with English listeners identifying Spanish vowels via assimilation to English vowels, and in the case of Spanish /a/ and /u/ via *multiple-category assimilation* to more than one English vowel. Spanish listeners correctly identified English speakers' vowels at higher rates than Spanish speakers' vowels. This is consistent with English speakers basing their Spanish vowel productions on English vowels, resulting in less intra-vowel variation. Spanish listeners' results were also consistent with part of the vowel space being uncommitted to any Spanish vowel.

## 1 INTRODUCTION

Much research in cross-language speech perception and production has focussed on learners of an L2 with a relatively large number of vowels compared to their L1. In contrast, few studies have focussed on learners of an L2 with a relatively small number of vowels compared to their L1. In an L2 with few vowels, each individual vowel might be expected to cover a larger proportion of the perceptual vowel space than covered by individual vowels in an L1 with many vowels. The vowel space covered by a single L2 vowel may overlap with several L1 vowel categories, and L2 learners may assimilate instances of the L2 vowel to multiple L1 vowels. Adapting the terminology of Best's Perceptual Assimilation Model [1], Escudero & Boersma [3] called this *multiple-category assimilation* (MCA). The learners may initially construct a representation of the L2 vowel with the same external boundaries as the several L1 sounds to which it was assimilated. Using the terminology of Flege's Speech Learning Model (SLM) [4] this might be characterised as a diaphone category consisting of one L2 vowel and several L1 vowels. The situation may be further complicated if the perceptual space covered by one of the L1 vowels contributing to the MCA diaphone also partially overlaps with a second L2 vowel. An MCA-diaphone category may develop the distributional characteristics of the complete set of L1 and L2 vowels, à la SLM [4]. Alternatively, if learners are able to switch between L1 and L2 modes of perception, then the diaphone may simply serve as a bootstrap in the initial stages of learning, after which the distributional

properties of the L2 vowel would gradually come to define the category. Escudero & Boersma [3] argued for the latter. They presented Dutch learners of Spanish with Spanish /i/ and /e/ embedded in both Spanish and Dutch carrier sentences, and asked them to identify the vowels in terms of Spanish /i/ and /e/, and Dutch /i/, /ɪ/ and /e/. Inexperienced learners had a Spanish /i/-/e/ boundary in approximately the same location as their Dutch /ɪ/-/e/ boundary, but for progressively more experienced learners the boundary approximated and reached the native Spanish listeners' /i/-/e/ boundary.

If the majority of instances of the L2 vowel in an MCA diaphone are perceived as more similar to one of the L1 vowels, then in production the learners may initially substitute this L1 vowel for the L2 vowel. Since the L1 vowel covers a smaller portion of the perceptual vowel space, in production it might be expected to have formant values with a tighter distribution than the L2 vowel.<sup>1</sup> The learners' tightly-clustered vowel may fall completely within the native listeners' perceptual space for that vowel and thus be correctly identified, although it may be off centre and thus sound accented. As learners gain in experience with the L2, the distribution of the production category would be expected to approximate the perceptual category.

The current paper presents an initial analysis of perception data from a pilot study investigating the perception and production of Spanish vowels by Canadian-English learners of Spanish. It addresses the questions of whether inexperienced Canadian-English learners of Spanish use MCA diaphones in their perception of Spanish vowels, and whether they produce Spanish vowels with tighter distributions than native Spanish speakers' vowels.

Rochet [6] argued that perceptual vowel category boundaries extend to be contiguous with adjacent categories leaving no uncommitted space. The current paper also addresses the question of whether Spanish has uncommitted vowel space.

## 2 METHODOLOGY

Spanish and English speakers read sentences in which the five Spanish vowels appeared utterance-finally following /s/. Each sentence was written so as to be syntactically and pragmatically well formed in Spanish. The sentences were:

Ha dicho que sí.	/i/
No los sé.	/e/
Dime una cosa.	/a/
¿Qué es eso?	/o/
La forma proclítica es "su".	/u/

The vowels /i/, /e/, and /u/ occurred in stressed position, and /a/ and /o/ were in post-stressed position. Each sentence occurred three times in random order on a written list from which the speakers read. The speakers were recorded in a soundproofed room using a Sony MZS-R5ST Mini Disc recorder and a Sony ECM-MS907 microphone. Recordings were transferred to computer via a Roland ED UA-30 USB Audio Interface and saved as 22.05 kHz 16bit files.

The 60 /sV/ syllables containing the target vowels were extracted from the recorded sentences and presented to Spanish and English listeners in random order via MEDS software [5]. Listeners responded by clicking on one of six boxes on the screen. Five boxes contained the orthographic letters, “A E I O U” representing the five Spanish vowel categories. Listeners were instructed to indicate the Spanish vowel that was closest to the vowel they heard. The sixth box contained the word “ninguna” [none] (henceforth *N*) and listeners were instructed to click on this box only if the vowel did not sound like any Spanish vowel. Listeners were told to use a vowel letter response instead of *N* if the token sounded even a little like a Spanish vowel. In a separate test using the same procedure, English participants identified the same tokens in terms of English vowel categories. Vowel responses were identified by the keywords “heat, hit, hate, het, hat, hot, hut, hotel, hood, none” representing the 10 English vowels /i ɪ e ε æ ɒ ʌ o u/ plus “not like any English vowel” respectively.

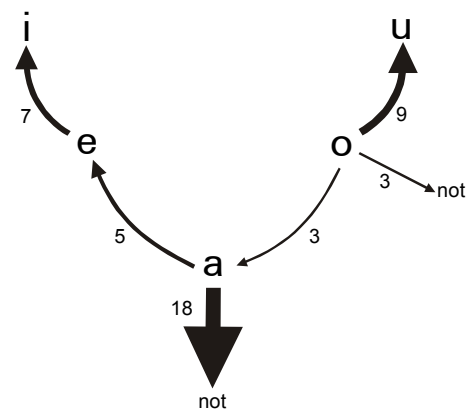
Spanish speakers in the production experiment were eight women, two from Mexico, and one each from Argentina, Chile, Colombia, Cuba, Spain, and Venezuela. Spanish listeners in the perception experiment were three women who had also participated as speakers, the Argentinian, Columbian, and Spaniard, and a man from Spain. Canadian English speakers were 12 women who had studied Spanish for between 0.5 and 5 years (mean 2.1). Listeners were seven women, who had also participated as speakers, and two men. They had studied Spanish for between 0.5 and 4 years (mean 1.8). None of the Canadian English participants had lived in a Spanish speaking country, and were therefore classified as relatively inexperienced learners of Spanish. All of the participants lived in the greater Vancouver area at the time of the experiments.

### 3 RESULTS AND DISCUSSION

**3.1 SPANISH SPEAKERS - SPANISH LISTENERS**  
Spanish listeners’ misidentifications of Spanish speakers’ vowels are shown in Figure 1. (The term “misidentified” is used to indicate that the vowel identified by the listeners was not the same as the vowel intended by the speakers. “Correctly identified” will be used to indicate that the vowel identified by the listeners was the same as that intended by the speakers.) The gross trend was for low and mid vowels to be misidentified as higher vowels.

In terms of individual vowels, the most prominent result relates to /a/ which was identified as *N* at a rate of 18%. This may suggest that there is an uncommitted area in the Spanish

listeners vowel space. Preliminary acoustic analysis suggests that /a/ was relatively high in the vowel space and not significantly lower than /e/. The uncommitted area is therefore likely in the centre of the vowel space. The /a/ tokens were also extremely short, some having only one or two glottal pulses. The Spanish listeners may have genuinely perceived the short vowels as not corresponding to any Spanish category, or may simply have chosen the *N* response because the shortness of the vowels made them difficult to identify. Whilst this result does support the existence of an uncommitted space, it is almost certainly an artifact of the test: the vowels were presented in an isolated syllable which listeners would expect to be stressed, but the /a/ vowels were extracted from unstressed syllables.



**Figure 1:** Misidentification of Spanish speakers’ vowels by Spanish listeners. The Arrows are proportional in width to the number of misidentifications and are labelled as percentages of the total number of responses for each vowel. Only misidentifications with a rate of > 2.5% are shown. The label “not” refers to *N* responses.

**3.2 SPANISH SPEAKERS - ENGLISH LISTENERS**  
English listeners’ misidentifications of Spanish speakers’ vowels in terms of Spanish categories are shown in Figure 2, and their identification of Spanish speakers’ vowels in terms of English categories are shown in Figure 3.

English listeners gave some Spanish /e/ identifications to Spanish speakers’ /i/ whereas Spanish listeners had 100% correct identification for /i/. This suggests that the English listeners’ boundary between Spanish /e/ and /i/ was higher than that of Spanish listeners. This might be expected if English listeners assimilated Spanish /e/ to higher English /e/. This assimilation pattern is also suggested by the fact that Spanish speakers’ /e/ were identified as English /e/ at a rate of 81%.

Although English listeners identified Spanish speakers’ /i/ as Spanish /i/ at a rate of 90% and as Spanish /e/ at a rate of 10%, they were identified as English /i/ at a rate of 100% with no English /e/ responses. In addition, a greater percentage of Spanish speakers’ /e/ (13%) were identified as English /i/ than were identified as Spanish /i/ (6%). This suggests that the English listeners’ English /e/-/i/ boundary was lower than their Spanish /e/-/i/ boundary. To summarise, English listeners’ had a higher English /e/-/i/

boundary than the Spanish listeners' Spanish /e-/i/ boundary, but the English listeners' Spanish /e-/i/ boundary was higher again. This result could not be predicted by simple transfer of the English boundary to the English listeners' Spanish perception, nor to Spanish learning, in which case the English listeners' boundary would be expected to converge with the Spanish listeners' boundary. It may be that in terms of initial spectral values the English listeners had the same /e-/i/ boundary for both English and Spanish but responded differently due to differences in diphthongisation: Spanish /e/ is essentially a monophthong whereas Canadian English /e/ is realised as a rising diphthong [ej]. When presented with a non-diphthongised vowel near their /e-/i/ boundary the English listeners may have identified it as Spanish /e/, but the same vowel may have been identified as non-diphthongised English /i/ rather than as diphthongised English /e/. This would require the listeners to be able to switch between English and Spanish modes of perception.

Whereas Spanish listeners misidentified Spanish speakers' /a/ as /e/ at a rate of 5%, English listeners showed the reverse in misidentifying /e/ as /a/ at a rate of 5%. This suggests that the English listeners' boundary between /a/ and /e/ was higher than that of Spanish listeners. The English listeners may have assimilated Spanish /a/ to the more fronted and higher English /æ/, or may have assimilated Spanish /e/ to the higher English /e/. Results for vowel identification in terms of English categories are consistent with both these assimilations: 81% of Spanish speakers' Spanish /e/ were identified as English /e/, and only 5% as English /æ/; and 70% of Spanish speakers' Spanish /a/ were identified as English /æ/. Assimilation of Spanish /e/ to English /e/ rather than /æ/ may in part have been due to the prohibition in English on lax vowels in final open syllables.

The English listeners misidentified Spanish /o/ as Spanish /a/ at a greater rate than had the Spanish listeners (13 versus 3%). This suggests that the English listeners' boundary between /a/ and /o/ was higher than that of the Spanish listeners. This might be expected if English listeners had assimilated Spanish /a/ to English /ʌ/ or /ɒ/. Spanish /a/ were identified as English /ɒ/ at a rate of only 5%, but were identified as English /ʌ/ at a relatively high rate of 14%. It therefore seems more likely that the English listeners' Spanish /a/ category was based primarily on MCA to English /æ/ and /ʌ/. The rate of identification of Spanish /o/ with Spanish /a/ was similar to the rate of identification of Spanish /o/ with English vowels which may form the English listeners' MCA diaphone for Spanish /a/ (12.9% compared to 10.8%, the sum of percentages of /ɒ/, /æ/, /ʌ/, and /ɛ/). This identification pattern may have been mediated through English schwa. As mentioned above, the Spanish speakers produced utterance final non-stressed /a/ that were short and centralised. Whereas Spanish listeners had a tendency to reject these vowels, English listeners likely heard them as schwas (the majority as /æ/-like schwas).

Spanish speakers' /u/ were identified as Spanish /u/ at a rate of 100%, and as English /u/ at 91% and English /ʊ/ at 8%.

This suggests an MCA of Spanish /u/ to English /u/ and /ʊ/.

In summary, the English listeners' perception of Spanish speakers' Spanish vowels was consistent with Spanish /i/, /e/, and /o/ perceptual categories based on assimilation to English /i/, /e/, and /o/ respectively, and Spanish /a/ and /u/ perceptual categories based on MCA to English /æ/+/ʌ/+/ɒ/+/ɛ/, and /ʊ/+/u/ respectively.

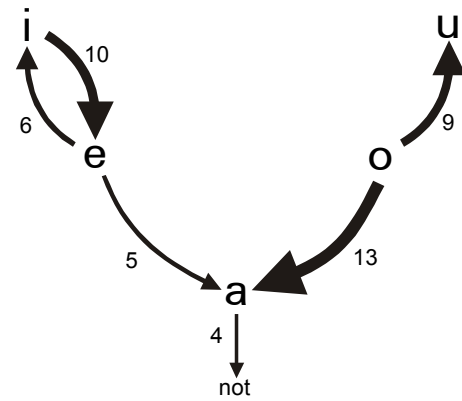


Figure 2: Misidentification of Spanish speakers' vowels by English listeners.

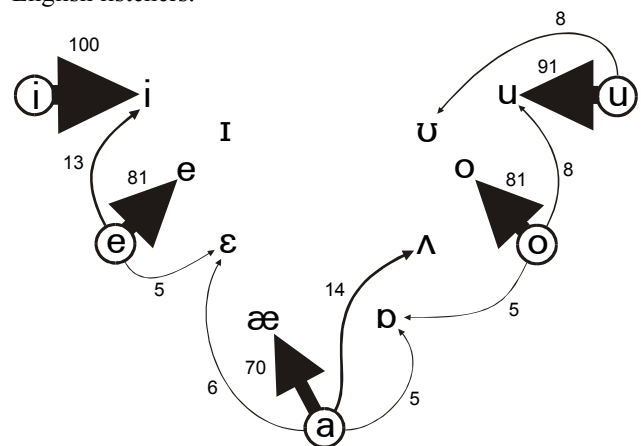


Figure 3: English listeners' identification of Spanish speakers' vowels in terms of English vowel categories. Symbols for the Spanish speakers' vowels are enclosed in circles. The Arrows are proportional in width to the number of identifications and are labelled as percentages of the total number of responses for each vowel. Only identifications with a rate of > 5% are shown.

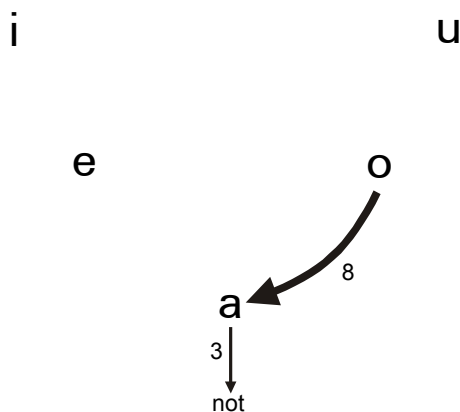
### 3.3 ENGLISH SPEAKERS - SPANISH LISTENERS

Spanish listeners' misidentifications of English speakers' Spanish vowels are shown in Figure 4. The Spanish listeners correctly identified the English speakers' vowels at higher rates than the Spanish speakers' vowels (overall 96% compared to 91%).<sup>2</sup> These results suggest that the English speakers' Spanish vowel productions varied less than those of the Spanish speakers, and are compatible with the hypothesis in the introduction that, in production, English listeners would initially substitute a single English vowel for each Spanish vowel. English production substitutes for Spanish vowels would be the English vowels to which the Spanish vowels were primarily assimilated in perception.

From Figure 3, these substitutes would be English /i/, /e/, /æ/, /o/, and /u/ for Spanish /i/, /e/, /a/, /o/, and /u/ respectively.

The higher correct identification rate for the English speakers' Spanish /a/ compared to the Spanish speakers' /a/ can also be accounted for by the fact that the English speakers produced /a/ which were longer and lower in the vowel space than those of the Spanish speakers, i.e., the English speakers exhibited less vowel reduction than the Spanish speakers.

The English speakers' Spanish /o/ productions were identified as /a/ at a rate of 8%. This would be at odds with the substitution of higher English /o/ for Spanish /o/. However, since English listeners also identified 8% of English speakers' /o/ as English /æ/, this is likely the result of vowel reduction leading to schwa realisations of /o/.



**Figure 4:** Misidentification of English speakers' Spanish vowels by Spanish listeners.

#### 4 CONCLUSIONS

The results were consistent with the following statements which are offered as tentative conclusions pending further analysis and research:

- English listeners perceived Spanish /i/, /e/, and /o/ via assimilation to English /i/, /e/, and /o/ respectively, and perceived Spanish /a/ and /u/ via *multiple-category assimilation* to English /æ/+/ʌ/+/ɒ/+/ɛ/, and /ʊ/+/u/ respectively.

- In production, English speakers substituted English /i/, /e/, /æ/, /o/, and /u/ for Spanish /i/, /e/, /a/, /o/, and /u/ respectively. These were the English vowel categories to which the Spanish vowels were primarily assimilated in perception.

- The centre of the perceptual vowel space is uncommitted to any Spanish vowel.

Future analysis will include statistical modelling to determine relative locations of perceptual boundaries for Spanish and English listeners, and comparison with production data. Future research will be longitudinal and will systematically investigate the effect of position in word

and stress on vowel production and perception. English and Spanish carrier sentences will be used to incline listeners towards Spanish and English modes of perception.

#### NOTES

1. A tighter distribution would be required to avoid excessive overlap with other vowels in a more crowded vowel space. It should be noted, however, that Bradlow [2] failed to find evidence that English speakers produced more tightly clustered vowels than did Spanish speakers.

2. The greater number of arrows for Spanish speakers in Figure 1 compared to English speakers in Figure 4 are not due to sample size differences. Spanish speakers would reach the 2.5% criterion for display with a smaller absolute number of responses because the total number of responses is smaller; however, if the display threshold were set at an absolute value of 4 the same arrows would be displayed.

#### REFERENCES

- [1] C. T. Best, "A direct realist view of cross-language speech perception," in *Speech perception and linguistic experience: Issues in cross-language research*, W. Strange, Ed., pp. 171-204. Baltimore: York Press, 1995.
- [2] A. R. Bradlow, *Language-specific and universal aspects of vowel production and perception: A cross-language study of vowel inventories*, Ithaca NY: DMLL, Cornell University, 1993.
- [3] P. Escudero and P. Boersma, "The subset problem in L2 perceptual development: Multiple-category assimilation by Dutch learners of Spanish," in *Proceedings of the 26th Annual Boston University Conference on Language Development*, B. Skarabela, S. Fish, and A. H.-J. Do, Eds., pp. 208-219. Somerville MA: Cascadilla, 2002.
- [4] J. E. Flege, "Second language speech learning theory, findings, and problems," in *Speech perception and linguistic experience: Issues in cross-language research*, W. Strange, Ed., pp. 233-277. Baltimore: York Press, 1995.
- [5] R. A. Kendall, *Music experiment development system: MEDS (version 2001-B)* [Computer software]. Los Angeles, 2001.
- [6] B. L. Rochet, "Perception and production of second-language speech sounds by adults," in *Speech perception and linguistic experience: Issues in cross-language research*, W. Strange, Ed., pp. 379-410. Baltimore: York Press, 1995.