Working Papers of the Cornell Phonetics Laboratory
1993, vol. 8, pp. 29-85.

A cross-language comparison of vowel production and perception: language-specific and universal aspects*

Ann R. Bradlow

1. Introduction

This paper presents a concurrent examination of the acoustic and perceptual characteristics of the vowels of two languages with very different vowel systems, namely English with a relatively large inventory of eleven stressed monophthongs and Spanish with a relatively small inventory of five stressed monophthongs. This investigation attempts to identify language-specific and universal aspects of both vowel production and perception by exploring the effect of inventory size on the acoustic and perceptual characteristics of similar vowel categories across different vowel systems. Specifically, this study investigates the acoustic characteristics of the four common vowels across the two languages, namely /i/, /e/, /o/, and /u/, as well as the perceptual characteristics of the /u/-/o/ contrast in the two languages. By looking at the phonetics of vowels that occupy similar positions in the vowel spaces of the two languages the effects of the different vowel inventory structures are most noticeable, and thus provide an indication of the language-specific and universal aspects of vowel production and perception.

English and Spanish were chosen for this study because of the large difference between the sizes of their vowel inventories: English has more than double the number of stressed monophthongal vowels than Spanish. Furthermore, the five-vowel system of Spanish is statistically very common, whereas the eleven-vowel system of English is uncommonly large (Maddieson, 1984). Despite these differences regarding inventory size, the vowel systems of these two languages are similar in that they vary along the same dimensions (neither language has contrastive rounding, length, or nasalization). In other words, the principle difference between these vowel systems is in the number of vowel categories. Thus, this English-Spanish comparison represents a comparison of an unusually large vowel inventory with a smaller cross-linguistically common vowel inventory. The expectation is that this difference between the two vowel inventories will highlight both the differences and similarities attributable to language-specific and/or universal aspects of vowel production and perception. Charts for the monophthongal vowels of each of these

---

* The work presented in this paper is based on my thesis research. I thank Allard Jongman, Abigail Cohn, John Kingston, Carol Krumhansl, and Joan Sereno for many helpful comments.

languages are given in Figure 1 below. This display follows the traditional analyses of these vowels systems as having three phonological heights, and three phonological divisions along the front-back dimension.

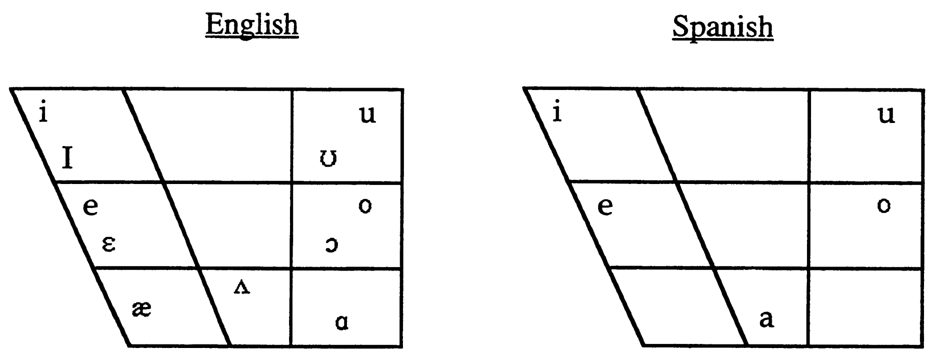

Figure 1. Charts of the English and Spanish monophthongs

The remainder of this paper is structured as follows. Section two presents an acoustic comparison of the English and Spanish vowel spaces. Section three presents a comparison of the acoustic and perceptual characteristics of the /u/-/o/ contrast within each of the two languages. Section four presents a cross-language perceptual comparison of the /u/-/o/ contrast in English and Spanish. Finally, section five presents the general conclusions and some discussion of the results.

2. Acoustic comparison of the English and Spanish vowel spaces

2.1. Introduction

This section explores the effect of inventory size on the acoustic realization of vowels in General American English and Madrid Spanish. First, I compare the locations of the common vowel categories of English and Spanish in the acoustic space. Through this comparison I determine if there is any systematic, language-specific effect that plays a role in determining the arrangement of vowels in the acoustic space. Second, the areas in the acoustic space covered by the English and Spanish vowels are compared, in order to see whether the English acoustic vowel space is "expanded" relative to the Spanish acoustic vowel space. This kind of "expansion" effect would be consistent with a general principle of dispersion that assumes cross-linguistic variation regarding the boundaries of the general acoustic vowel space . Third, this acoustic comparison focuses on the tightness of within-category clustering in English and Spanish, in order to determine whether the larger inventory of English will tend towards tighter within-category clustering as a means of minimizing the potential for confusion between neighboring categories. Finally, the acoustic data of the present study are compared with other published acoustic data on the vowels of English and Spanish.

2.2. Method

Four male speakers of General American English and four male speakers of Madrid Spanish read words exemplifying the vowel contrasts of their respective languages. All of the English speakers have spent most of their lives in the Ithaca, New York area. All of the Spanish speakers come from Madrid: two of these have spent significant amounts of time in the U.S., the other two have spent almost all of their lives in the Madrid area. The target vowels all occurred between either /p/ or /b/, and /t/. The word lists and frame sentences are given below in Table 1.

Table 1. English and Spanish word lists and frame sentences.

English word list:		Spanish word list:		
beat	/bit/	bita	/bita/	'bit, nautical term'
bit	/bIt/	beta	/beta/	'beta'
bait	/bet/	bata	/bata/	'bath robe'
bet	/bɛt/	bota	/bota/	'boot'
bat	/bæt/	puta	/puta/	'prostitute'
but	/bʌt/			
boot	/but/			
put	/pʊt/			
boat	/bot/			
bought	/bɔt/			
pot	/pɑt/			

English frame sentence:	Spanish frame sentence:
Say _____ again.	Escribe____bien.
	'Write____well.'

As shown in Table 1, the English words were all monosyllabic, and in accordance with Spanish phonotactics and syllabification, the Spanish words were disyllabic. The words were embedded in frame sentences and each sentence was repeated five times in random order, giving a total of 20 tokens (four speakers x five repetitions) for each vowel category of each language. All recordings were made with a Marantz portable cassette recorder PMD222 and an AKG D310 microphone. The English recordings were made in a sound-attenuated booth in the Phonetics Laboratory at Cornell University. The Spanish recordings were made in a quiet room in Madrid. The recordings were digitized with a sampling rate of 12000 Hz and low-pass filtered at 6000 Hz. All measurements were made using the Entropics WAVES+ speech analysis software on a SUN workstation. Both LPC spectra and spectrograms were used to determine the formant frequencies. The LPC spectra were calculated from a 25 ms Hanning window in the vowel steady-state. In most cases an analysis order of 14 was used; however, in a small number of cases the analysis order was lowered to 12 so that the data were smoothed to yield a sharper peak. The steady-state portion of the vowel was determined by first placing two vertical cursors in the spectrogram: the first cursor marked the end of the formant transitions coming out of the initial consonant, and the second cursor marked the beginning of the formant transitions into the final consonant. Figure 2 gives a sample spectrogram of an English /bit/ token. The two vertical lines show the cursors that mark the end of the vowel onset and the start of the vowel offset, respectively. The period between the two cursors thus indicated the portion of the vowel which showed no (or very little) formant movement.

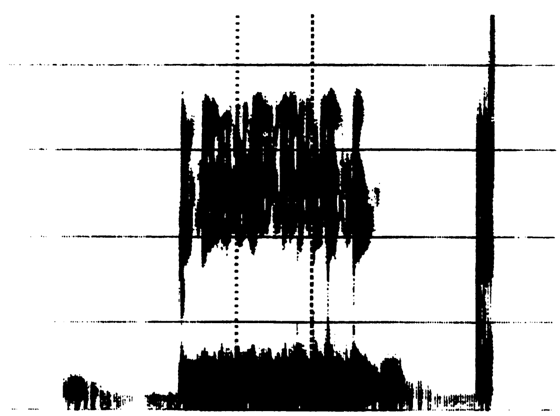

Figure 2. Sample spectrogram of an English /bit/ token. The vertical lines mark the end of the vowel onset and the start of the vowel offset, respectively. Each horizontal line represents 1000 Hz.

A 25 ms Hanning window was placed approximately in the middle of this steady-state period. The formant values were then read from the LPC spectrum and checked with readings from the spectrogram. In the cases of English /e/ and /o/, which are sometimes diphthongized, the formant measurements were taken from the portion of the vowel before the offglide. The auditory quality of this portion of the vowel was checked, in order to insure that the formant measurements being recorded were those appropriate for /e/ or /o/ rather than for the /j/ or /w/ offglides.

2.3. Results

Figure 3 shows results of the acoustic analysis of the data from the four male English speakers. In this plot, as in subsequent plots, F1 is shown on the horizontal axis, and F2 is shown on the vertical axis. Both axes are represented by a linear frequency scale. The points in the plot represent the mean F1 and F2 measurements for the 20 tokens of each vowel. The bars indicate one standard deviation from the mean in each direction.

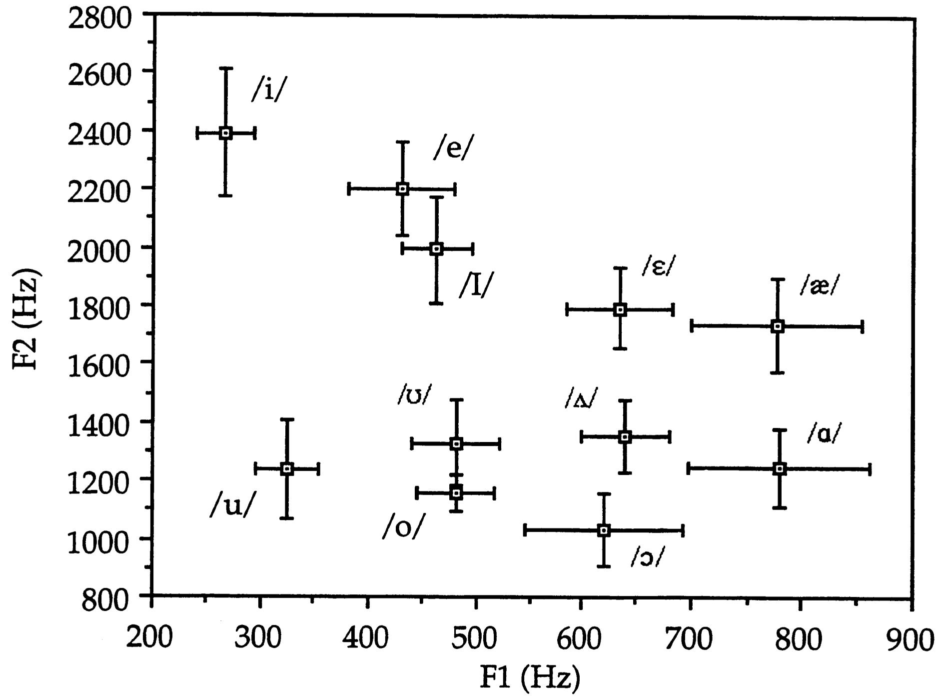

Figure 3. English acoustic data. Each point represents the mean F1 and F2 values of 20 vowel tokens. The bars represent one standard deviation from the mean in each direction.

Table 2 gives the pooled mean F1 and F2 values with standard deviations for the English vowels.

Table 2. English CVC mean F1 and F2 in Hz with standard deviations (4 male speakers, 5 repetitions of each vowel).

vowel	F1 (sd)	F2 (sd)
i	268 (26)	2393 (220)
I	463 (32)	1995 (184)
e	430 (49)	2200 (159)
ε	635 (50)	1796 (134)
æ	777 (77)	1738 (159)
ʌ	640 (41)	1354 (129)
ɑ	780 (81)	1244 (133)
ɔ	620 (74)	1033 (126)
o	482 (36)	1160 (63)
ʊ	481 (41)	1331 (151)
u	326 (29)	1238 (171)

Figure 4 is an F1 by F2 plot of the Spanish vowels on a linear frequency scale. The means and standard deviations for the Spanish data are given in Table 3. In this case, as for the English data, the points in the figure and the values in the table represent the mean F1 and F2 for the 20 tokens of each vowel category.

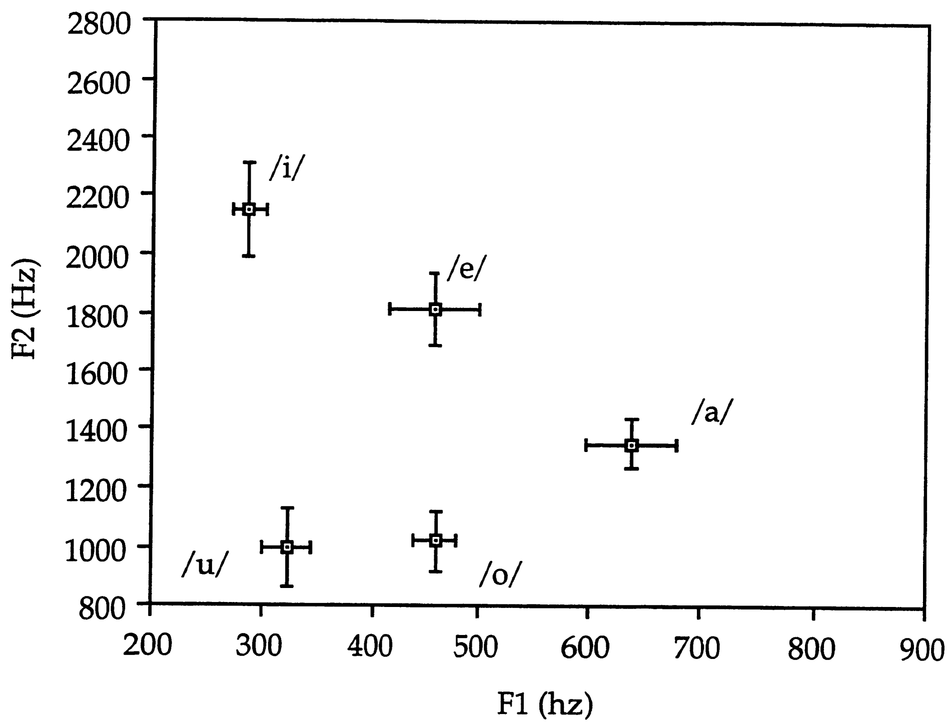

Figure 4. Spanish acoustic data. Each point represents the mean F1 and F2 values of 20 vowel tokens. The bars represent one standard deviation from the mean in each direction.

Table 3. Spanish mean vowel formants in Hz with standard deviations (4 male speakers, 5 repetitions of each vowel).

vowel	F1 (sd)	F2 (sd)
i	286 (15)	2147 (160)
e	458 (42)	1814 (121)
a	638 (40)	1353 (85)
o	460 (20)	1019 (101)
u	322 (22)	992 (134)

2.4. Discussion

2.4.1. Base-of-articulation

The first issue addressed by this acoustic study concerns the general placement of English and Spanish vowels in the acoustic vowel space. In this regard we might expect both a language-specific effect, which will cause similar vowel categories across two languages to differ in a systematic way due to a consistent language-specific adjustment of the articulators (that is, a base-of-articulation effect), as well as a general expansion effect for languages with crowded inventories relative to languages with uncrowded inventories. These two possible effects on vowel location would necessarily interact with each other; thus, it is important to examine each individually in order to assess the extent of the two separate effects. In this section, then, I focus on assessing the presence (and extent) of any language-specific base-of-articulation effect, which may result in a general shift in some direction of all the vowels of one language relative to the equivalent vowels of the other language.

As shown in Figure 5, a comparison of the locations of the common vowels across English and Spanish, namely /i/, /e/, /o/, and /u/, indicates a general upward shift in the F2 dimension of the English vowels relative to the Spanish vowels. The low vowel is excluded from the set of common vowels because for these Spanish speakers, the low vowel is central in the front-back dimension (as represented by the IPA symbol /a/); whereas, for these English speakers, there is a low front vowel, (IPA /æ/) and a low back vowel, (IPA /ɑ/), but no low central vowel.

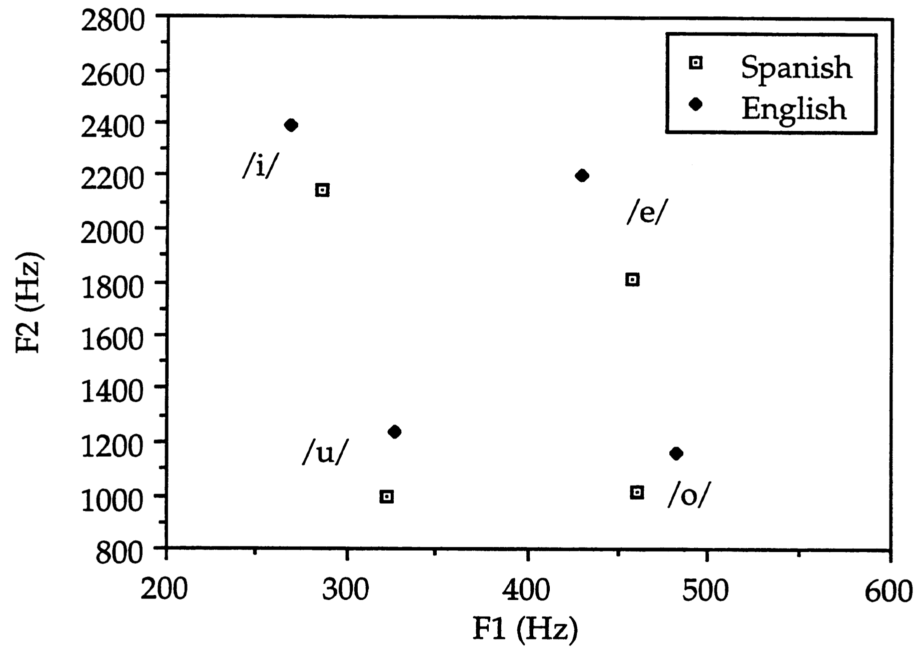

Figure 5. A comparison of the locations of /i/, /e/, /o/, and /u/ in English and Spanish.

As shown in Table 4, the results of two-tailed, unpaired t-tests on the 20 individual tokens for each of these vowels in English and in Spanish show that the English versions are all significantly higher in the F2 dimension than the corresponding Spanish versions.

Table 4. Results of a two-tailed, unpaired t-test on the mean F2 values of the shared vowels of English and Spanish.

	English	Spanish	mean diff.	t(18)	P-value
i	2393	2147	246	4.051	.0002
e	2200	1814	386	8.655	<.0001
o	1160	1019	141	5.306	<.0001
u	1238	992	246	5.052	<.0001

The effect of this difference in F2 between the English and Spanish vowels is that the English front vowels, namely /i/ and /e/, are located more peripherally in the acoustic space than the Spanish front vowels. This effect on the front vowels would thus seem to concur with the result we might expect from an expansion effect. However, we also find that the English back vowels, namely /u/ and /o/, are raised in the F2 dimension relative to the Spanish back vowels, resulting in a more central location for the English back vowels. This result for the back vowels is in contradiction to the result we might expect from an expansion effect, since such an effect should result in English back vowels with lower F2 frequencies than their Spanish counterparts. However, the consistent upward shift in the F2 dimension of the English vowels relative to the Spanish vowels is in accordance with the result we would expect from a language-specific base-of-articulation effect. Specifically, this result suggests that the vowels of English are articulated with a fronted tongue position relative to the tongue position for the Spanish vowels.

An alternative explanation for the higher F2 of the English vowels relative to the Spanish vowels, is that the four individual English speakers in this study all have shorter vocal tracts than the four individual Spanish speakers, resulting in generally higher formants for the English speakers than for the Spanish speakers. In such a situation, the present data would also show a systematic raising of the F1 frequencies for the English speakers relative to the F1 frequencies for the Spanish speakers. However, if the present data do not show a similar effect for F1 as for F2, then we can rule out the possibility that the difference in F2 for the two groups of speakers is a function of vocal tract length. Accordingly, a comparison of the F1 frequencies for the English and Spanish vowels was performed.

The results of the comparison of the F1 frequencies across English and Spanish indicate that for the front vowels, English /i/ has a lower F1 value than Spanish /i/, and that English /e/ tends towards a lower F1 than Spanish /e/. For the back vowels, English /o/ has a significantly higher F1 than Spanish /o/, but English /u/ and Spanish /u/ do not show any difference in F1. These results are summarized in Table 5 below.

Table 5. Results of a two-tailed, unpaired t-test on the mean F1 values of the shared vowels of English and Spanish.

	English	Spanish	mean diff.	t(18)	P-value
i	268	286	-19	-2.771	.0086
e	430	458	-28	-1.941	.0597
o	482	460	22	2.378	.0225
u	326	322	4	.440	.6628

Thus, in the F1 dimension we do not find a consistent effect as we do in the F2 dimension across all four common vowels. Furthermore, since these data do not indicate a systematic raising of all the English formants relative to the Spanish formants, an effect of different vocal tract lengths between the two groups of speakers can be ruled out. We do, however, find a consistent trend towards a lowering in the F1 dimension of the English front vowels, /i/ and /e/, relative to the phonemically equivalent Spanish front vowels. In the case of the back vowels, the trend in this F1 comparison is more consistent with an expansion effect than with a base-of-articulation effect: even though there is no F1 difference between English and Spanish /u/, the English mid back vowel, /o/, is more peripheral, that is, it has a significantly higher F1, than the Spanish /o/.

In the comparison of the vowel category locations for the four vowels common to English and Spanish thus far, we have seen that the English vowels are all significantly higher in the F2 dimension, and that, in the F1 dimension, there is no such consistent language-specific effect on vowel location, although there is a general trend for the English front vowels to be lower in the F1 dimension than their Spanish counterparts.

2.4.2. English CVCV tokens

Recall that the English vowel tokens all occurred in monosyllabic words whereas the Spanish tokens all occurred in disyllabic words. It is thus possible that the consistent shift in F2 of the English vowels relative to the Spanish vowels is due to a more general effect, such as the structure of the test words, rather than due to a language-specific base-of-articulation property. For instance, it is possible that the tautosyllabic coronal, with its characteristically high F2 locus, in the English test words raised the F2 of the target vowel, especially in the case of the back vowels which are generally characterized by low second formants. Therefore, in order to rule out an effect of word-structure, a new set of English recordings was made using disyllabic non-words that matched the syllabic structure of the Spanish test words. This new set of data thus allowed for a direct comparison of the tokens with the same syllabic structure as produced by speakers of the two languages.

Three of the four male speakers of General American English from the first experiment produced disyllabic non-word versions of the tokens used in the original data set. This set of words is shown in Table 6 below.

Table 6. English disyllabic non-words, with a sample from the reading list.

/bita/	Reading list sample:
/bIta/	1. beat
/beta/	Say bea.ta again.
/bɛta/	2. put
/bæta/	Say pu.ta again.
/bʌta/	
/pɑta/	
/bɔta/	
/bota/	
/pʊta/	
/buta/	

These test tokens were embedded in the same frame sentence as used in the first experiment; and the speakers read each sentence five times in random order. The speakers read these sentences from a list that was constructed as follows: the word from the original list (see Table 2) that contained the target vowel appeared on one line, and the frame sentence with the target non-word appeared on the following line. In order to reinforce the disyllabicity of the test tokens, and to avoid the production of a flap for the medial /t/, the target non-word was typed with a period separating the two syllables (e.g. bea.ta), and the speakers were instructed that the period indicated a syllable boundary and to avoid producing a flap for the intervocalic /t/. This procedure was effective in eliciting tokens that matched their Spanish counterparts at both the segmental and intonational levels. A sample of the list that the speakers read from is included in Table 6 above.

This set of data was collected and analyzed identically to the original set of English data. Figure 6 gives an F1 by F2 plot of the mean English vowels from the CVCV tokens, along with the mean vowels from the CVC tokens as produced by the same three individuals.

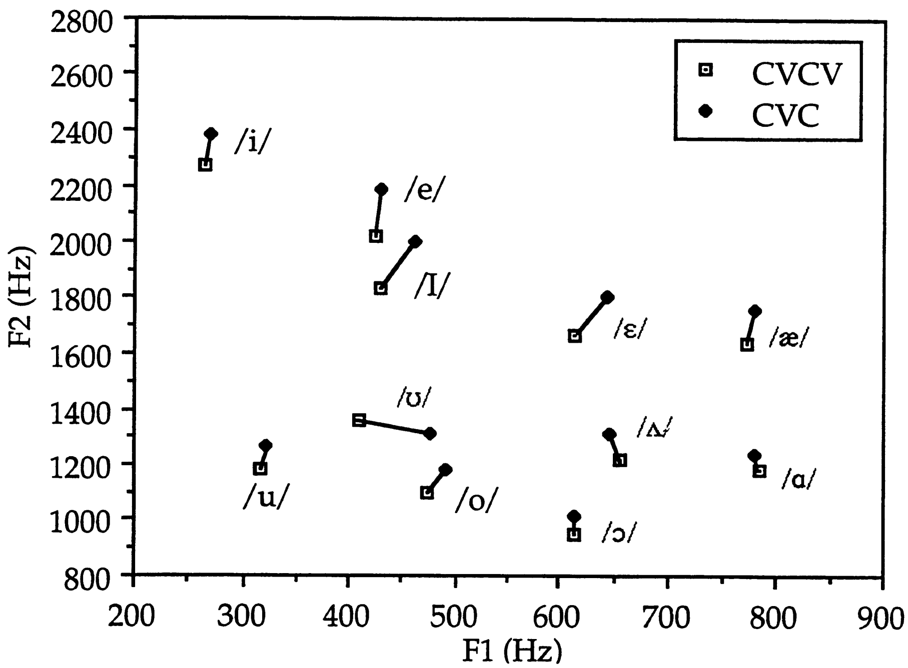

Figure 6. Comparison of the English CVCV and CVC acoustic data. Each point represents the mean F1 and F2 value of 15 tokens.

Table 7 gives the mean F1 and F2 values and standard deviations for the English di- and monosyllabic tokens shown in Figure 5. The data from the fourth speaker in the CVC condition was omitted from this analysis so that the comparison of these two data sets is constrained to a comparison of the data from the same three individuals under each condition.

Table 7. Mean F1 and F2 frequencies for the English mono- and disyllabic tokens (3 male speakers, 5 repetitions in each condition).

	CVCV		CVC	
vowel	F1 (sd)	F2 (sd)	F1 (sd)	F2 (sd)
i	264 (30)	2268 (179)	270 (4)	2380 (254)
I	429 (44)	1831 (148)	461 (36)	2002 (212)
e	424 (38)	2020 (132)	429 (54)	2185 (176)
ɛ	615 (56)	1665 (129)	644 (53)	1806 (154)
æ	773 (70)	1640 (154)	779 (87)	1759 (180)
^	655 (55)	1216 (70)	645 (44)	1309 (117)
ɑ	783 (150)	1182 (131)	780 (91)	1236 (153)
ɔ	614 (74)	945 (83)	613 (80)	1018 (142)
o	473 (23)	1094 (65)	491 (37)	1178 (61)
ʊ	411 (41)	1361 (92)	475 (45)	1314 (169)
u	316 (38)	1183 (140)	322 (28)	1267 (182)

Table 8 shows the results of a two-tailed paired comparison of the F1 and F2 measurements from the three English speakers that participated in both the di- and monosyllabic experiments.

Table 8. Comparison of the F1 and F2 frequencies in the CVCV and CVC contexts.

	F1			F2		
vowel	mn. diff	t (14)	P-value	mn. diff	t (14)	P-value
i	-6.2	-.801	.4367	-112.333	-3.894	.0016
I	-31.867	-2.384	.0319	-171.067	-7.958	<.0001
e	-4.733	-.418	.6825	-164.667	-7.212	<.0001
ɛ	-28.333	-3.113	.0076	-141.533	-8.331	<.0001
æ	-6.400	-.394	.6996	-118.533	-6.171	<.0001
ʌ	10.133	.751	.4653	-92.867	-4.007	.0013
ɑ	3.400	.151	.8825	-53.667	-2.531	.0240
ɔ	.933	.049	.9618	-72.667	-2.834	.0133
o	-18.133	-1.470	.1638	-83.400	-3.292	.0053
ʊ	-64.200	-3.461	.0038	46.933	1.439	.1721
u	-5.600	-.668	.5153	-83.933	-2.256	.0406

The comparison of the two English data sets shows that in general, the locations of the target vowels in the acoustic space are affected by the syllabic structure of the word in which they appear. Specifically, with the exception of one vowel, namely /Ü/, the F2 frequency in the CVC condition is significantly higher than in the CVCV condition. Additionally, for /I/, /ɛ/, and /ʊ/, the F1 frequency is significantly higher in the CVC condition than in the CVCV condition. Thus, it appears to be the case that a syllable-final /t/ raises the F2 frequency of the preceding vowel relative to the F2 frequency in an open syllable.

Given this effect, we now turn to a comparison of the English CVCV tokens with the Spanish CVCV tokens to see if the systematic, cross-language F2 difference that was obtained in the previous experiment persists under conditions of similar word-structure. Table 9 shows the results of the two-tailed, unpaired t-test comparing the F2 values of /i, e, o, u/ from the disyllabic English tokens to the corresponding values from the disyllabic Spanish tokens.

Table 9. Comparison of the mean F2 values of the shared vowels of English and Spanish in disyllabic tokens.

	English	Spanish	mean diff.	t(5)	P-value
i	2268	2147	121	2.110	.0425
e	2020	1814	206	4.823	<.0001
o	1094	1019	75	2.521	.0167
u	1183	992	191	4.098	.0003

Table 10 shows the results of similar tests comparing the F1 values of /i, e, o, u/ of these disyllabic English tokens to the corresponding values from the disyllabic Spanish tokens.

Table 10. Comparison of the mean F1 values of the shared vowels of English and Spanish in disyllabic tokens.

	English	Spanish	mean diff.	t (5)	P-value
i	264	286	-22	-2.844	.0076
e	424	458	-34	-2.459	.0194
o	473	460	13	1.702	.0981
u	316	322	-6	-.567	.5742

In this comparison, we again find that all the English vowels have significantly higher F2 values than their Spanish counterparts, and that the English front vowels are lower in F1 than the Spanish front vowels. Thus, these data indicate that the effect we found in the comparison of the English monosyllabic words and the Spanish disyllabic words cannot be accounted for by the difference in word-structure alone. Rather, these results suggests that indeed the English vowels differ systematically from the Spanish vowels, in a manner that is consistent with the notion of a language-specific base-of-articulation effect.

The crucial point in this comparison of the English and Spanish vowels is that both the front and the back vowels of English are higher in F2 than their Spanish counterparts. This systematic difference for all of the shared vowels of the two languages, indicates a language-specific base-of-articulation effect, rather than an expansion effect due to the crowdedness of the English vowel space relative to the Spanish vowel space. Furthermore, the observed F2 difference cannot be accounted for by a difference in vocal tract length between the two groups of speakers, since we do not find similar shifts for all formants. Finally, an effect of word-structure can be discounted since the observed formant differences across the two languages persist when word-structure is controlled for. Thus, the comparison of the vowel locations across English and Spanish indicate the presence of a language-specific factor in the phonetic realization of phonemic vowel categories.

2.4.3. A comparison of two five-vowel systems

In order to test the independence of this base-of-articulation effect from an effect of inventory size, a comparison of two phonemically equivalent five-vowel systems was performed. For this comparison, the present Spanish data were compared to data for Greek vowels taken from Jongman, Fourakis and Sereno (1989). In that study, the authors recorded four male native speakers of Modern Greek producing four repetitions of words exemplifying the five phonemic vowel contrasts of the language. The words were chosen such that the target vowels were preceded by a bilabial consonant and followed by an alveolar consonant, thus facilitating a direct comparison with the present Spanish data which used the same consonantal context. The words were embedded in frame sentences, and the vowel formant measurements were taken from LPC spectra, as described in that paper. Table 11 gives the mean F1 and F2 values with standard deviations for the Greek data.

Table 11. Greek mean vowel formants in Hz with standard deviations (4 males, 5 repetitions). Data from Jongman et al. (1989).

vowel	F1 (sd)	F2 (sd)
i	310 (6)	2040 (203)
e	474 (12)	1641 (157)
a	680 (63)	1229 (89)
o	476 (5)	854 (62)
u	339 (39)	879 (216)

Tables 12 and 13 show the results of two-tailed, unpaired t-tests on the F2 and F1 values, respectively for the individual tokens for each of the five vowel categories in Greek and Spanish.

Table 12. Comparison of the mean F2 values of the vowels of Greek and Spanish.

	Greek	Spanish	mean diff.	t(34)	P-value
i	2040	2147	-107	-1.763	.0869
e	1641	1814	-173	-3.742	.0007
a	1229	1353	-124	-4.250	.0002
o	854	1019	-166	-5.746	<.0001
u	879	992	-113	-1.926	.0625

Table 13. Comparison of the mean F1 values of the vowels of Greek and Spanish.

	Greek	Spanish	mean diff.	t(34)	P-value
i	310	286	24	5.813	<.0001
e	474	458	16	1.490	.1454
a	680	638	42	2.454	.0194
o	476	460	16	3.164	.0033
u	339	322	17	1.615	.1155

The results of this comparison show that for the non-high vowels the Spanish vowels are significantly higher in the F2 dimension than their Greek counterparts; and, in the case of the two high vowels, /i/ and /u/, there is a strong trend for the Spanish vowels to be higher in F2 than the Greek vowels. In the F1 dimension, there is a general trend for the Spanish vowels to have lower values than the Greek vowels; however, this F1 difference reaches statistical significance for only three of the five vowels, namely, /i/, /a/, and /o/.

The general result of this comparison of the Greek and Spanish vowels is that these two phonemically equivalent vowel systems show a systematic difference regarding the acoustic realization of distinct vowel categories: the vowels of Spanish are generally higher in F2 than the corresponding vowels of Greek. Thus we can conclude that a language-specific, base-of-articulation property plays an important role in determining the location of vowel categories in the acoustic space; and, that this property functions independently of the general size and structure of the vowel inventory. In particular, as shown in Figure 7, the present data indicate that the phonemically equivalent vowels of English, Greek, and Spanish all differ systematically with respect to one another in the F2 dimension. (The English data in this plot are taken from the CVCV data set.)

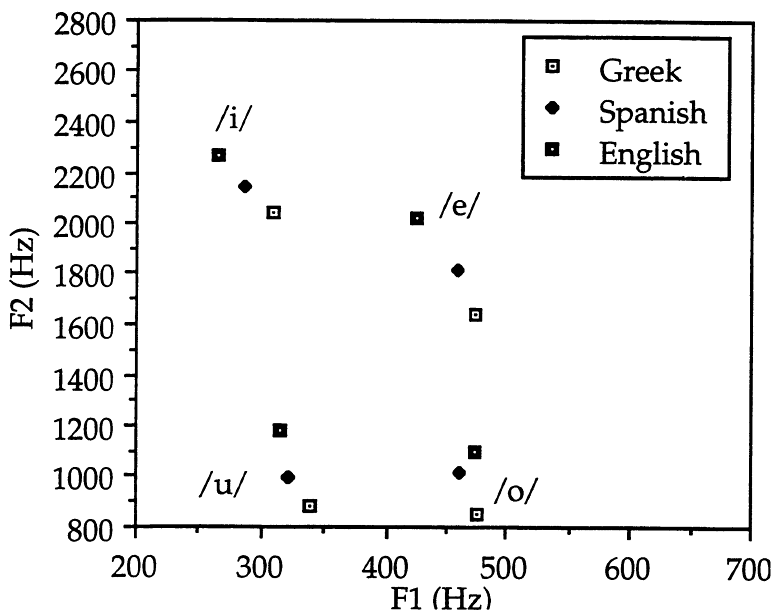

Figure 7. Comparison of the /i/, /e/, /o/, and /u/ locations in Greek, Spanish, and English. All tokens in all languages appear in CVCV contexts.

As shown in Figure 7, in the case of each of the four vowels common to Greek, Spanish, and English, the English vowel is systematically higher in F2 than the Spanish vowel, which is in turn systematically higher in F2 than the Greek vowel. These observations thus indicate that, with respect to these three languages, the general base-of-articulation for English involves the most fronted tongue position, the Spanish base-of-articulation involves an intermediate tongue position, and the Greek base-of-articulation involves the least fronted tongue position.

2.4.4. Expansion of the acoustic vowel space

I now turn to a comparison of the range, or area in the acoustic vowel space, covered by the vowel categories of English and Spanish. Based on the dispersion principle, we expect that the relative crowdedness of the English vowel inventory will cause an expansion of the English acoustic vowel space relative to the Spanish acoustic vowel space. In the comparison of the locations of the English and Spanish vowels in the acoustic space we saw that the English vowels are systematically shifted upward in the F2 dimension relative to the Spanish vowels. However, this difference between the two languages regarding their bases-of-articulation, does not preclude an expansion effect. In other words it is possible that the English vowels are both higher in F2, and cover a greater area than the Spanish vowels.

In order to compare the general range of the English and Spanish vowels, the area covered by the quadrilaterals defined by the mean F1 and F2 values of the four common vowels was calculated. (In this comparison, and in all subsequent comparisons, the original English CVC data set was used.) As an additional point of comparison, the corresponding area covered by the Greek /i/, /e/, /o/, and /u/ was included in this analysis. Figure 8 shows these quadrilaterals in the F1 by F2 spaces of each of the three languages, and Table 14 shows the results of this comparison.

Table 14. Comparison of the /i/-/e/-/o/-/u/ area in English, Spanish, and Greek.

	difference	% difference
English vs. Spanish	+18,832 Hz^2	+12.7%
English vs. Greek	+23,749 Hz^2	+16.6%
Spanish vs. Greek	+4,917 Hz^2	+3.4%

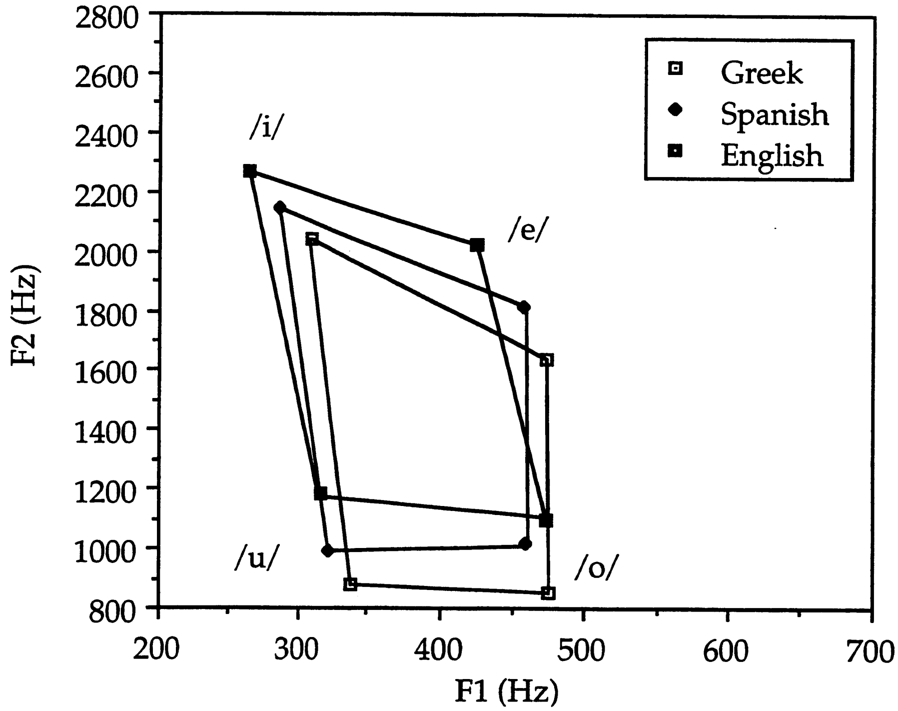

Figure 8. The /i/-/e/-/o/-/u/ areas in English, Spanish, and Greek.

The results of this comparison of the /i/-/e/-/o/-/u/ area in English, Spanish, and Greek show that the English acoustic vowel space covers more area than either the Spanish or the Greek acoustic vowel spaces. The area covered by English /i/-/e/-/o/-/u/ is 12.7% greater than the corresponding area in Spanish, and 16.6% greater than the corresponding area in Greek. Whereas the /i/-/e/-/o/-/u/ areas in Spanish and Greek differ by only 3.4%. This result indicates that the relatively crowded vowel space of English is expanded in the acoustic domain in order to accommodate more than double the number of distinct vowel categories in Spanish and Greek.

2.4.5. Tightness of within-category clustering

In order to compare the tightness of within-category clustering for the vowels of English and Spanish, both the standard deviations and the coefficients of variation of the F1 and F2 values for the four common vowel categories across the two languages were compared. The principal difference between the two measures of within-category variance is that the standard deviation gives a measure of the average deviation of the individual tokens from the category mean whereas the coefficient of variation gives a measure of average deviation taking into account the category mean. The latter is calculated as the category standard deviation divided by the category mean, and is thus useful when comparing samples with different means. Tables 15(a) and 15(b) give the results of these comparisons.

Table 15(a). A comparison of the F1 and F2 standard deviations across English and Spanish /i, e, o, u/.

	English	Spanish	diff.	t(3)	P-value
F1	35	25	10	1.308	.2387
F2	153	129	24	.691	.5154

Table 15(b). A comparison of the F1 and F2 coefficients of variation across English and Spanish /i, e, o, u/.

	English	Spanish	diff.	t(3)	P-value
F1	.094	.065	.029	2.165	.0735
F2	.089	.094	-.005	-.202	.8468

The results of both of these comparisons indicate that there is no significant difference in the tightness of within-category clustering across the four common vowels of English and Spanish.

Another statistical analysis that gives an indication of within-category clustering is the d'_{rms} statistic, as described in Miller (1989). This statistic is calculated as the ratio of the mean between-category distance over the mean within-category distance. The mean between-category distance is determined by calculating the distance of each point in the F1 x F2 space from every point in all of the other categories; the mean within-category distance is determined by calculating the distance of each point in the vowel space from every other point within the same category. Since the ratio of these two values provides a measure of the distances between the categories relative to the distances between points within the categories, the larger the d'_{rms} the better the overall clustering by category. In the application of this statistic to the four common vowels of English and Spanish, we find that the d'_{rms} for Spanish is slightly higher then the d'_{rms} for English (see Table 16 below).

As a way of evaluating the significance of this difference between the d'_{rms} for English and Spanish, we can compare it to the d'_{rms} differences that Miller (1989) found in his comparison of various scales for representing vowels, for example, the mel, bark, and log frequency scales. In that study, Miller used the d'_{rms} measure as a means of quantifying the efficiency of the various scales in representing the distinct vowel categories of English as distinct regions in the acoustic vowel space. The d'_{rms} values that he found for the various scales ranged from 2.87 to 11.08; and, when the plots of the English vowel data on the various scales are compared visually it is clear that this d'_{rms} range represents small differences in within-category clustering. Thus, the small (0.487) difference that we observe in the d'_{rms} values for the four common vowels of English and Spanish can therefore be considered a non-significant difference in within-category clustering.

Table 16. English and Spanish d'_{rms} and d'_{mn}.

	English	Spanish
d'_{rms}	4.127	4.614
d'_{mn}	0.780	0.764

Miller also suggests a second statistic, d'_{mn}, which provides a measure of the overall separation between categories. This statistic is calculated from the ratio of the smallest distance between category means over the mean within-category distance. For this statistic, the larger the value the greater the overall separation between categories. When applied to the four common vowels of English and Spanish, we find that the English data have a slightly larger d'_{mn} than the Spanish data. (These values are also given in Table 16 above.) However, this difference (0.016) can be considered insignificant, given that in Miller's comparison of the various scales for representing the vowels of English, the d'_{mn} values range from 0.25 to 2.38. Thus, the results of these clustering analyses corroborate the results of the comparisons of the standard deviations and coefficients of variance, indicating that the vowels of English are not more tightly clustered than the vowels of Spanish.

The present data also provide a means for testing theoretical predictions regarding differences in clustering between categories within each language. Specifically, according to Quantal Theory (Stevens 1972, 1989) the point vowels, /i, a, u/ should show less within-category variability since they occur in quantal regions of the articulatory vowel space. However, in a test of this prediction, Pisoni (1980) found that the standard deviations of the English point vowels were not significantly smaller than the standard deviations of the non-point vowels. In accordance with that result, based on the present data, we find that the F1 and F2 standard deviations of English /i, a, u/ are not significantly smaller than the F1 and F2 standard deviations of the eight English non-point vowels. Similarly, we find no difference between the F1 and F2 standard deviations for the Spanish point and non-point vowels. The results of these analyses are summarized in Tables 17(a) and 17(b).

Table 17(a). A comparison of the F1 and F2 standard deviations for the point vs. non-point vowels in English.

	point	non-point	mean diff.	t(9)	p-value
F1	45	50	5	.331	.7481
F2	175	138	-37	-1.429	.1867

Table 17(b). A comparison of the F1 and F2 standard deviations for the point vs. non-point vowels in Spanish.

	point	non-point	mean diff.	t(3)	p-value
F1	26	31	5	.422	.7014
F2	126	111	-15	-.522	.6375

Thus, in general these data indicate that the tightness of within-category clustering is not dependent on the size of the vowel inventory. Furthermore, we find that there is no difference between the point vowels and non-point vowels with regards to within-category variation. In other words, there does not appear to be any general determiner of the clustering of individual tokens around the category mean.

2.4.6. Comparison with other data in the literature

There are several sources of acoustic data on the vowels of English and Spanish available in the literature; for example, Delattre (1969), Holbrook and Fairbanks (1962), Lehiste and Peterson (1961), Peterson and Barney (1952), and Zue (1988) provide acoustic measurements of English vowels; and Delattre (1969), Mendez (1982), Quilis and Esgueva (1983) provide acoustic measurements of Spanish vowels. These published data are all consistent with the present acoustic data regarding the location in the acoustic space of the phonemic vowel categories.

For the English data, five of the sources provide formant measurements from vowels in monosyllabic words that end in an alveolar consonant. Three of these studies have an initial bilabial consonant (Bradlow, Lehiste & Peterson, and Zue); the other two (Holbrook & Fairbanks, Peterson & Barney) had initial [h]. Delattre's study used polysyllabic words with the vowel appearing in a stressed position between various consonants. Figure 9 shows the vowel space of English as reported in the literature, together with the data from the present study. For the Spanish data there was more variation in the context of the target vowels: Delattre's word list contained trisyllabic words with various consonantal contexts (*achican, alternan, agradan, recobran, ocupan)*; Mendez used the words *si, paso, tu*; and Quilis and Esgueva used disyllabic words and non-words with the target vowel between two bilabial consonants (e.g. *pipa, biba, pepa, beba* etc.). Figure 10 shows the vowel space of Spanish as reported in the literature, together with the data from the present study.

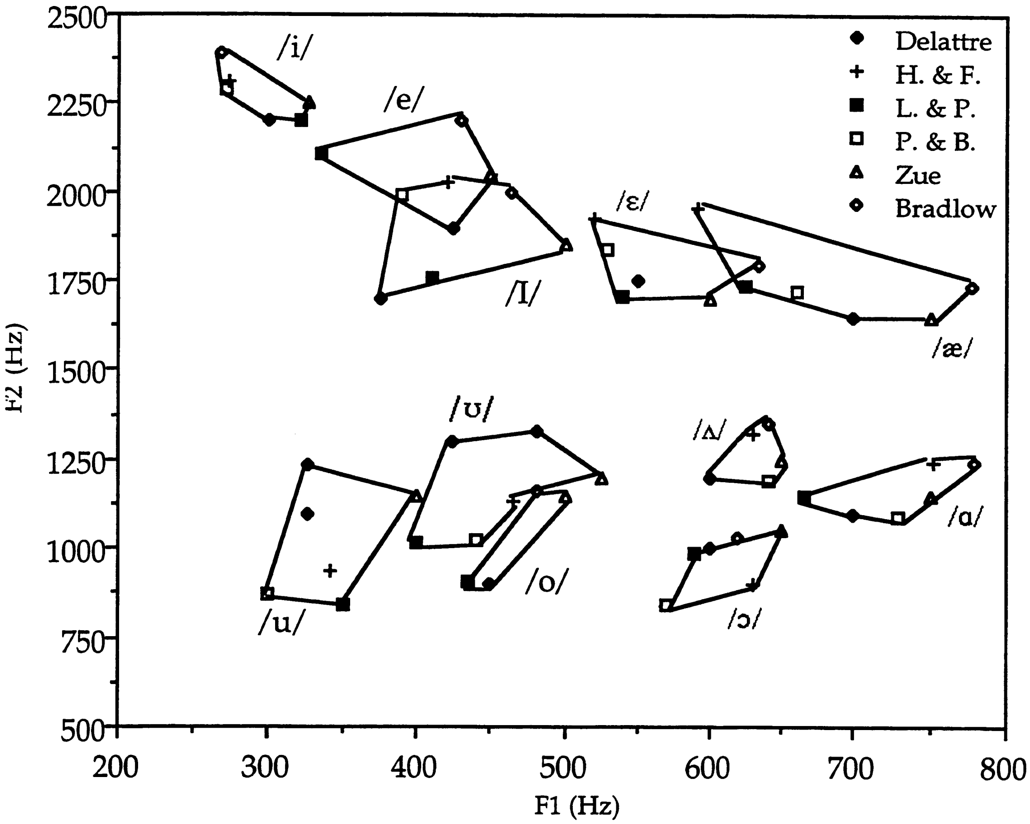

Figure 9. English acoustic data from various sources.

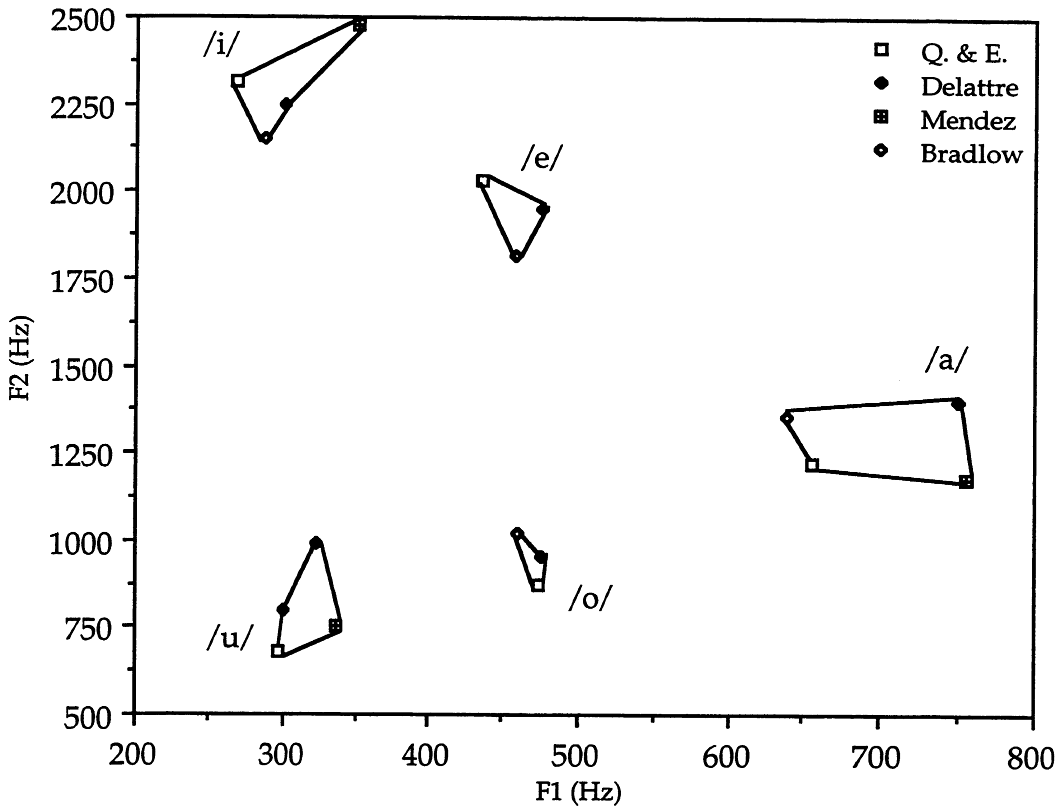

Figure 10. Spanish acoustic data from various sources.

As a further test of the robustness of the results of the present comparison of the English and Spanish vowels, a comparison of the present data with the data from the seven male Spaniards included in the Quilis and Esgueva (1983) study was performed. This particular set of data was selected for this analysis because it is the most comparable to the present data with respect to dialectal background of the speakers, measurement techniques, and consonantal environment of the target vowels. In that study the vowel formants were measured from LPC spectra and spectrograms, and the target vowels appeared between two bilabial consonants. In the present study, the target vowels were always followed by an alveolar consonant. This difference between the two sources of vowel formant measurements regarding the following consonant may have exerted an effect on the vowel formants; thus, the comparison of the Quilis and Esgueva data with the present English and Spanish data is not an ideally controlled comparison. Nevertheless, this comparison serves as an opportunity to assess the validity of the claims made on the basis of the present data, which come from a limited number of subjects.

Figure 11 shows a comparison of the mean locations for /i/, /e/, /o/, and /u/ in the F1 by F2 space between the Quilis and Esgueva Spanish data set (Spanish'), the present Spanish data set (Spanish), and the present English data set. Table 18 shows the means and standard deviations for the Quilis and Esgueva data represented in this plot. The data for the present English and Spanish data shown in this plot are given in Tables 2 and 3 respectively.

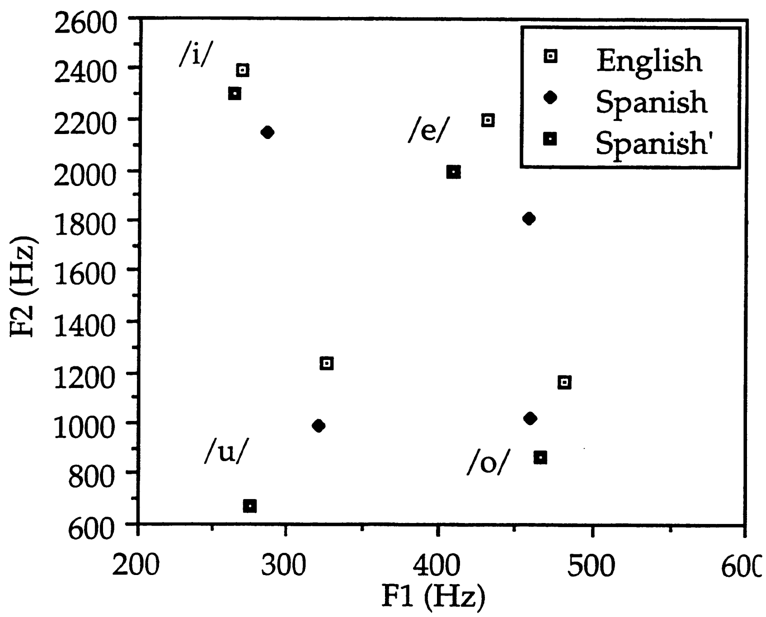

Figure 11. Comparison of /i/, /e/, /o/, and /u/ in English (present data), Spanish (present data), and Spanish data from Quilis and Esgueva (1983), labelled Spanish' in the figure legend.

Table 18. Mean F1 and F2 values with standard deviations for the seven male Spaniards from Quilis and Esgueva (1983).

	F1 (sd)	F2 (sd)
i	263 (44)	2302 (152)
e	408 (51)	1996 (126)
a	648 (32)	1163 (64)
o	466 (71)	862 (79)
u	275 (43)	671 (63)

Tables 19(a) and 19(b) show comparisons of the F1 and F2 values for the Quilis and Esgueva Spanish data with the present Spanish data (Table 19a), and with the present English data (Table 19b).

Table 19(a). Results of a two-tailed, unpaired t-test on the mean F1 and F2 values for all five vowels across the present Spanish data and the Quilis and Esgueva (1983) Spanish data.

	F1			F2		
	mn. diff	t (32)	P-value	mn. diff	t(32)	P-value
i	23.000	2.149	.0399	-155.050	-2.699	.0113
e	50.357	3.135	.0037	-182.171	-4.262	.0002
a	-9.900	-.769	.4474	190.121	-7.034	<.0001
o	-5.350	-.323	.7485	157.079	4.880	<.0001
u	47.079	4.205	.0002	320.757	8.289	<.0001

Table 19(b). Results of a two-tailed, unpaired t-test on the mean F1 and F2 values for /i/, /e/, /o/, and /u/ across the present English data and the Quilis and Esgueva (1983) Spanish data.

	F1			F2		
	mn. diff	t (32)	P-value	mn. diff	t(32)	P-value
i	4.250	.344	.7329	91.150	1.263	.2162
e	22.357	1.286	.2078	203.879	4.001	.0003
o	16.400	.891	.3794	298.279	12.243	<.0001
u	50.679	4.136	.0002	566.557	11.795	<.0001

The results of this comparison indicate that, for /i/, /e/, and /u/, the tokens from the present Spanish data set have significantly higher F1's than the tokens from the Quilis and Esgueva data set. For the other two vowels, /a/ and /o/, the trend is in the opposite direction, that is, for the tokens from the present data to be lower in F1 than the tokens from the Quilis and Esgueva data. In the F2 dimension, the two front vowels from the present Spanish data set are significantly lower than their counterparts from the Quilis and Esgueva data set; however, for the other three vowels, the difference is reversed with the Quilis and Esgueva data being significantly lower in F2. In general, the results of the comparison between the two sets of Spanish data show no clear pattern of variation in the F1 dimension; whereas in the F2 dimension we find that the Quilis and Esgueva vowels are all more peripheral than the corresponding vowels of the present data set. A possible source of this difference between the two data sets is the nature of the lists from which the speakers in the two studies read. In the Quilis and Esgueva study, the list consisted of tokens, most of which were non-words, that were constructed so as to isolate (that is, to stress) the target vowel; whereas in the present study, the target vowels all appeared in words that were read with normal stress. Thus, it is possible that the observed difference between the two sources of data is due to different stress patterns of the speakers, rather than due to measurement error or artifactual characteristics of the individual speakers.

In the comparison of the present English data with the Quilis and Esgueva Spanish data for the four shared vowels of English and Spanish, the general trend in the F1 dimension is for the English data to be higher than the Quilis and Esgueva Spanish data; although, this difference is statistically significant only for /u/. In the F2 dimension, the English vowels are also generally higher than the Quilis and Esgueva Spanish vowels. This difference in F2 is statistically significant for /e/, /o/, and /u/, and for /i/ there is a strong trend in the same direction. Thus, this comparison corroborates the results of the comparison between the present English and Spanish data sets, and the basic finding of a base-of-articulation effect is supported.

2.5. General summary and conclusions of the acoustic comparison

In summary, the present data indicate that the English acoustic vowel space is significantly shifted upward in the F2 dimension relative to the Spanish acoustic vowel space; whereas, in the F1 dimension no consistent differences are found between the vowels of the two languages. Furthermore, in the comparison of the two five-vowel systems of Greek and Spanish, a consistent upward shift is observed in the F2 dimension of the Spanish vowels relative to the Greek vowels. These consistent shifts of the vowels of one language relative to similar vowels in another language indicate that the location of vowels in the acoustic space is determined, at least to some extent, by a language-specific, base-of-articulation property that functions independently of the size and structure of the phonemic vowel system.

We also observed from these data that the English vowels cover a broader range than the Spanish vowels, thus providing evidence in favor of the prediction of Dispersion Theory that larger vowel inventories will result in an expansion of the acoustic vowel space. Furthermore, these cross-language acoustic data showed no difference between the tightness of within-category clustering in English and Spanish, suggesting that larger vowel inventories do not necessarily cause less variation in the acoustic realization of individual tokens within distinct vowel categories. Additionally, the prediction of Quantal Theory that point vowels will show less token variability than non-point vowels was tested by looking at within-category clustering. Contrary to this prediction, in both English and Spanish the data indicate no difference between the degree of clustering for the point vowels and for the non-point vowels.

These findings suggest that the location of the vowel categories of a given language in the acoustic vowel space is determined by a number of factors, including a language-specific, base-of-articulation property. In addition, languages with relatively crowded vowel inventories, such as English, have an expanded vowel space relative to languages with relatively uncrowded vowel inventories, such as Spanish. These two effects on the location of vowels in the acoustic space interact with each other resulting in an acoustic vowel system for a given language with both a language-specific distinctive quality to it, as well as an arrangement of vowel categories which allows for efficient production and perception of the linguistically necessary contrasts.

3. Comparison of acoustic and perceptual /u/-/o/ categories

3.1. Introduction

The findings of the acoustic study presented above suggest that the acoustic realization of vowels is determined by both a language-specific base-of-articulation property, as well as by a general principle of sufficient dispersion between categories This section extends the comparison of the vowels of English and Spanish to the perceptual domain. In particular, this section focuses on the relationship between the acoustic and perceptual characteristics of the /u/-/o/ contrast within each of the two languages.

In the comparison of the produced and perceived vowel spaces, we might expect that the vowel spaces in each of the domains are subject to a variety of different language-specific and universal factors, which may or may not be similar in their effects on the acoustics and perception of the vowels. One possibility is that the perceived categories correspond closely with the acoustic production data, indicating a close, language-specific connection between vowel acoustics and perception. In this case we would expect to find that the perceptual boundary between two vowel categories that are adjacent in the acoustic vowel space of a given language would coincide with the acoustic boundary.

Alternatively, it may be the case that a listener's ability to identify and discriminate speech sounds is subject to a set of auditory constraints that differ in their effect on perceptual category boundary placement from the articulatory constraints that effect acoustic category boundary placement. In this case we would find a certain amount of variation between the two sets of categories. For example, we may find that, in general, we are able to discriminate finer distinctions than the phonemic vowel categories require. This is the result that we may find in the case of languages with relatively uncrowded vowel spaces, which could be said to make use of a subset of the possible vowel contrasts for linguistic purposes.

The reverse situation is also logically possible, that is that the perceived categories are more restrictive than their corresponding categories in the production domain. This would indicate a potential for confusion between vowel categories. It is possible that this potential, should it exist, is realized most readily for isolated vowels and that there are other mechanisms within the sound pattern of a language which counteract this confusion. These may take the form of neutralizations, or mergers between categories, or perhaps the introduction of some other distinctive feature, such as length, source characteristics etc. This is the result that we could expect for languages with very crowded vowel spaces, and would lead us to believe that these languages are approaching some kind of limit on discriminability. In fact, this may be the case in French, with 11 non-nasal monophthongs, where there seems to be a neutralization between mid vowels in open and closed syllables: only tense mid vowels appear in open syllables whereas their lax counterparts only appear in closed syllables.

This section, then, will present data from perceptual experiments on the /u/-/o/ contrast in English and Spanish; and will compare these data to the acoustic data from Section Two for this vowel pair, in order to get an indication of how the acoustic and perceptual vowel spaces of each language correspond to each other.

3.2. Acoustic data

The acoustic data for this part of the study were taken from the set of acoustic data discussed in Section Two of this paper. The details of the methods of recording and formant measurement are described in that section. For the purposes of this comparison between the acoustic and perceptual vowel categories, the F1 and F2 measurements for

English and Spanish vowels were plotted on the perceptually motivated mel scale, rather than on the linear frequency scale used for the cross-language acoustic comparison in Section Two. The use of the mel scale insures that pure tones that are equidistant in mels are also equidistant in pitch. Thus, it insures that, when converted to mels, a change of a given frequency for all formants is perceptually equivalent across all the formants. This scale is particularly appropriate for the stimulus arrangement used in the perceptual experiments; and, in order to facilitate the production-perception comparison, the acoustic production data were also represented on this scale. The exact equation for converting frequencies from hertz to mels is $M = (1000/\log 2)\log(F/1000+1)$ (Fant 1973).

Figure 12 shows M1 x M2 plots of the 20 tokens for the English /u/, /o/, and /Ü/ categories. The ellipses are drawn such that they are centered around the category mean M1 and M2 values, and their dimensions are defined by horizontal and vertical distances equal to two standard deviations from the mean in the M1 and M2 dimensions, respectively. Figure 13 shows similar plots for the Spanish /u/ and /o/ categories.

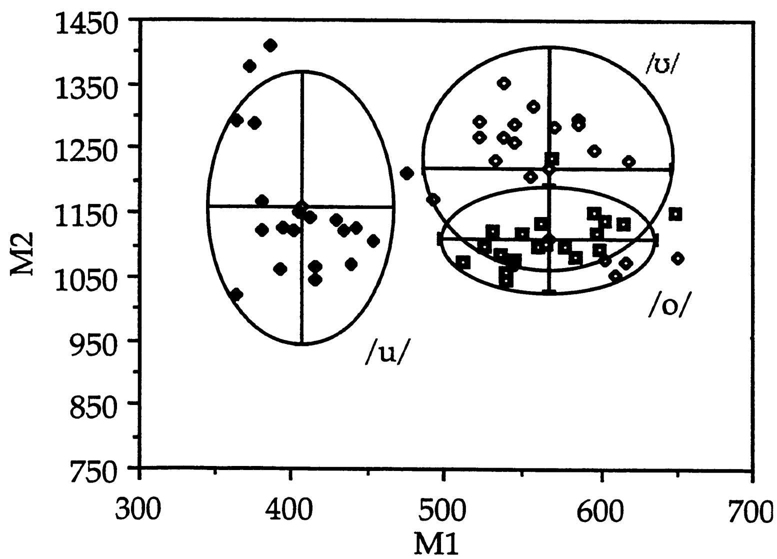

Figure 12. M1 x M2 plot of English /u/, /o/, and /ʊ/ with ellipses covering two standard deviations from the category means.

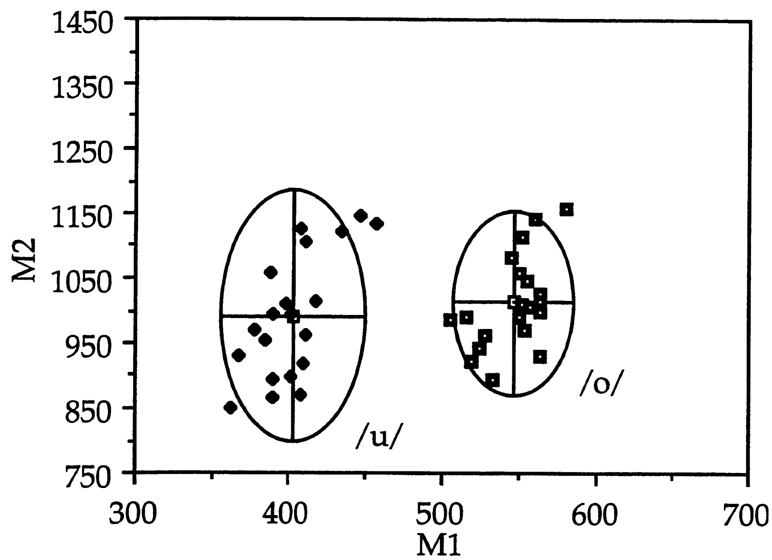

Figure 13. M1 x M2 plot of Spanish /u/ and /o/ with ellipses covering two standard deviations from the category means.

In the production data represented in these plots, it is clear that, in English, the neighboring /ʊ/ category overlaps considerably with the /o/ category. Thus, under the assumption of a close production-perception link, we would expect this overlap to result in a limit on the range of the perceptual /o/ category due to the presence of the neighboring /ʊ/ category. In contrast, the Spanish produced /u/ and /o/ categories are well separated, and we would expect that a close production-perception link would be observed by a Spanish /u/-/o/ perceptual boundary that falls approximately in the region of the acoustic space between the two ellipses in Figure 13. The perception data presented below will provide a means of testing this expectation.

3.3. Methodology for the perception experiments

3.3.1. Stimuli

For these experiments, an adaptation of the general phonetic prototype methodology developed by P. Kuhl and her co-workers at The University of Washington (Grieser and Kuhl 1989, Kuhl 1991) was used. This methodology involves an arrangement of stimuli in concentric circles that radiate from a central "prototype" or mean vowel, and the subject's task is to give each stimulus an identification label, as well as a goodness rating. The starting point for the generation of the stimuli in this methodology is the conversion to mels of the mean vowel formants for the relevant vowel categories. In the present experiments, the mean locations in the M1 x M2 space of /u/ and /o/ in each of the two languages, as measured from the acoustic study discussed in Section Two above were taken as the central points around which the other stimuli were generated.

The set of stimuli for each language consists of two groups: one group around each of the two mean vowels that participate in the contrast under investigation. For each group the stimuli are arranged in concentric circles with radii of 30, 60, 90, and 120 mels from the mean, and there are eight evenly spaced stimuli on each ring. For each group there are 32 stimuli around the mean plus the mean itself, giving a total of 33 stimuli per group. Thus, there are a total of 33x2=66 stimuli per set. Figures 14 and 15 show the English and Spanish /u/-/o/ stimulus sets, respectively. In the English set, the location of the neighboring vowel, /ʊ/, is shown for reference.

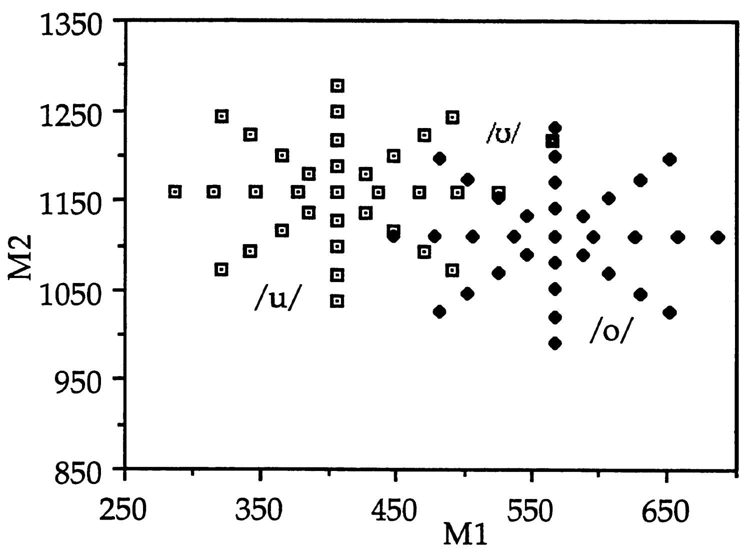

Figure 14. The English /u/-/o/ stimulus set.

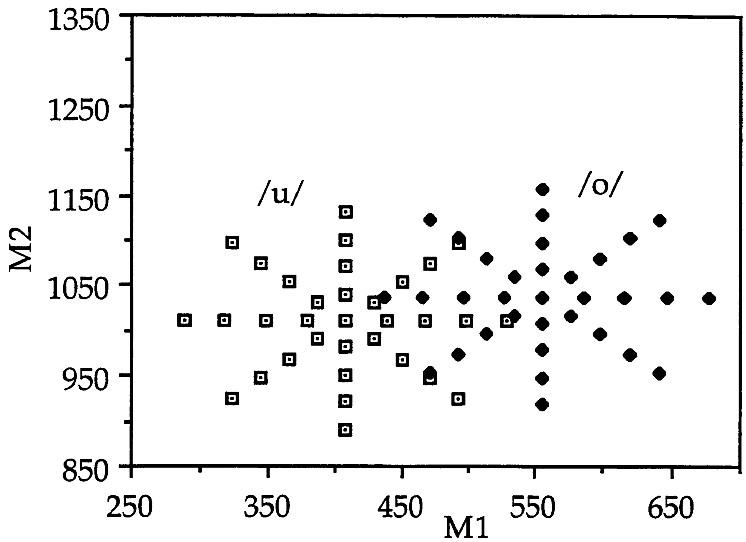

Figure 15. The Spanish /u/-/o/ stimulus set.

3.3.2. Synthesis of the stimuli

Once the M1 and M2 values of each of the 66 vowels in each set of stimuli were specified, these values were converted back to hertz, and used to create synthetic speech stimuli with the Klatt synthesizer on an IBM/AT computer. The program KLPC, written by K. Johnson and Y. Qi, was used to specify the synthesis parameters. The higher formants, F3, F4, and F5, were kept constant across all stimuli at 3010 Hz, 3300 Hz, and 3850 Hz, respectively. The formant bandwidths (B) were set as follows: for F1 B=60 Hz, for F2 B=90 Hz, for F3 B=150 Hz, for F4 B=200 Hz, for F5 B=200 Hz. The duration of all stimuli was 250 ms. The amplitude envelope went from 0 dB to 60 dB over the first 30 ms and was then kept constant throughout the remainder of the vowel duration. The fundamental frequency was kept constant at 100 Hz. The sampling rate of the synthesized stimuli was 10,000 samples per second.

3.3.3. Procedure

In each experiment, subjects responded to each of the 66 stimuli five times, giving a total of 330 trials. In each trial, subjects gave the stimulus an identification label as well as a goodness rating on a scale of one to five. A rating of "5" indicated that the stimulus was a “very good and clear” exemplar of its identification label; a rating of "1" indicated that it was a “very bad and unclear” exemplar of its identification label. At the start of the test, there was a practice run in which subjects responded to 40 stimuli selected so as to be representative of the entire stimulus range.

The English test was run in the Phonetics Laboratory at Cornell University using the BLISS experiment running software package (Mertus 1989). This package played the randomized stimuli directly from hard-disk, and subjects listened over headphones (Sony MDR-7506) while seated in a sound-attenuated booth. Subjects were tested individually, and responded by pressing buttons on button boxes designed for use with the BLISS software package. The buttons were marked "u" and "o," and the subjects were told to identify the stimuli either as "u" as in the word *boot*, or as "o" as in the word *boat*. Subjects had to respond with their identification label within three seconds, as measured from the onset of the stimulus. They then heard a recorded male voice saying the word "rating" and they had to record the rating by pressing numbered buttons within three seconds, as measured from the onset of the recorded prompt.

The Spanish test was run in a quiet room in Madrid. The subjects listened with headphones (Sony MDR-7506) to a tape (played on a Marantz Portable cassette recorder PMD222) of the 40 practice trials, followed by the 330 test stimuli. For each trial, they identified the stimulus by circling the appropriate identification label, and gave it a rating by circling a number from one to five on prepared answer sheets. The 330 trials were divided into 33 blocks of 10 trials each. The inter-trial interval within a block was three seconds and the inter-block interval was five seconds. For the labeling task the subjects were told to identify the stimulus as "u" as in the words *puta* and *tu*, or as "o" as in the words *bota* and *lo*. For the rating task they were told that a rating of "1" indicated a "very bad and unclear" exemplar, and "5" indicated a "very good and clear" exemplar.

3.3.4. General experimental design

For all of the perceptual experiments there were two separate subject groups and two separate conditions. Subjects were either native English speakers or native Spanish speakers, and they participated in either the "homogenous" condition or the "mixed" condition. In the homogenous condition the subjects responded to synthetic stimuli that were based on their native language; in the mixed condition the subjects responded to stimuli that were based on the other language.

Following this design, there are four separate sets of data. First, there are the English homogenous results; second, there are the Spanish homogenous results; third, there are the results of the English listeners' responses to the Spanish stimuli; and finally, there are the results of the Spanish listeners' responses to the English stimuli. The experimental task and the instructions given to the subjects were identical in both the homogenous and the mixed conditions, and the subjects were not informed as to which set of stimuli they were tested on. In all cases the results of the perceptual experiments are compared with the acoustic characteristics of the listeners' native produced vowel categories in the /u/-/o/ region of the vowel space.

3.3.5. Subjects

A total of 28 native English listeners and 20 native Spanish listeners participated in these experiments. A group of seven English listeners participated in each of the two experiments with English listeners; and, a group of five Spanish listeners participated in each of the two experiments with Spanish listeners. Each individual subject participated in only one perceptual experiment. Within each of the two language groups, across all of the perceptual experiments, the dialect of the subjects was always consistent with the dialect of the speakers in the acoustic study. None of the individual English subjects participated in the acoustic study, whereas two of the Spanish subjects did. All subjects were paid for their participation in this study.

3.4 Results and discussion

3.4.1. English subjects with English stimuli

Figure 16 shows the results of the English subjects' responses to the identification task with the English /u/-/o/ stimuli; Figure 17 shows the results of the corresponding rating task. In these plots the stimuli that received a majority of /u/ labels are represented in boldface type, and those that received a majority of /o/ labels are represented in outlined type. In the identification plot (Figure 16) the numbers represent the size of the majority and the enclosed regions are drawn so as to include all stimuli that were identified with less than 80% consistency. In this plot, the total number of responses to each stimulus was 35, since there were seven listeners who responded to each stimulus five times; and, the enclosed region includes all stimuli that were identified consistently less than 28 out of 35 times, that is with less than 80% consistency. In the rating plot (Figure 17), the numbers represent the mean rating for all trials of that stimulus that received an identification label that was consistent with the majority identification label for that stimulus. In both plots the ellipses represent the area in the acoustic space covered by the individual tokens from the production data for the English /u/, /o/, and /ʊ/ categories. These ellipses were drawn at a distance of two standard deviations in each direction from the mean vowel.

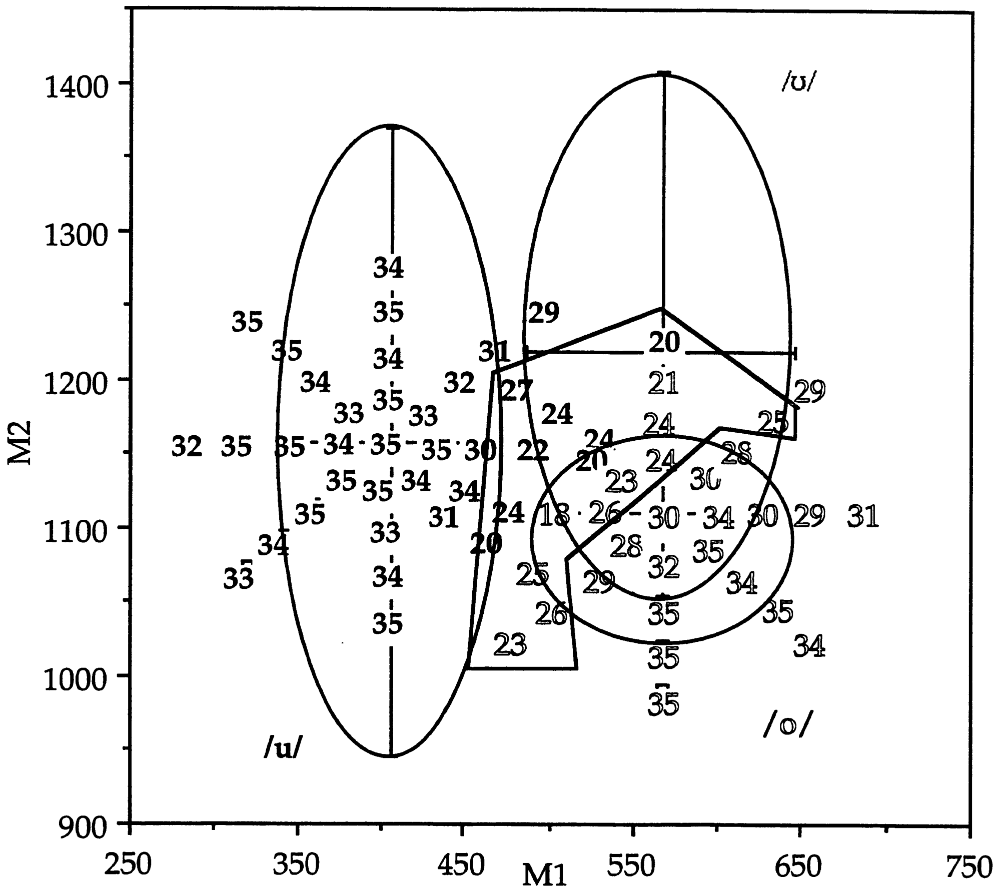

Figure 16. Results of the English /u/-/o/ identification task.

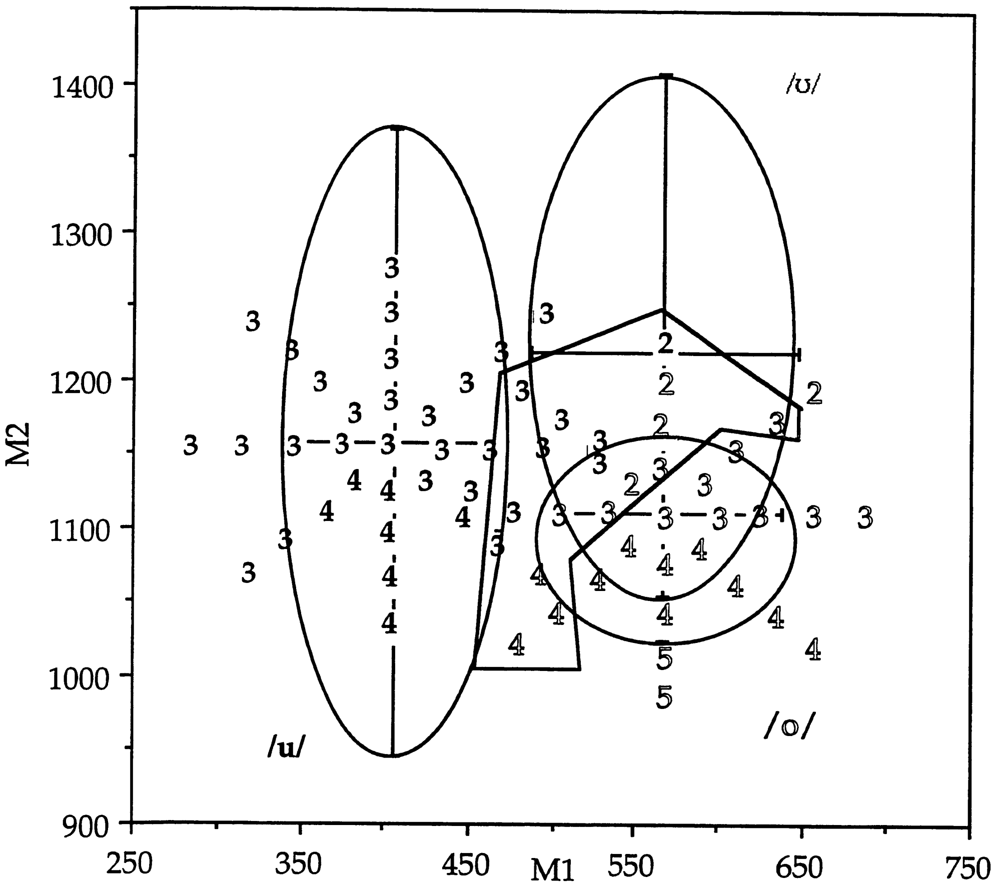

Figure 17. Results of the English /u/-/o/ rating task.

In Figure 16 we observe that the upper F1 edge of the ellipse that represents the produced /u/ category coincides almost exactly with the lower F1 edge of the enclosed region, which represents the region of perceptual uncertainty. We also find that the stimuli that received at least an 80% majority of /o/ labels, that is the stimuli that are represented in outlined type and that are not included in the enclosed region, are all lower in the F2 dimension than the upper F2 edge of the ellipsis that represents the produced /o/ category; conversely, all stimuli with F2 values above the upper edge, and with F1 values above the lower F1 edge of the produced /o/ category, are included in the region of perceptual uncertainty. Thus, we find that the stimuli that lie uniquely within the region of the acoustic space that corresponds to the produced /ʊ/ category are included in the uncertainty region. This results in a region of perceptual uncertainty that represents not only the perceptual boundary between /u/ and /o/, but also the region associated with the neighboring /ʊ/ category. Thus, in general we observe a very close link between the arrangement of the phonemic /u/, /o/, and /ʊ/ vowel categories in the produced and in the perceived vowel space.

We also observe that in the English /u/-/o/ rating task, the higher ratings are generally concentrated around the stimuli with the lower F2 values. This corresponds to the region of the stimulus range that is at the outer edge of the general vowel space implied by these stimuli, and suggests that a low F2 value is a perceptually salient feature for both of these back vowels. This feature of the distribution of higher ratings results in high ratings even for stimuli that lie outside of the /u/ and /o/ acoustic categories. Thus, in the English /u/-/o/ rating task (Figure 17) we find that the highest rating of five occurs for the stimuli with the lowest F2 values, even though these stimuli lie outside of the lower F2 boundary of the English acoustic /o/ region.

3.4.2. Spanish subjects with Spanish stimuli

Figures 18 and 19 show the results of the Spanish subjects' response to the identification and rating tasks with the Spanish /u/-/o/ stimuli, respectively. As in the English plots (Figures 16 and 17), the tokens that received a majority of /u/ identification labels are represented in bold type, and those that received a majority of /o/ labels are in outlined type. In the Spanish identification plot (Figure 18) the numbers represent the size of the identification majority; in the Spanish rating plot (Figure 19) the numbers represent the mean ratings for the stimuli whose identification was consistent with the majority identification label for that stimulus. In the Spanish identification plot (Figure 18), since there were five listeners who responded to each stimulus five times, the total number of responses to each stimulus was 25, and the enclosed region includes all stimuli that were identified consistently less than 20 out of 25 times (that is, with less than 80% consistency). In these plots, the ellipses represent the area covered by the tokens from the production for the Spanish /u/, /o/ categories. These ellipses were drawn at distances of two standard deviations in each direction.

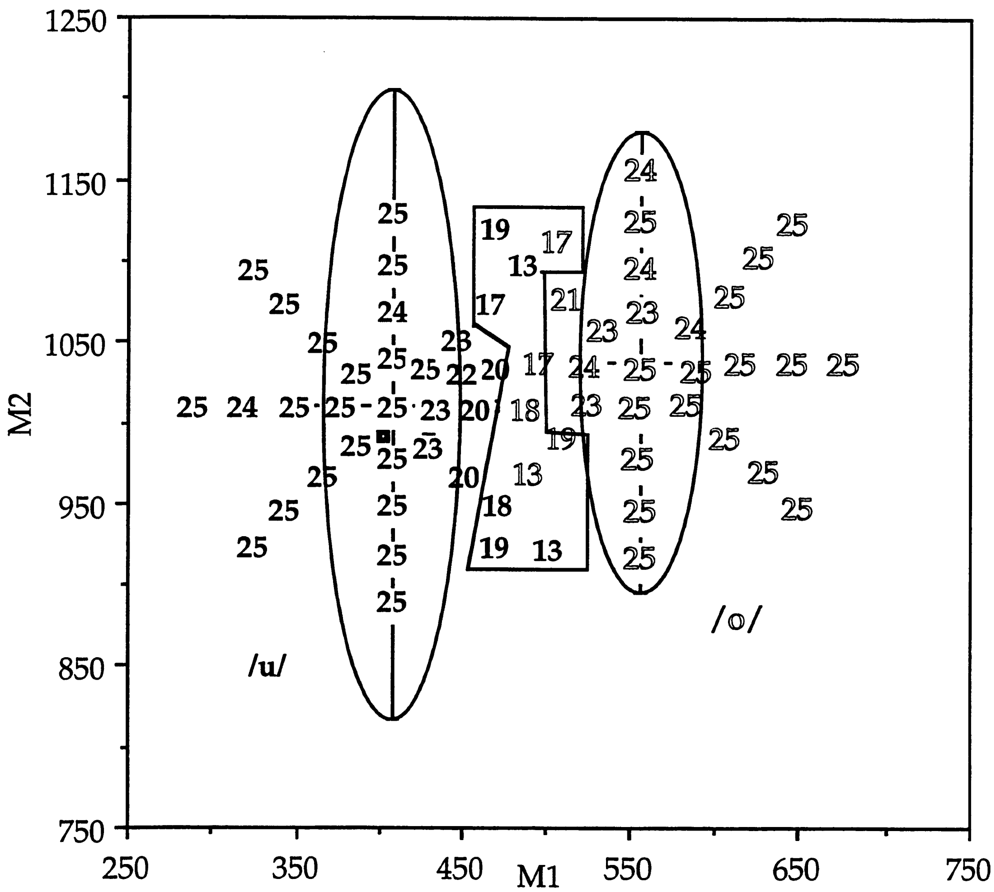

Figure 18. Results of the Spanish /u/-/o/ identification task.

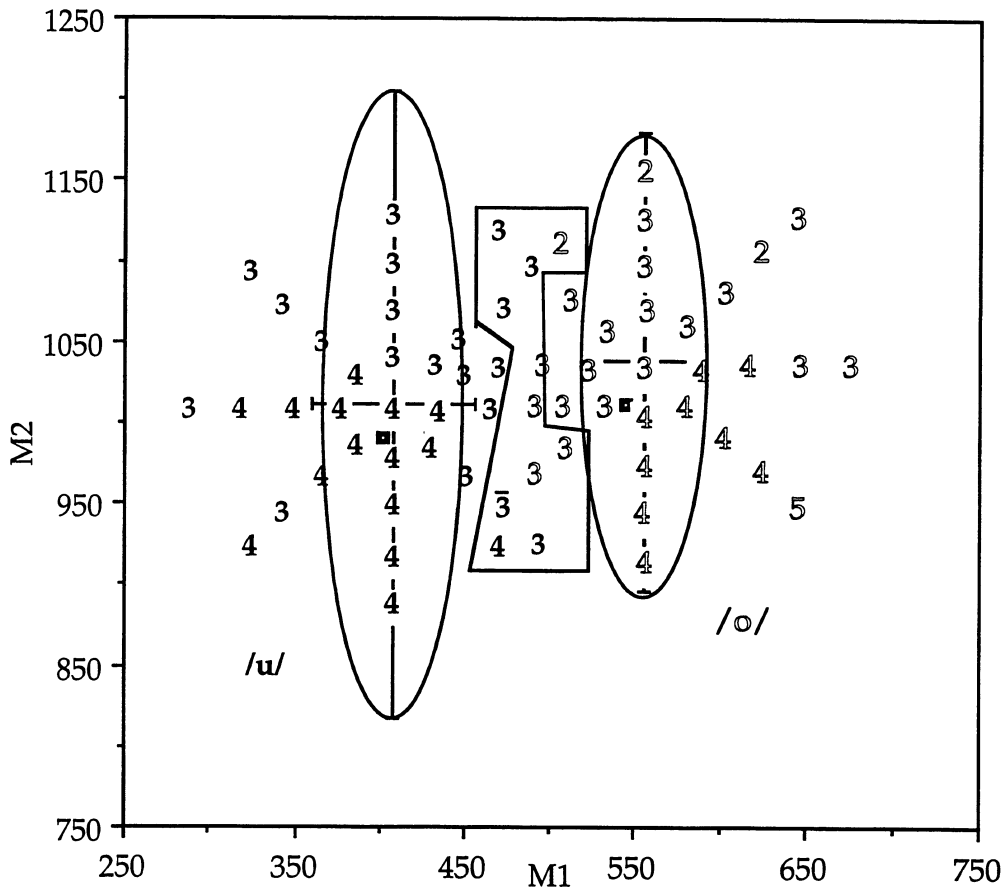

Figure 19. Results of the Spanish /u/-/o/ rating task.

In the Spanish /u/-/o/ identification task (Figure 18) the edges of the perceptual boundary correspond almost exactly to the edges of the region in the produced vowel space between the /u/ and /o/ categories. As expected, the perceptual responses of the listeners show evidence of only two categories in this region of the vowel space, and these two perceptual categories, namely /u/ and /o/, have the F1 and F2 characteristics of the /u/ and /o/ categories that were produced by speakers of the same dialect as the listeners.

In the rating task (Figure 19) the higher ratings are concentrated at the outer (low F2) edge of the vowel space. Furthermore, we find that these higher ratings occur for all stimuli that are at the outer edge of the vowel space implied by these stimuli, even when they lie outside of the range of the produced /u/ and /o/ categories. This pattern of distribution for the higher ratings at the outer edge of the stimulus range indicates that, in general, these Spanish listeners show a preference for back vowels with relatively low F2 values.

Thus far, the first general result of the comparison of the production and perception of /u/ and /o/ in English and Spanish is that listeners judge the vowel stimuli in a way that reveals a very close production-perception link. In the English case, we found that the stimuli that lay outside of the produced /u/ and /o/ categories, but within the neighboring /ʊ/ category, were identified with less consistency than the stimuli that corresponded clearly to either the /u/ or /o/ categories. This effect of the neighboring /ʊ/ category resulted in a broad extent of the region of perceptual uncertainty. In the Spanish case, we found that the perceptual uncertainty region corresponded very closely to the region between the produced /u/ and /o/ categories. In both cases, the location and extent of the uncertainty region could be explained by reference to the general arrangement of the vowel categories in the produced vowel space.

The second general result is that, in both the English and the Spanish /u/-/o/ rating tasks, there is a concentration of higher ratings at the outer edge of the vowel space. For these back vowels, the outer edge corresponds to the lower F2 region of the stimulus range. This general result indicates that for the listeners from both subject groups, a perceptually salient feature of back vowels is a relatively low F2 value. Note also that this concentration of high ratings is at the outer edge of the vowel space, rather than at the outer edge of the stimulus range. For example in the Spanish /u/-/o/ rating task (Figure 19) the higher ratings occur for stimuli with low F2 values, that is for the stimuli that are both at the lower F2 edge of the general Spanish vowel space and at the lower F2 edge of the stimulus range. However we find either average or relatively low ratings for the stimuli with relatively high F2 values, which are at the upper (that is, outer) F2 edge of the stimulus range but on the inside of the general Spanish vowel space. Thus, the general pattern of rating distribution suggests that listeners judge stimuli with reference to the acoustic characteristics of the general vowel space that is defined by the range of stimuli they are presented with.

I now turn to the results of the experiments that involved the presentation of the Spanish /u/-/o/ stimuli to English listeners and the English /u/-/o/ stimuli to Spanish listeners to see if these general results persist even under the condition of non-native stimuli.

3.4.3. English subjects with Spanish stimuli.

Figures 20 and 21 show the results of the English listeners' identification and rating of the Spanish stimuli, respectively. As before, the stimuli that received a majority of /u/ labels are in bold; those that received a majority of /o/ labels are outlined. The numbers in Figure 20 represent the majority sizes (out of 35), and in Figure 21 the numbers represent the mean ratings. The irregularly shaped box includes all tokens that were identified with less than 80% consistency; the ellipses are drawn at a distance of two standard deviations from the /u/, /o/, and /ʊ/ category means.

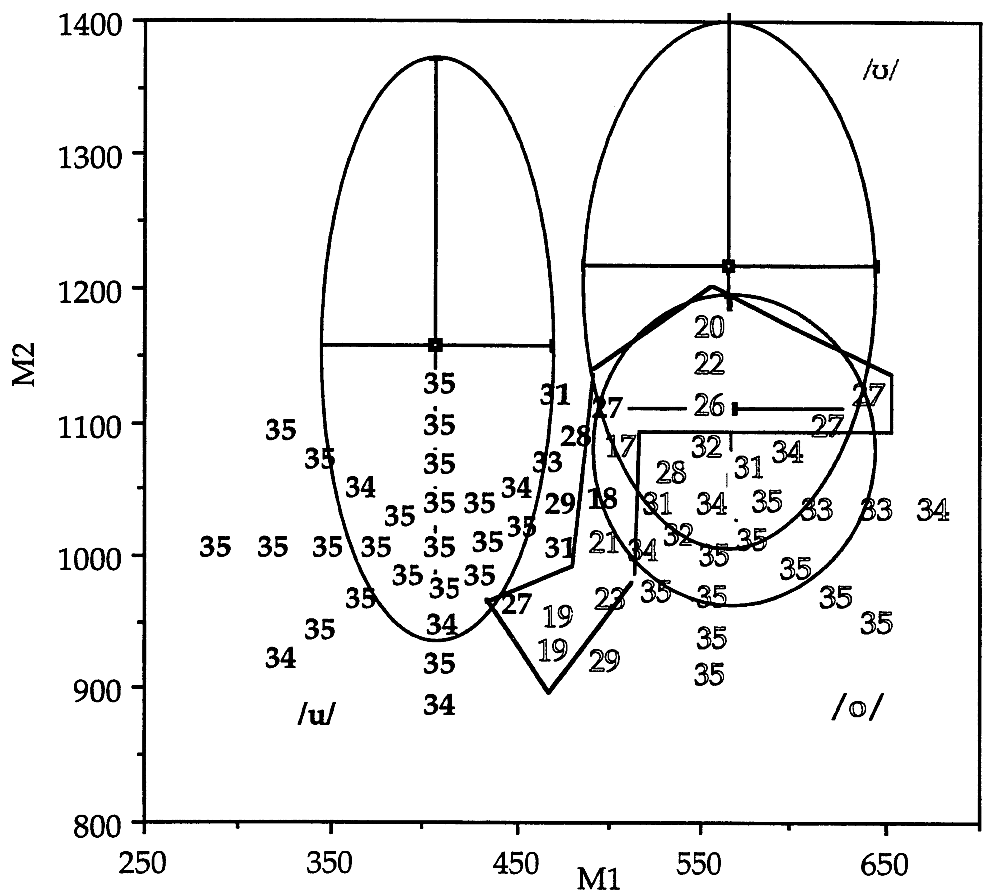

Figure 20. English listeners' identification of Spanish /u/-/o/ stimuli.

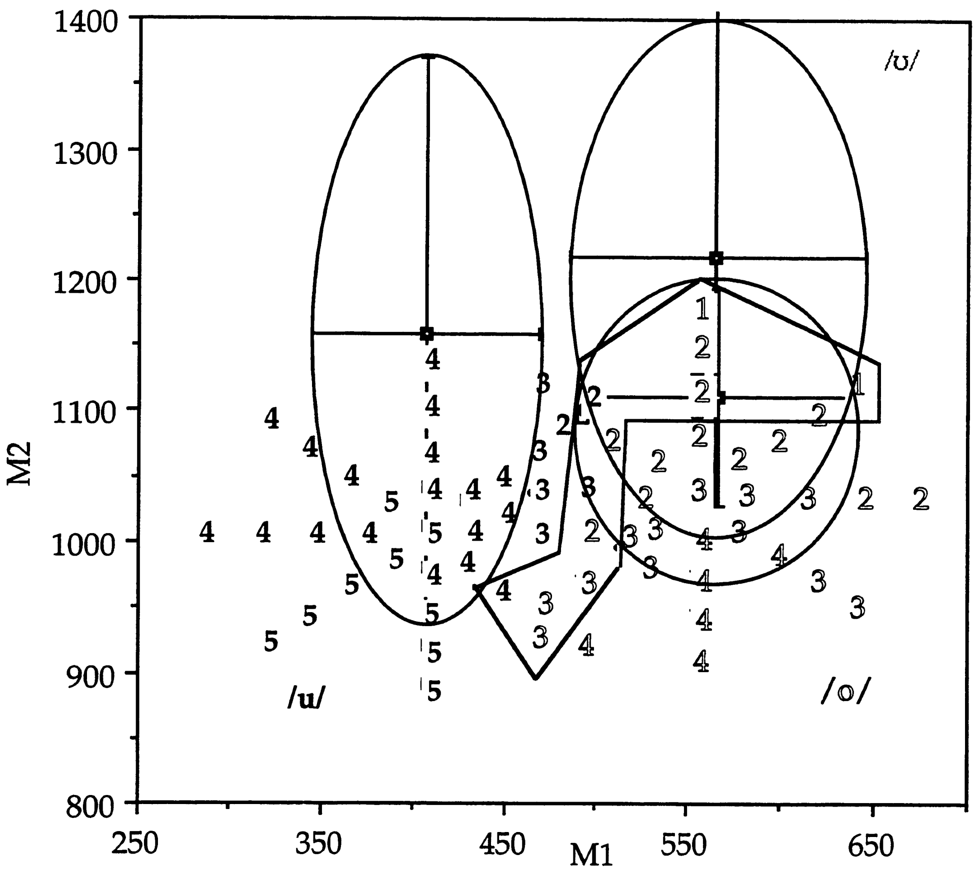

Figure 21. English listeners' rating of Spanish /u/-/o/ stimuli.

The relationship of the Spanish /u/-/o/ stimuli to the vowels in this region of the English produced vowel space is such that several of the stimuli lie outside of the bounds of the general English vowel space, that is several of the stimuli are lower in F2 than the lower F2 boundary of the native English acoustic vowel space. Furthermore, although several of the stimuli around the mean Spanish /o/ lie in the intersection of the English produced /o/ and /ʊ/ categories, none are higher in F2 than the upper F2 bound of the English produced /o/ category. In the perceptual experiment with English listeners and English stimuli, we found that listeners included in the uncertainty region the stimuli that were generated around the mean /o/, and that had F2 values above the upper F2 edge of the produced /o/ category. In other words, in general, these were the stimuli that were within the /ʊ/ range and outside the /o/ range.

In this experiment, we find that the general shape of the uncertainty region, as shown in Figures 20 and 21, matches the general shape of the uncertainty region in the experiment with English subjects and English stimuli (Figures 16 and 17). In both cases we see an effect of the neighboring /ʊ/ category in the extent of the uncertainty region. However, unlike the stimuli in the experiment with English listeners and English stimuli, none of the stimuli in this experiment lie outside of the produced /o/ region. Thus, in the case of the English listeners responding to the Spanish /u/-/o/ stimuli, which are generated around means that lie outside of the bounds of the native English vowel space, the listeners have shifted their perceptual vowel spaces to match the stimuli that they are presented with. This shift results in an uncertainty region whose general shape matches the uncertainty region of the English listeners' responses to the English /u/-/o/ stimuli, even though these stimuli are located in a different region of the vowel space relative to the native English-produced /u/-/o/ categories.

In the results of the rating task with the English listeners and Spanish /u/-/o/ stimuli, we find that the highest ratings are concentrated at the outer (low F2) edge of the vowel space defined by this stimulus range. Furthermore, the distribution of high ratings does not correspond to the location of the native English /u/ and /o/ categories; rather, the highest ratings occur for the stimuli with the lowest F2 values. This result is consistent with the pattern that we saw in the rating task with English listeners and English stimuli, and reinforces the notion that a low F2 value is a perceptually salient feature for back vowels regardless of their relationship to category means.

Thus, the results of this /u/-/o/ perceptual experiment with English listeners and Spanish stimuli fit well into the general interpretation of vowel judgment that we have developed so far. The listeners in this case judge the vowel stimuli with reference to their native vowel space by shifting their perceptual vowel space downward in the F2 dimension to match the stimuli that they were presented with. Furthermore, in the rating task the listeners' give relatively high ratings to the stimuli that lie at the periphery of the vowel space as defined by these stimuli.

3.4.4. Spanish subjects with English stimuli.

Figures 22 and 23 show the results of the identification and rating tasks, respectively for the Spanish listeners with English stimuli. The representation scheme in these plots is the same as in all earlier plots, with the stimuli that received a majority of /u/ tokens in bold, and those that received a majority of /o/ tokens in outline type. The box and ellipses represent the regions of uncertainty and the native Spanish produced /u/ and /o/ categories, respectively.

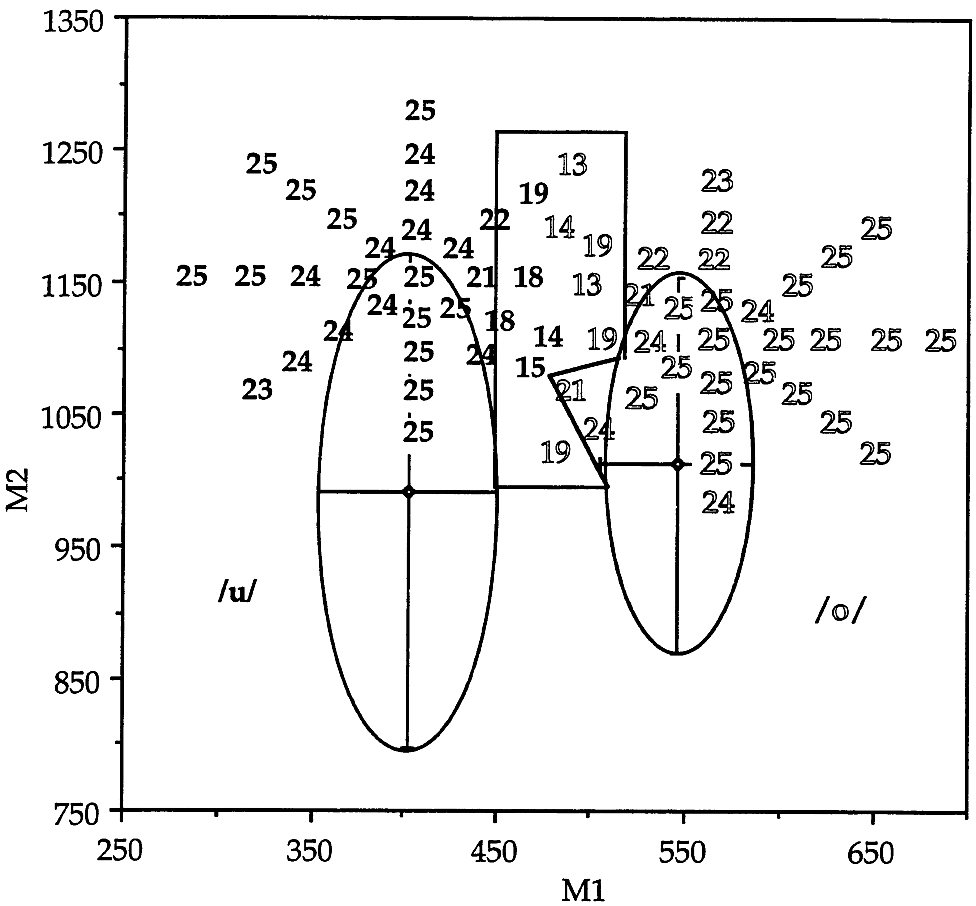

Figure 22. Spanish listeners' identification of English /u/-/o/ stimuli.

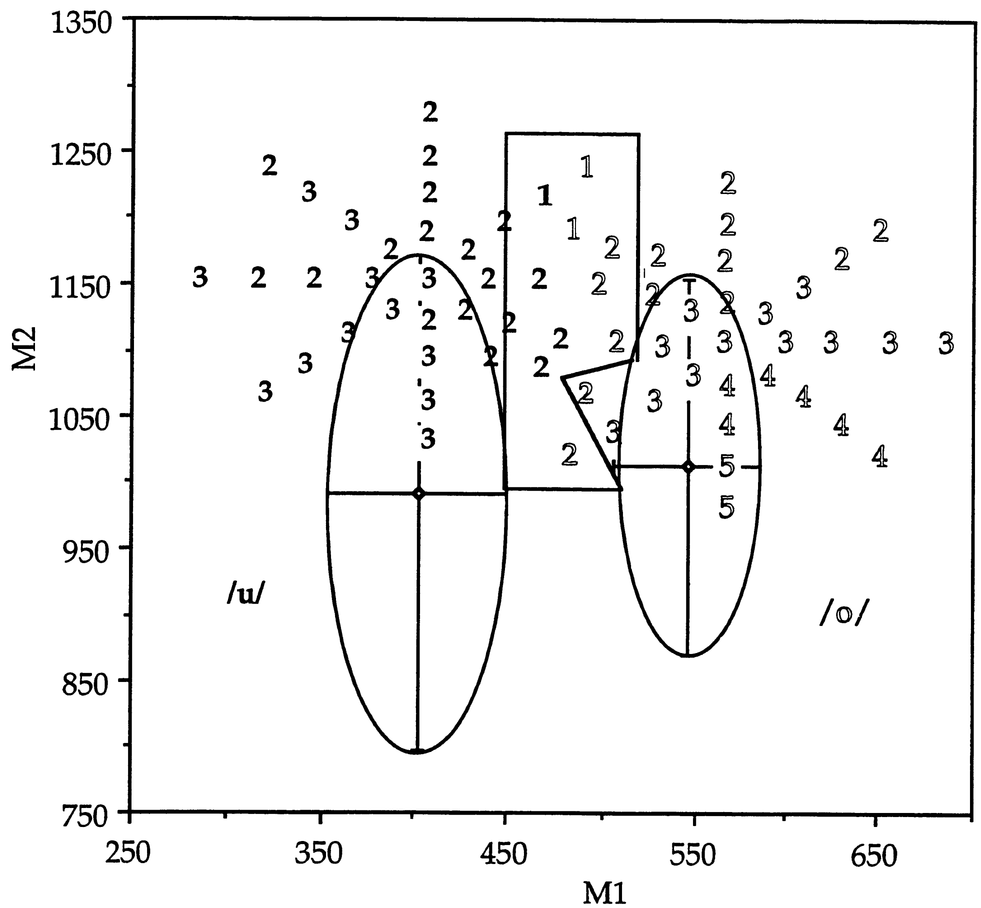

Figure 23. Spanish listeners' rating of English /u/-/o/ stimuli.

Recall that these English /u/-/o/ stimuli are generated around mean /u/ and /o/ locations that differ from the native Spanish mean /u/ and /o/ locations in that they are higher in the F2 dimension. Nevertheless, all of the stimuli lie well within the general range of the native Spanish vowel space. Thus, in the case of the Spanish listeners responding to the English /u/-/o/ stimuli, we expect to find that listeners treat the stimuli as points in their native vowel space, and locate the /u/-/o/ perceptual boundary between the Spanish /u/ and /o/ produced categories. Furthermore, since there are no neighboring Spanish vowel categories in the vicinity of these English stimuli, we do not expect to find a region of uncertainty that reflects anything other than the /u/-/o/ perceptual boundary. Indeed, the uncertainty region in Figures 22 and 23 is located between the Spanish produced /u/ and /o/ categories, and we find no evidence of any boundary besides the /u/-/o/ boundary. Thus, these data provide further evidence of a close production-perception link even for stimuli that are not centered on the produced category means.

In the rating task, as expected, we find a concentration of the higher ratings at the outer (low F2) edge of the stimulus range. In this case the lower F2 region corresponds to the region of the stimulus range that overlaps with the native Spanish produced /u/ and /o/ categories. However, when viewed in the context of the other rating results, this pattern of rating distribution is consistent with the notion that a perceptually salient feature for back vowels is a low F2 value relative to the stimulus range, regardless of the absolute location of the stimulus category centers.

The general results of these mixed condition experiments show clear evidence of a language-specific effect on listeners' responses to vowel stimuli. Specifically these data demonstrate a close connection between the listeners' perceptual vowel categories and their native vowel system as observed in the acoustic vowel categories produced by speakers of the same dialect. In the experiment with English listeners and Spanish /u/-/o/ stimuli, the general shape of the uncertainty region matched the general shape of the uncertainty region in the experiment with English listeners and English /u/-/o/ stimuli, even though the Spanish stimuli were generated around /u/ and /o/ means that are lower in F2 than the lower F2 edge of the native English vowel space. This result suggests that the English listeners shift their perceptual vowel space downward in the F2 dimension to match the Spanish stimuli, and in order to render the stimuli interpretable from an English listener's point of view. In the case of the Spanish listeners' responses to the English /u/-/o/ stimuli, the perceptual boundary region corresponded closely to the region between the native Spanish produced /u/ and /o/ categories, indicating that these listeners identified the non-native stimuli with reference to their native vowel system. Thus, these data provide additional evidence in favor of a language-specific production-perception link for vowels, as well as offering evidence that when presented with stimuli that lie outside of the general range of their native acoustic vowel space, listeners adjust their perceptual vowel space to match the stimuli.

4. Cross-language comparison of the /u/-/o/ perceptual boundary

This section compares the general orientation and location of the /u/-/o/ perceptual boundaries across English and Spanish, in order to determine whether there are any universal factors which play a role in determining the structure of vowel spaces cross-linguistically. One possible outcome of this comparison is that the precise location of the perceptual boundary between two vowel categories is determined on a language-specific basis, which is dependent on the size and structure of the entire vowel inventory and on the location of the vowel categories in the vowel space. For example, in English the vowel /ʊ/ is located in the /u/-/o/ acoustic neighborhood, thus we might expect that the English /u/-/o/ boundary would be shifted slightly relative to the corresponding Spanish boundary, in order to "make room" for this additional category.

The other possibility is that perceptual boundaries are determined by more general, universal factors that are not dependent on language-specific properties such as inventory size and structure. In this case, the boundary between two vowel categories would be in approximately the same place across different languages, even for two languages with such different vowel inventories as English and Spanish. With these possibilities in mind, we now turn to a comparison of the /u/-/o/ boundaries across these two languages.

Figure 24 shows the English (top panel) and Spanish (bottom panel) /u/-/o/ perceptual boundary regions that were observed in the experiments that presented English stimuli to English subjects and Spanish stimuli to Spanish subjects, that is in the homogeneous experimental condition. These figures, which correspond to Figures 16-17 and 18-19 above, show the stimulus groups with the regions of perceptual uncertainty, that is the set of stimuli that were identified with less than 80% consistency, enclosed in the boxes. In the English plot, the uncertainty region is divided into two portions: the portion to the left of the dividing line includes the stimuli that lie in the region between the produced categories in this corner of the English vowel space, and the portion to the right of the dividing line includes the stimuli that were identified with a relatively high degree of inconsistency due to the effect of the neighboring /ʊ/ category. Thus, in the comparison of the /u/-/o/ perceptual boundary across English and Spanish, we compare the portion of the English uncertainty region to the left of the dividing line with the box in the Spanish plot.

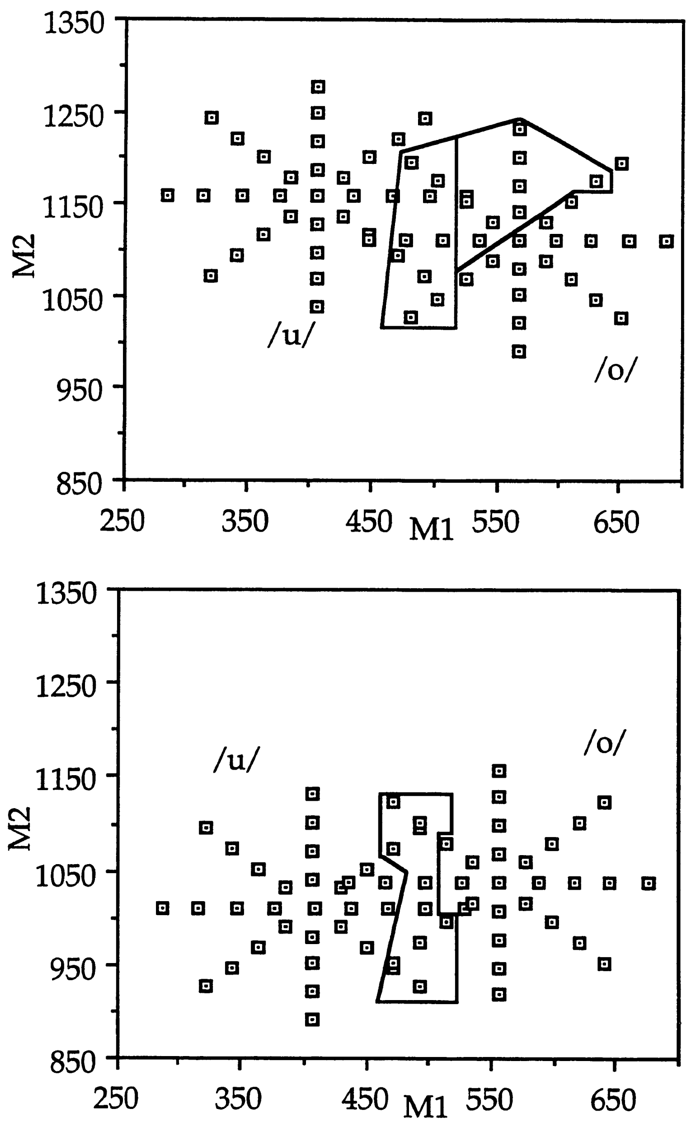

Figure 24. Comparison of the /u/-/o/ boundaries from the homogeneous experimental condition.(top panel shows English listeners' responses to English stimuli; bottom panel shows Spanish listeners' responses to Spanish stimuli)

In both cases we find that the /u/-/o/ boundary region is primarily in the F1 dimension; and, that it is centered around an F1 value of between 450 and 500 mels, which corresponds to a value of between approximately 366 and 414 hertz. Thus, we find that there is consistency across the two languages regarding the location of the /u/-/o/ boundary, even though the general phonemic vowel systems are quite different. In English we find a clear effect of the neighboring /ʊ/ category, which is absent from the Spanish vowel inventory; nevertheless, the English /u/-/o/ perceptual boundary region is located in approximately the same position as the Spanish /u/-/o/ perceptual boundary region.

Figure 25 shows the English (top panel) and Spanish (bottom panel) /u/-/o/ perceptual boundary regions that were observed in the experiments that presented Spanish stimuli to English subjects and English stimuli to Spanish subjects, that is in the mixed experimental condition. These figures correspond to Figures 20-21 and 22-23 above, and show the stimulus groups with the regions of perceptual uncertainty enclosed in the boxes. In the top figure (that is, the figure that shows the English listeners' responses to the Spanish stimuli), the uncertainty region is divided into two portions: the portion to the left of the dividing line includes the stimuli that correspond to the /u/-/o/ perceptual boundary, and the portion to the right of the dividing line includes the stimuli that were identified with a relatively high degree of inconsistency due to the effect of the presumed neighboring /ʊ/ category. Thus, in this comparison of the /u/-/o/ perceptual boundary across the English and Spanish listeners, we compare the portion of the English listeners' uncertainty region to the left of the dividing line with the box in the plot of the Spanish listeners' responses.

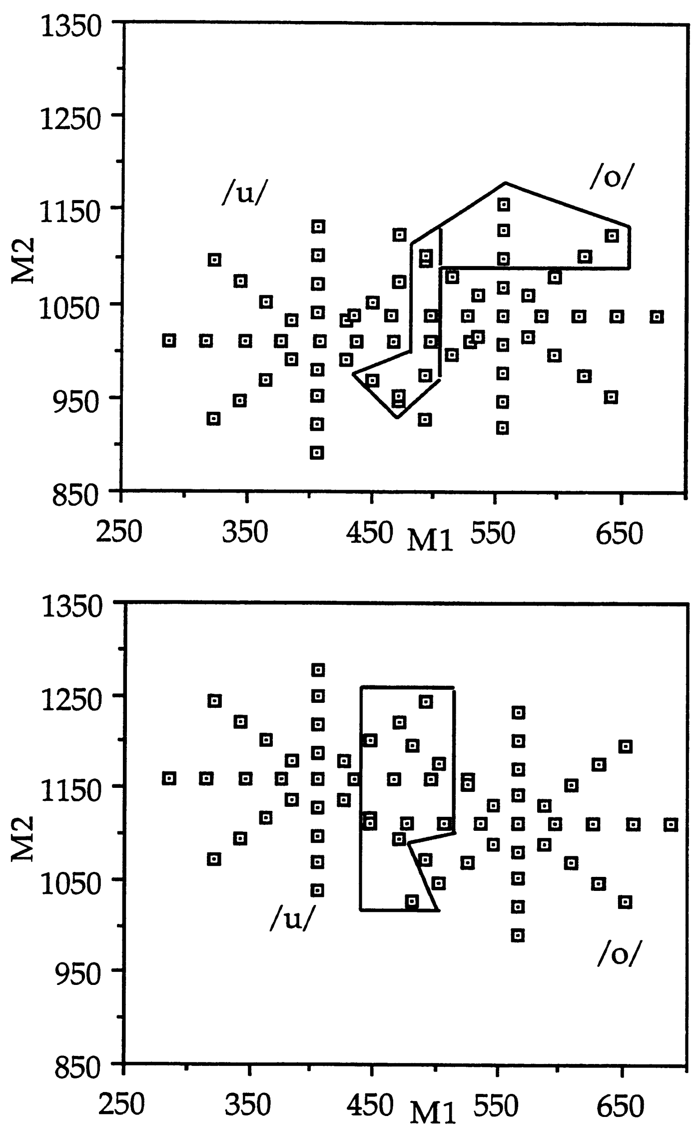

Figure 25. Comparison of the /u/-/o/ boundaries from the mixed experimental condition.(top panel shows English listeners' responses to Spanish stimuli; bottom panel shows Spanish listeners' responses to English stimuli)

As in the homogeneous experimental condition, in the mixed condition shown in Figure 25 above, we find that the /u/-/o/ perceptual boundary regions are primarily F1 boundaries centered around an F1 value of between 450 and 500 mels (366-414 Hz). Thus, in all cases of the /u/-/o/ perceptual experiments we find that the general orientation and location of the perceptual boundaries are relatively consistent. This suggests that there is a universal