A Perceptual Comparison of the /i/-/e/ and /u/-/o/ Contrasts in English and in Spanish: Universal and Language-Specific Aspects

Abstract

Two experiments were run to investigate language-specific and universal acoustic-perceptual characteristics of the /i/-/e/ and /u/-/o/ contrasts in two languages that differ widely in the sizes of their vowel inventories, namely, English and Spanish. In order to investigate the relationship between production and perception of these two vowel pairs, the first experiment compared the acoustic and perceptual characteristics of the /i/-/e/ and /u/-/o/ contrasts within each of the two languages. In order to investigate the responses of listeners to non-native stimuli, the second experiment presented native listeners of English and Spanish with synthetic /i/-/e/ and /u/-/o/ stimuli that were modeled after natural vowels as spoken by speakers of the other language. The results indicated a close production-perception link, as well as a degree of flexibility of the perceptual vowel space in response to stimuli that are structured around non-native vowel category locations. These results suggest that listeners perceive vowel stimuli with reference to a specific phonemic vowel system, and, if necessary, will adjust their perceptual vowel system to match the stimuli. Additionally, a comparison of the absolute locations of the /i/-/e/ and /u/-/o/ perceptual boundaries across the two languages indicated the presence of universally preferred, inventory-independent boundary locations.
Introduction

Cross-linguistic surveys of speech sound inventories indicate that certain vowels and vowel system configurations are preferred [e.g. Maddieson, 1984]. For example, the point vowels, /i/, /u/, and /a/, are present as phonemic categories in the vast majority of the languages included in Maddieson’s survey of 317 languages. Additionally, frequency distributions of vowel inventory sizes indicate that two thirds of the languages in this sample have between five and seven distinct vowel qualities. Furthermore, it is the same five to seven vowel categories that occur in language after language. Thus, we may hypothesize that there are certain universally preferred articulatory and/or perceptual characteristics that influence the organization of vowel categories. However, due to differences in vowel inventory size and structure we might also expect that vowel categories will differ across languages in their extent and location in the vowel space. For example, we might expect to observe cross-language differences in the acoustic and perceptual characteristics of the point vowels due to the presence of additional categories in one language relative to another. In order to investigate the relationship between the general, auditory and language-specific features that structure linguistic vowel systems, the present study was designed to compare the acoustic realization and perceptual categorization of shared vowel contrasts across languages with different vowel inventories.

The interaction of the universal and language-specific factors that determine a listener’s perceptual vowel space has received considerable attention. For instance in a cross-language study of vowel perception for English and Swedish subjects, Stevens et al. [1969] found that the identification functions of the two subject groups differed according to the vowel categories of the two languages: however, these differences did not carry over into the discrimination tests. Rather, in the discrimination tests, both Swedish- and English-speaking subjects exhibited similar regions of enhanced discriminability between certain vowel pairs from rounded and unrounded continua. The authors interpreted this result as indicating that certain aspects of speech perception are independent of language background, and that there are general auditory-perceptual mechanisms that play an important role in speech perception. They hypothesized further that linguistic categories across languages will tend to be located between these regions of high discriminability, and that these general perceptual mechanisms are a possible source of some of the cross-linguistic tendencies observed in the structure of vowel inventories.

The notion of universally preferred perceptual boundaries was formalized in Stevens’ [1972, 1989] Quantal Theory of Speech. This theory posits a nonmonotonic relationship between certain changes in the acoustic domain and the perceptual response of the listener. As evidence in favor of this proposition, Stevens [1989] cites an experiment by Chistovich and Lublinskaya [1979]. In this experiment, subjects were asked to match a one-formant target vowel to reference stimuli with two formants which vary in proximity to each other. Below a critical value of proximity the subjects picked a frequency which was intermediate between the two formants of the reference vowel; whereas, when the two formants were farther apart than the critical value, they matched the one formant of the target vowel to one of the two formants in the reference vowel. The specific critical distance found in these experiments was approximately 3–3.5 bark units. This result (the spectral center of gravity effect) was taken as an indication of a quantal aspect to the relationship between the acoustic and perceptual domains.
Syrdal and Gopal [1986] further explored this effect and its manifestations in vowel systems by developing the notion of a bark distance classification scheme for vowels. In this scheme, vowel categories are organized around the 3–3.5 barks found by Christovich and Lublinskaya [1979], such that the frequency differences between formant pairs form the basis for phonological vowel feature contrasts. For example, vowels for which the $F_1-F_0$ frequency difference is less than 3–3.5 bark are classified as high vowels; whereas those for which the $F_1-F_0$ frequency difference is greater than 3–3.5 bark are classified as nonhigh vowels [Syrdal and Gopal, 1986; see also Traunmüller, 1981; Hoemeke and Diehl, 1991]. This approach to vowel systems focuses on demonstrating the existence of universal, auditory principles that constrain the structure of phonemic vowel systems, and that lead to favored acoustic vowel categories across languages. The present experiments further explore the contribution of general, auditory features in the structure of vowel systems by comparing the categorization of similar vowels by listeners from different language backgrounds. Specifically, we explore the effect on vowel categorization of the presence in one language versus the absence in another language of a neighboring vowel category.

The present study was also designed to explore the extent of the correspondence in the acoustic domain between vowel production and perception. The finding of a close production-perception link in the acoustic domain would be suggestive of a language-specific factor in the general structure of vowel systems, and might provide some insight into the structure of the perceptual vowel space that listeners refer to when responding to vowel stimuli. Scholes [1967a] investigated the categorization of a set of 69 synthetic vowels that vary in both $F_1$ and $F_2$ by speakers of several different languages. In these experiments the listeners were instructed to categorize the synthetic stimuli according to words in their native language that exemplify the relevant vowel contrasts. The results of these studies showed a close correspondence between the phonemic system of the listeners and their categorization of the stimuli: in general, the categorization of the full set of 69 vowels matched the general structure of the vowel inventory of the listener's native language. In a follow-up study, Scholes [1968] compared the categorization of the set of 69 vowels by non-native speakers of English according to the vowels of their native language with their categorization according to the vowels of English. The motivation behind this study was to investigate native language interference in the perception of a foreign language. The results of this study showed that listeners' categorizations in terms of the English vowels were predictable from their categorization in terms of their native language. Scholes [1968, p. 102] concluded with the hypothesis that 'the categorization of synthetic vocalic stimuli by a non-native speaker of English using English responses will parallel his native language categorization of these stimuli insofar as the native language categories have phonetically similar English counterparts'. The present study builds on these findings by performing a direct comparison of the acoustic characteristics of naturally produced vowel tokens and perceptual category boundaries for speakers from two languages with different vowel inventories.

The extensive body of literature on cross-language speech perception in adults [for recent reviews see Polka, 1991; Best, 1993; Strange, 1994, and references therein] and work on infant language development [for a recent review see Jusczyk, in press] address directly the contributions of language-specific and universal factors to the structure of per-
ceptual categories. The cross-language studies on adults show different sensitivities to different classes of non-native categories, implicating both universal, acoustic factors and language-specific, phonemic factors in determining the structure of speech categories. For example, Beddor and Strange [1982] showed increased sensitivity by American listeners relative to Hindi listeners to within-category differences of the Hindi /baː/–/baː/ contrast, which is a non-native contrast in English; whereas, Miyawaki et al. [1975] showed reduced sensitivity by Japanese listeners relative to American listeners to the English /iː–ɪ/ contrast, which is neutralized in Japanese. Work on infants has shown that as early as 6 months of age infants from different language environments exhibit different sensitivities to phonemic vowel categories that are present or absent in the phoneme inventory of the language to which they are exposed. For example, Kuhl et al. [1992] found that Swedish and English babies differ in their responses to high front vowels in a manner that coincides with the different phonemic vowel categories across the two languages in that region of the acoustic vowel space. Yet, there is evidence that prior to this age infants exhibit sensitivity to vowel contrasts that are not native to the language of their environment [e.g., Werker and Polka, 1993]. Thus, the current data from cross-language and infant studies indicate a complex interaction of universal and language-specific factors in the structure of phonemic inventories.

The present study seeks to contribute to this body of literature by focusing on the acoustic and perceptual characteristics of two shared vowel contrasts, the /iː–ɪ/ and /uː–oʊ/ contrasts, within and across two languages with very different vowel systems. The languages selected for this study were English with a relatively large inventory of eleven stressed monophthongs, and Spanish with a relatively small inventory of five vowels. The main purpose of this study was to explore the effect of the presence in one language, versus the absence in the other, of additional neighboring vowel categories. Thus, this study sought to explore the effect of inventory differences on the perception of vowel contrasts that are native to both languages. The expectation being tested was that the perceptual categories are similar across the two languages due to acoustic similarities of the vowels in question; however, we also expect to observe differences in the extents of the perceptual vowel categories due to the effects of different neighboring categories. For example, in the case of the /iː–ɪ/ contrast, we expect to observe an effect of the neighboring /uː/ category in English that limits the range of vowels in the phonetic vowel space judged as good /iː/ 's and /ɪ/ /'s. In Spanish, we expect such an effect to be absent. Similarly, in the case of the /uː–oʊ/ contrast we expect to observe an effect of the neighboring English /uː/ category, whereas in Spanish there is no neighboring category to exert such an effect. We also expected to observe a production-perception link within each language, that is that the perceptual categories in each language coincide acoustically with the production categories as measured from naturally spoken vowels.

The general design of the study included two experiments. In experiment 1, synthetic /iː–ɪ/ and /uː–oʊ/ stimuli were presented to native English- and Spanish-speaking listeners. In this experiment subjects always responded to stimuli that were modeled after naturally spoken vowels from their native language. The observed perceptual categories were then compared to the acoustic categories produced by native speakers of the respective languages. These data provided a means of evaluating the extent of a production-perception link for these vowel contrasts in these languages. In experiment 2, synthetic /iː–ɪ/ and /uː–oʊ/
Table 1. English and Spanish word lists and frame sentences

<table>
<thead>
<tr>
<th>English</th>
<th>Spanish</th>
</tr>
</thead>
<tbody>
<tr>
<td>beat</td>
<td>/beit/</td>
</tr>
<tr>
<td>bit</td>
<td>/bi/</td>
</tr>
<tr>
<td>bait</td>
<td>/bat/</td>
</tr>
<tr>
<td>bet</td>
<td>/bet/</td>
</tr>
<tr>
<td>bat</td>
<td>/bat/</td>
</tr>
<tr>
<td>but</td>
<td>/but/</td>
</tr>
<tr>
<td>boot</td>
<td>/buit/</td>
</tr>
<tr>
<td>put</td>
<td>/put/</td>
</tr>
<tr>
<td>boat</td>
<td>/bot/</td>
</tr>
<tr>
<td>bought</td>
<td>/bot/</td>
</tr>
<tr>
<td>pot</td>
<td>/puit/</td>
</tr>
</tbody>
</table>

Frame sentence
Say _____ again.

Frame sentence
Escribe _____ bien.

‘Write _____ well.’

Experiment 1

Acoustic Data

The acoustic characteristics of English and Spanish /a, /e/, /u/, and /o/ were determined from formant measurements of vowels spoken by speakers of each of the two languages. Four male speakers of General American English and 4 male speakers of Madrid Spanish read words exemplifying the vowel contrasts of their respective languages. The target vowels were all preceded by a bilabial consonant and followed by an alveolar consonant. The words were embedded in frame sentences and each sentence was repeated 5 times in random order, giving a total of 20 tokens (4 speakers x 5 repetitions) for each vowel category of each language [see Bradlow, in press, for a detailed discussion of these data]. The target words and frame sentences are given in Table 1.

All recordings were made with a Marantz portable cassette recorder PMD222 and an AKG D310 microphone. The recordings were digitized with a sampling rate of 12,000 Hz and low-pass-filtered at 6,000 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. The formant values

Phonetics 1996; 53: 55–85
were then read from the LPC spectrum and checked with readings from the spectrogram. In the case of  

diplongized English /e/ and /ə/ the formant  
measurements were taken from the portion of the vowel before  
the offglide. The auditory quality of this portion of the  
the vowel was checked, in order to insure that the formant  
measurements being recorded were those appropriate  
for /e/ or /ə/ rather than for the /i/ or /u/ offglides. A  
comparison of the formant values for both the English  
and Spanish vowels in this study with those from previous  
studies showed that the present measurements are similar to other data reported in the literature [e.g.  
Peterson and Barney, 1952; Delattre, 1969; Mendez,  
1982].

For the purposes of the comparison between the  
acoustic and perceptual vowel categories, the \( F_1 \) and  
\( F_2 \) measurements for English and Spanish vowels  
were plotted on the perceptually motivated mel scale.

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. As noted  
by Kuhl [1991], 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  
\[ M = (1,000/\log 2) \log ((F/1,000)+1) \] [Pam, 1973].

Figure 1 shows \( M_1 \times M_2 \) plots of the 20 tokens for  
the English and Spanish vowels in the /i/-/e/ (fig. 1a,  
b) and /u/-/ə/ (fig. 1c, d) regions of the vowel space.  
In the English plots the neighboring /i/ and /e/ categories  
are included in these regions; whereas in Spanish  
there are no other vowels in these regions of the vowel
space. The ellipses were drawn such that they are centered around the category mean \( M_1 \) and \( M_2 \) values, and their dimensions are defined by horizontal and vertical distances equal to two standard deviations from the mean in the \( M_1 \) and \( M_2 \) dimensions, respectively.

As shown in these plots, there is considerable overlap between the English /æ/ and the neighboring /u/ category. Similarly, the neighboring English /æ/ category overlaps considerably with the English /ə/ category. The presence of these neighboring categories in English might be expected to affect the perception of the /æ/-/æ/ and /u/-/ə/ contrasts for English listeners; however, the perception of these contrasts by Spanish listeners should be unaffected by other vowels in the system. Thus, under the assumption of a close production-perception link, we would expect these overlaps in English to result in a limit on the range of the perceptual /æ/ and /ə/ categories due to the presence of the neighboring /u/ and /æ/ categories, respectively. In contrast, the Spanish acoustic categories in the /u/-/æ/ and /u/-/ə/ regions of the vowel space are well separated, and we would expect that a close production-perception link would be reflected by Spanish /u/-/æ/ and /u/-/ə/ perceptual boundaries that fall approximately in the region of the acoustic space between the two ellipses in the Spanish plots in figures 1b and 1d, respectively. The perception data presented below provide a means of testing this expectation. It is important to note that the English category overlap observed in figure 1 is dependent on the fact that this plot shows only two dimensions and ignores differences in durational and dynamic spectral properties between categories. A representation that takes these additional factors into account might reveal less overlap than shown in this figure. For the purpose of the present acoustic-perceptual comparison, however, only static \( F_1 \) and \( F_2 \) characteristics were considered.

### Perception Data

#### Stimulus Arrangement

An adaptation of the general phonetic prototype methodology developed by Grieser and Kuhl [1989] and Kuhl [1991] was used. This methodology involves an arrangement of stimuli in concentric circles that radiate from a central vowel. The stimuli vary only in \( F_1 \) and \( F_2 \) frequencies, 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 formant frequencies for the relevant vowel categories. In the present experiment, the mean locations in the \( M_1 \times M_2 \) space of the vowels under investigation in each of the two languages, as measured from the acoustic data, were taken as the central points around which the other stimuli were generated.

For each language, two sets of stimuli were generated: an /u/-/æ/ set and an /u/-/ə/ set. Each set of stimuli for each language consisted of two groups: one group around each of the two mean vowels that participate in the contrast under investigation. For each group the stimuli were arranged in concentric circles with radii of 30, 60, 90, and 120 mels from the mean, and there were eight evenly spaced stimuli on each ring. For each group there were 32 stimuli around the mean plus the mean itself, giving a total of 33 stimuli per group. Thus, there were a total of 33 x 2 = 66 stimuli per set.

Since the stimuli varied only in \( F_1 \) and \( F_2 \) target frequencies, this methodology isolated the effects of static spectral properties from effects of other characteristics of naturally produced vowels, such as duration and spectral change. For the purposes of the present experiment, which was designed as an initial attempt to explore the effect of inventory crowdedness on the structure of perceptual vowel categories, this methodology was deemed appropriate. The rationale here was to investigate whether static \( F_1 \) and \( F_2 \) features alone are sufficient to reveal an effect of neighboring categories, and in so doing to draw attention to potentially universal and language-specific characteristics of vowel systems. Should such an effect be demonstrated, follow-up experiments that investigate the role of other vowel characteristics would be warranted.

### Stimulus Synthesis

Once the \( M_1 \) and \( M_2 \) 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, \( F_3 \), \( F_4 \), and \( F_5 \), were kept constant across all stimuli at 3,010, 3,300, and 3,850 Hz, respectively. The formant bandwidths (B) were set as follows: for \( F_1 \) B = 60 Hz, for \( F_2 \) B = 90 Hz, for \( F_3 \) B = 150 Hz, for \( F_4 \) B = 200 Hz, for \( F_5 \) B = 200 Hz. The duration of all stimuli was 250 ms. The amplitude envelope went from 0 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.

**Subjects**

A total of 14 native English listeners and 10 native Spanish listeners participated in experiment 1. A group of 7 English listeners participated in each of the two conditions with English stimuli (one /ɪ/-/ɛ/ condition and one /ʌ/-/ə/ condition); and a group of 5 Spanish listeners participated in each of the two conditions with Spanish stimuli (also one /ɪ/-/ɛ/ condition and one /ʌ/-/ə/ condition). Each individual subject participated in only one condition. Within each of the two language groups, across all conditions, the dialect of the subjects was always consistent with the dialect of the speakers on whose productions the stimuli were based. Like the English speakers who provided the acoustic models, the English subjects had all spent all of their adult years in the Northeastern region of the United States. Similarly, the Spanish subjects had all spent most of their adult years in Madrid, as had the Spanish speakers. Several of the subjects in both groups had studied one or more foreign language, however none described him or herself as ‘fluent’ in a second language. English subjects’ ages ranged from 18 to 38, with a mean of 21.4 years. Spanish subjects’ ages ranged from 22 to 47, with a mean of 30.6 years. All subjects were paid for their participation, and none reported any history of speech or hearing impairment.

**Procedure**

In each condition, subjects responded to each of the 66 stimuli in one stimulus set 5 times, giving a total of 330 trials per condition. In each trial, subjects gave the stimulus an identification label as well as a goodness rating on a scale of 1–5. 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 tests were run in the Phonetics Laboratory at Cornell University using the BLISS experiment running software package [Mertus, 1989] on a PC (Swan 386/25). The randomized stimuli were played directly from the hard disk, and subjects listened over headphones (Sony MDR-7506) while seated in a sound-attenuated booth. Before the first trial subjects adjusted the output intensity to a comfortable level which then remained fixed for the duration of the session. Subjects were tested individually and responded by pressing buttons on button boxes designed for use with BLISS software. For the /ɪ/-/ɛ/ condition with English listeners, since the subjects were not familiar with the IPA symbols, the buttons were labeled ‘E’ and ‘A’ (rather than ‘i’ and ‘e’), and the subjects were told to identify the vowel as ‘E’, as in the name of the letter and in the word beat, or as ‘A’, as in the name of the letter and in the word boat. For the /ʌ/-/ə/ condition with English listeners, the buttons were marked ‘u’ as in the word boot, or as ‘o’ as in the word boat. Subjects had to respond within 3 s with their identification label, as measured from the onset of the stimulus. They then heard a recorded male voice saying the word ‘rating’ and they entered their rating by pressing numbered buttons (1–5) within 3 s, 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. Before beginning the practice trials, subjects were given the chance to adjust the output intensity to a comfortable level which then remained fixed for the duration of the session. For each trial, they identified the stimulus by circling the appropriate identification label, and gave it a rating by circling a number from 1 to 5 on prepared answer sheets. (For the Spanish subjects there was no explicit rating prompt, however the answer sheets made it clear that a rating was required.) The 330 trials were divided into 33 blocks of 10 trials each. The intertrial interval within a block was 3 s and the interblock interval was 5 s. For the /ɪ/-/ɛ/ labeling task the Spanish subjects were told to identify the stimulus as ‘i’ as in the words buta and si, or as ‘e’ as in the words beta and te. For the /ʌ/-/ə/ labeling task the Spanish 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.

The English /ɪ/-/ɛ/ Contrast

Figure 2 shows the results of the English subjects’ identifications (fig. 2a) and ratings (fig. 2b) of the English /ɪ/-/ɛ/ stimuli. In this case, 7 subjects responded to each stimulus 5 times, yielding a total of 7 x 5 = 35 responses to each stimulus. In this figure, and in all other /ɪ/-/ɛ/ plots throughout the article, stimuli that
received a majority of /i/ identification labels are represented in boldface type and stimuli that received a majority of /e/ responses are represented in italics. The numbers in the identification plots represent the size of the majority, with possible maxima of 35 (where there were 7 subjects) or 25 (where there were 5 subjects). For example, a boldface 35 in fig-

Fig. 2. English listeners' identification (a) and ratings (b) of the English /i/-/e/ stimuli.
overlaps considerably with the acoustic /u/ category. This overlap would lead us to expect that in the perceptual domain, the /e/ category would show some effect of this neighboring category. However, in the identification task, subjects were equally consistent in their /e/ labeling as they were in their /i/ labeling, despite the overlap of the /e/ and /i/ categories that we observe in the acoustic vowel space: for both the perceptual /i/ and the /e/ groups the average size of the identification majority was 95%. Yet, in the rating task subjects generally gave higher ratings for the /i/ tokens than for the /e/ tokens: the mean rating for the 30 tokens that received at least an 80% majority of /i/ identification labels is 3.73; whereas the mean rating for the 24 tokens that received at least an 80% majority of /e/ identification labels is 3.10. (An unpaired t test on these ratings shows that this difference between the /i/ and /e/ ratings is highly significant \(t(64) = -4.841, p<0.0001\).) These generally lower ratings for the stimuli in the /e/ perceptual category relative to the ratings for the stimuli in the /i/ perceptual category reflect an uncertainty, which may be linked to the overlap of the /i/ and /e/ categories in the acoustic domain. Thus, although the identification data do not show any clean effects of this overlap, the rating data apparently do show a link between the overlap of the acoustic categories and listeners’ judgments of vowel quality. An alternative explanation for the lower ratings for /e/ than for /i/ is that the lack of diphthongization for the /e/ stimuli leads the listeners to judge them as worse exemplars of the English /e/ category. However, as discussed below, this explanation is not supported by the rating data from the /i/-/e/ condition.

In general then, the results of the rating task reflect preferences, whereas the consistency of the identifications reflects discriminability of one category from the other. This difference between ratings and identification consistency
is shown by the only moderate Spearman rank-order correlation between ratings and size of identification majorities of $R^2 = 0.536$. This suggests that listeners are generally equally consistent in accepting tokens as exemplars of a particular category throughout the area of the perceptual space appropriate for that category, regardless of how they rate those tokens. However, when asked to rate the goodness of individual stimuli, subjects exhibit preferences for stimuli in certain regions. In the English /i/-/e/ case, listeners label stimuli as /i/ or /e/ with equal consistency across the two vowel categories; however, the results of the rating task show lower ratings for /e/ than for /i/, presumably due to the effect of the neighboring /i/ category.

An additional observation concerning the results of the English /i/-/e/ rating task is that the range of high /i/ ratings extends well beyond the bounds of the acoustic /i/ region. In other words, the listeners gave high goodness ratings of 4 or 5 to the /i/ stimuli with lower $F_1$ values than the low $F_1$ boundary of the acoustic /i/ region. The effect of this rating pattern is that the stimuli at the extremes of the stimulus range receive high goodness ratings even when they are outside of the bounds of the corresponding acoustic category. Furthermore, in contrast to the data presented by Grieser and Kuhl (1989) and Kuhl (1991), the present data do not show a pattern of rating distribution in which the ratings decrease as the stimuli diverge from the central vowel. Rather, these data show a narrower range of ratings, and the ratings do not decrease with distance from the central vowel. Differences in both the stimuli and subjects' tasks may have resulted in the observed differences between the present data and the data from Kuhl's laboratory. In addition to using synthetic stimuli based on vowels from a different set of speakers, the stimuli in the present study were shorter (250 versus 500 ms) and had flat, rather than rise-fall, $F_0$ contours. Furthermore, the task in this experiment involved identification as well as rating, rather than just rating.

The Spanish /i/-/e/ Contrast

The identification and rating plots for the Spanish /i/-/e/ test are given in figure 3. As before, the box encloses all stimuli that received a consistent label for less than 80% of the trials. Since there were 5 Spanish listeners, the maximum for the identification majorities in the top panel of figure 3 is 25; the box encloses all stimuli that received a consistent label less than 20 out of 25 trials. The ellipses correspond to the ellipses in figure 1b and represent the general area covered by the native Spanish /i/ and /e/ acoustic vowel categories.

In the identification plot in figure 3 we observe that the Spanish /i/-/e/ region of uncertainty (the region enclosed by the box) is shifted towards the /i/ end of the stimulus range. This shift is reflected by the fact that 42% of the total number of tokens received a majority of /i/ identification labels, and 58% received a majority of /e/ identification labels. We also observe that this region of uncertainty coincides very closely with the region between the two ellipses, which correspond to the regions in the acoustic space covered by the produced /i/ and /e/ categories. In other words, the location of the perceptual boundary between Spanish /i/ and /e/ and the location of the Spanish produced /i/ and /e/ categories appear to be linked quite closely to each other.

In the results of the Spanish /i/-/e/ rating task we observe that, unlike in the case of the English /i/-/e/ rating task, the mean rating for all tokens that received more than 80% /i/ labels does not differ significantly from the mean rating for all tokens that received more than 80% /e/ labels ($t(64) = -0.877$, p = 0.3840). (These means are 2.973 and 3.102,
respectively.) However, the rating pattern in the Spanish case shows a concentration of high ratings at the peripheries of the vowel space, that is for the stimuli with relatively high $F_2$ values. Conversely, the lower ratings appear on the inside edge of the vowel categories, that is the low $F_2$ edge. Thus, in the Spanish /i/-/e/ case, the rating pattern seems to re-
fect the fact that for front vowels a perceptually salient feature is the relatively high F2; consequently, stimuli with higher F2 values were judged as better exemplars of these front vowel categories than stimuli with lower F2 values.

We also find that although subjects prefer front vowels with high F2 over front vowels with lower F2, they were equally consistent in their identification of vowels at the outer and inner edges of the vowel categories. This observation is expressed statistically by the moderate Spearman rank-order correlation between ratings and size of identification majorities of $R^2 = 0.522$. Thus, even though the membership of a particular vowel stimulus to an identification group may be undisputed among listeners, this stimulus might receive a low rating since the listeners all agree that it is a poor exemplar of the category.

The English /u/-/o/ Contrast

Figure 4 shows the results of the English subjects’ responses to the identification and rating tasks with the English /u/-/o/ stimuli. These plots follow the same representation scheme as the /u/-/o/ plots except that in the /u/-/o/ 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 italics. In this plot, the total number of responses to each stimulus was 35, since there were 7 listeners who responded to each stimulus 5 times, and the ellipses represent the area in the acoustic space covered by the native English /u/, /o/, and /u/ acoustic categories as shown in figure 1.

In figure 4 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 italics and that are not included in the enclosed region, are all lower in the F2 dimension than the upper F1 edge of the ellipse 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 /o/ 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 /o/ category. Thus, in general we observe a very close link between the arrangement of the phonemic /u/, /o/, and /u/ vowel categories in the produced and in the perceived vowel space, and we observe a clear effect of the neighboring /u/ category.

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 we find that the highest rating of 5 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. Furthermore, these English rating data do not show generally higher ratings for the stimuli labeled /u/ than for those labeled /o/.
even though naturally spoken English /o/ tokens are normally diphthongized. This finding supports the interpretation of the lower /e/ ratings as indicating an effect of the overlap between /e/ and /u/, rather than as indicating an effect of the lack of an /e/ offglide. If the lower /e/ ratings were a result of the absence of diphthongization, then we would expect similarly
lower /o/ ratings in the /a/-/o/ condition. Thus, we can conclude that, whereas in the /a/-/e/ condition the effect of the neighboring /a/ category is seen in the lower /e/ ratings, in the /a/-/o/ condition the effect of the neighboring /o/ category is seen in the extent of the uncertainty region.

The Spanish /u/-/o/ Contrast

Figure 5 shows the results of the Spanish subjects' response to the identification (fig. 5a) and rating tasks (fig. 5b) with the Spanish /u/-/o/ stimuli. In these plots the total number of responses to each stimulus was 25, since there were 5 listeners who responded to each stimulus 5 times. The ellipses in these plots represent the native Spanish /u/ and /o/ acoustic categories as shown in figure 1d.

In the Spanish /u/-/o/ identification task 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 /F_1/ and /F_2/ characteristics of the /u/ and /o/ categories that were produced by speakers of the same dialect as the listeners.

In the rating task the higher ratings are concentrated at the outer (low /F_2/) 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 /F_2/ values.

Thus, as in the /a/-/e/ case, the comparison of the production and perception of /a/ and /o/ in English and Spanish indicates 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 /a/ and /o/ categories, but within the neighboring /a/ category, were identified with less consistency than the stimuli that corresponded clearly to either the /a/ or /o/ categories. This effect of the neighboring /a/ 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.

Furthermore, and again as in the /a/-/e/ case, in both the English and the Spanish /u/-/o/ rating tasks, there is a concentration of higher ratings at the outer edge of the entire vowel space implied by the stimuli. For these back vowels, the outer edge corresponds to the lower /F_2/ 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 /F_2/ value. Note also that this concentration of high ratings is at the outer edge of the entire vowel space, rather than at the outer edge of the stimulus range. For example in the Spanish /u/-/o/ rating task (fig. 5) the higher ratings occur for stimuli with low /F_2/ values, that is for the stimuli that are both at the lower /F_2/ edge of the general Spanish vowel space and at the lower /F_2/ edge of the stimulus range. However we find either average or relatively low ratings for the stimuli with relatively high /F_2/ values, which are at the upper (that is, outer) /F_2/ edge of the stimulus range but on the inside of the general Spanish vowel space. Thus, the gen-
Fig. B. Spanish listeners' identification (a) and ratings (b) of the Spanish /u/-/o/ stimuli.

geral pattern of rating distribution as seen in figures 2–6 suggests that listeners judge stimuli with reference to the acoustic characteristics of the entire vowel space implied by the stimuli they are presented with: based on the range of stimuli they hear they infer the structure of the entire acoustic-phonetic vowel space and judge the stimuli accordingly.
The general results of these acoustic-perceptual comparisons of the /i/-/e/ and /u/-/o/ vowel contrasts within each of the two languages reveal a close language-specific production-perception link, as well as a tendency in both languages for subjects to give higher goodness ratings to vowels at the extremes of the entire implied vowel space. In order to investigate whether there are cross-linguistic consistencies in the general perceptual characteristics of these category boundaries, we now turn to a comparison of the /i/-/e/ and /u/-/o/ perceptual boundaries in English and Spanish.

Cross-Language Comparison of the /i/-/e/ Boundaries

Figure 6 shows the English (fig. 6a) and Spanish (fig. 6b) /i/-/e/ stimulus groups with the regions of uncertainty enclosed by the boxes. The general orientation of the boxes indicates that in both languages the primary distinction between /i/ and /e/ occurs along the F1 dimension. However, we do observe that for one English /i/ stimulus with an F1 value in the range of the perceptual boundary region and an F2 value above approximately 1750 mels (indicated by the arrow in the top panel of figure 6), the relatively high F2 overrides the F1 boundary effect. Similarly, for the one Spanish stimulus with an F1 value in the range of the boundary region and an F2 value below 1,442 mels (indicated by the arrow in the bottom panel of figure 6), the relatively low F2 of /e/ overrides the F1 boundary effect. Thus, although the primary distinction between /i/ and /e/ in both English and Spanish is in the F1 dimension, the F2 frequency can also make an important contribution to this perceptual (and phonological) distinction.

We also observe that for both languages the region of uncertainty is not in the center of the range of stimuli. In the English identification task, 44% of the stimuli received /i/ labels at least 80% of the time, and 36% received a clear majority of /e/ labels. In the Spanish identification task, 35% were labeled /i/ at least 80% of the time, and 50% were labeled /e/ at least 80% of the time. In each case the remainder of the stimuli are inside the uncertainty region. In other words, for the English stimuli, the perceptual boundary between /i/ and /e/ is shifted towards /e/; whereas, for the Spanish stimuli, this perceptual boundary is shifted towards /i/. Thus, in neither of these two languages is the perceptual boundary located in a position which results in an equal di-
vision of the stimuli into the two phonemic categories; rather, the location of the perceptual boundary is shifted such that in each language the boundary region is centered around an \( F_1 \) value of between 400 and 475 mels (320 and 390 Hz). This relatively consistent location of the /\( \text{i} \}/-/\( \text{e} \)/ boundary across the two languages, despite the differences between the two general vowel systems, suggests that there may be a natural, auditory-perceptual boundary in this region of the perceptual space. Specifically, we observe that there is a stable boundary in the high (low \( F_1 \)), front (high \( F_2 \)) corner of the vowel space with an orientation around an \( F_1 \) value of between 400 and 475 mels.

**Cross-Language Comparison of the /\( \text{u} \}/-/\( \text{o} \)/ Boundaries**

Figure 7 shows the English (fig. 7a) and Spanish (fig. 7b) /\( \text{u} \}/-/\( \text{o} \)/ stimulus groups with the regions of uncertainty 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 /\( \text{u} \)/ category. Thus, in the comparison of the /\( \text{u} \}/-/\( \text{o} \)/ 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.

In both cases we find that the /\( \text{u} \}/-/\( \text{o} \)/ boundary region is primarily in the \( F_1 \) dimension; and, that it is centered around an \( F_1 \) value of between 450 and 500 mels (366 and 414 Hz). Thus, we find that there is consistency across the two languages regarding the location of the /\( \text{u} \}/-/\( \text{o} \)/ boundary, even though the general phonemic vowel systems are quite different. In English we find a clear effect of the neighboring /\( \text{u} \)/ category, which is absent from the Spanish vowel inventory; nevertheless, the English /\( \text{u} \}/-/\( \text{o} \)/ perceptual boundary region is located in approximately the same position as the Spanish /\( \text{u} \}/-/\( \text{o} \)/ perceptual boundary region.

Thus, for both the /\( \text{i} \}/-/\( \text{e} \)/ and /\( \text{u} \}/-/\( \text{o} \)/ contrasts the general orientation and location of the perceptual boundaries are relatively consistent across English and Spanish. This suggests that there is a universal factor in the de-
termination of these perceptual boundaries that is independent of the size and structure of the general vowel inventory of a given language. This result, in conjunction with the results of the production-perception comparisons within each language, suggests that there are both language-specific and universal factors that determine listeners' judgments of vowels. The language-specific factor is seen by the close production-perception link within each language; the universal factor is seen by the consistent orientation and location of the boundaries across languages. We now turn to the results of the experiment that involved the presentation of the Spanish stimuli to English listeners and the English stimuli to Spanish listeners to see if these general results persist even under the condition of non-native stimuli.

Experiment 2

Method

The stimuli and experimental procedures for experiment 2 were the same as those in experiment 1, except that in experiment 2 English listeners responded to the Spanish stimulus sets and Spanish listeners responded to the English stimulus sets. In this experiment both the English and the Spanish subjects were run in the Cornell Phonetics Laboratory, and in place of a recording of the English word 'rating', a short, high-pitched tone prompted the subjects to enter the rating after the identification label. Subjects remained naive as to the source of the stimuli, and the instructions remained identical across the two experiments. A group of 14 English listeners participated in this experiment: 7 participated in the /ɪ/–/ɪ/ condition and 7 in the /ɛ/–/ɛ/ condition. A group of 10 Spanish listeners participated in this experiment, with 5 in each of the two conditions. None of these subjects also participated in experiment 1. Within each language the native dialect of the listeners is the same as the dialect of the speakers on which the stimuli and acoustic categories shown in the figures are based. The English subjects had all spent all of their adult years in the Northeastern region of the United States, and were monolinguals. The Spanish subjects had all spent most of their adult years in the Madrid region of Spain, but were now living in Central New York. Thus, although the Spanish subjects were all native speakers of Madrid Spanish, they were also proficient in English. English subjects' ages ranged from 18 to 27, with a mean of 20.7 years. Spanish subjects' ages ranged from 25 to 32, with a mean of 24.9 years. All subjects were paid for their participation, and none reported any history of speech or hearing impairment.

English Subjects with Spanish /ɪ/–/ɛ/ Stimuli

Following the same representation scheme as all other /ɪ/–/ɛ/ plots, figure 8 shows the results of the English listeners' identification and ratings of the Spanish /ɪ/–/ɛ/ stimuli. In this case, the total number of responses to each stimulus is 25, and the ellipses correspond to the native English acoustic /ɪ/ and /ɛ/ categories as shown in figure 1a.

The major difference between the Spanish /ɪ/–/ɛ/ stimulus group and the English listeners' native /ɪ/ and /ɛ/ categories is in the F2 dimension: the English listeners' native categories are significantly higher in the F2 dimension than the Spanish /ɪ/ and /ɛ/ stimuli [Bradlow, in press]. In fact, the majority of stimuli generated around Spanish mean /ɛ/ are located outside of the area in the acoustic space that corresponds to English /ɛ/, and many of these stimuli lie in the region that corresponds to English /ɪ/. Consequently, in their labeling of the Spanish stimuli, the English listeners show a relatively high degree of inconsistency. This is clear in figure 8 from the broad extent of the uncertainty region, that is the region enclosed by the irregularly shaped box. The stimuli that are included in this box are those that fall in the region between mean English /ɪ/ and mean English /ɛ/, as well as those that are outside of the range of the English /ɛ/. This response pattern suggests that the English listeners identify these stimuli with
Fig. 8. English listeners' identification (a) and ratings (b) of the Spanish /i/-/e/ stimuli.

reference to their native acoustic vowel space as it is represented by the categories produced by native speakers. In general, the listeners employ a strategy that treats these stimuli as vowels in their native vowel space, and make no adjustment of their perceptual vowel space to match the category locations indicated by the stimuli. Thus, this comparison of the re-

---

Phonetics 1996;53: 55-85

Bradlow
sponses of the English and Spanish listeners to the same Spanish /ɪ/-/e/ stimuli shows an effect of language background that can be directly related to the acoustic characteristics of the listeners' native vowel system.

However, if we consider the English listeners' /ɪ/-/e/ boundary to be just that portion of the uncertainty region to the left of the line in figure 9, then we find that the universal boundary that obtained in experiment 1 persists in this case as well. The placement of this line is such that it marks the lower F1 boundary of the /ɪ/ category from the production data and corresponds to the upper F1 boundary of the /ɪ/-/e/ perceptual boundary region that we observed in the condition with English listeners and English stimuli. Specifically, this /ɪ/-/e/ boundary region is primarily in the F1 dimension and is centered around an F1 value of between 400 and 475 mels.

In general then, the results of the identification task with English subjects and Spanish synthetic /ɪ/-/e/ stimuli indicate that listeners identify these non-native stimuli with reference to their native vowel space. Furthermore, we find additional evidence that there is a natural boundary in this region of the acoustic vowel space.

In the rating plot, we observe that there is a concentration of high ratings at the outer edge of the group of stimuli that received a majority of /ɪ/ labels. In other words, the highest ratings were given to /ɪ/ stimuli with relatively low F1 values and relatively high F2 values. For the few stimuli that received a clear majority of /e/ labels, we also find that the higher ratings were given to those stimuli at the outer edge of the stimulus range. This general distribution of high ratings concurs with the general pattern of rating distribution that was observed in experiment 1, that is that the higher ratings are concentrated at the outer edge of the general vowel space implied by the stimuli. In this case, since the entire stimulus group is lower in F2 than the English listeners' native vowel space, the stimuli at the outer edge (that is, those with relatively high F2 values) are also the stimuli that correspond best with the listeners' native /ɪ/ and /e/ categories. Thus, the general pattern of rating distribution in this condition corresponds both with the proximity of the stimulus to the listeners' native vowel categories and to the outer edge of the entire vowel space implied by these stimuli.

Spanish Subjects with English /ɪ/-/e/ Stimuli

Figure 10 shows the Spanish listeners' identifications and ratings of the English /ɪ/-/e/ stimuli. In this case the total number of responses per stimulus is 25, and the ellipses correspond to the native Spanish /ɪ/ and /e/ acoustic categories as shown in figure 1b.

The major difference between these English stimuli and the Spanish native /ɪ/ and /e/ categories is that these stimuli are significantly higher in the F2 dimension. In fact, sev-
Fig. 10. Spanish listeners' identification (a) and ratings (b) of the English /i/-/e/ stimuli.

ceral of these English stimuli lie beyond the upper $F_2$ edge of the entire Spanish acoustic vowel space. Responses of the Spanish listeners to these stimuli show that the location of the region of uncertainty is shifted toward the /i/ end of the stimulus range. The location of this uncertainty region relative to the distribution of the stimuli is similar to the uncertainty
region location that we observe in the response of the Spanish listeners to the Spanish stimuli in figure 3. In that case, the region of uncertainty was also shifted towards the /l/ end of the stimulus range. Thus, when presented with a stimulus range that extends beyond the range of their native vowel space, it appears as though the Spanish listeners shifted their perceptual /l/ and /l/ categories to match the location of the /l/ and /l/ categories indicated by the stimuli.

This uncertainty region location is also consistent with the region between the native Spanish acoustic /l/ and /l/ categories, even though the extent of these acoustic categories does not overlap fully with the stimulus range. However, this uncertainty region location is not consistent with the English listeners’ location of the /l–l/ boundary between these stimuli. In the latter case, the boundary location is shifted slightly toward the /l/ end of the stimulus range, and it has a considerably broader range than the boundary region of the Spanish listeners. Thus, as in the case of the English listeners’ response to the Spanish /l–l/ stimuli, we find a language-specific effect in that the two groups of listeners with different language backgrounds show different response patterns to similar stimuli. Furthermore, we find a link between the listeners’ responses to non-native vowel stimuli and the structure of their native vowel spaces: whereas the English listeners always exhibit an effect of the neighboring /l/ category, the Spanish listeners always locate the /l–l/ boundary such that the perceptual /l/ and /l/ categories mirror the production categories by having a broader extent in the F1 dimension for /l/ than for /l/.

With respect to the natural boundary that we obtained in the other /l–l/ perceptual tests, we find that the general characteristics of the natural boundary persist in the case of Spanish listeners’ responses to the English /l–l/ stimuli. As expected, the Spanish listeners’ /l–l/ boundary between the English /l/ and /l/ stimuli is primarily an F1 boundary, and is centered around an F1 value of between 400 and 475 mels. In general then, we find that the structure of the listeners’ native vowel space and this natural boundary combine to determine the precise location of the /l–l/ boundary for a given stimulus range.

In the rating task with the Spanish listeners and the English stimuli, we again find that the higher ratings occur for the stimuli at the outer edge of the entire vowel space implied by these stimuli (fig. 10). In this case, however, we also find relatively low ratings for the stimuli with relatively low F1 values that are within the range of the Spanish listeners’ native /l/ and /l/ categories. This result suggests that, regardless of the locations of the listeners’ native vowel categories, there is a tendency for vowels at the extremes of the entire vowel space implied by the current stimulus range to get relatively high ratings.

**English Subjects with Spanish /l–l/ Stimuli**

Figure 11 shows the results of the English listeners’ identification and rating of the Spanish /l–l/ stimuli, with a total of 35 responses per stimulus. The ellipses correspond to the native English /l/, /l/, and /l/ acoustic categories (fig. 1c).

The relationship of the Spanish /l–l/ 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, 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 /l/ lie in the intersection of the English produced /l/ and
Fig. 11. English listeners' identification (a) and ratings (b) of the Spanish /u/-/o/ stimuli.

/lo/ categories, none are higher in F₂ than the upper F₂ bound of the English produced /lo/ category. In the condition with English listeners and English /u/-/o/ stimuli, we found that listeners included in the uncertainty region the stimuli that were generated around the mean /lo/, and that had F₂ values above the upper F₂ edge of the produced /lo/ category. In other

---

Phonetica 1996;53:55–85
Bradlow
words, in general, these were the stimuli that were within the /u/ range and outside the /o/ range.

In the present condition, we find that the general shape of the uncertainty region, as shown in figure 11, matches the general shape of the uncertainty region in the condition with English subjects and English /u/-/o/ stimuli (fig. 4). In both cases we see an effect of the neighboring /u/ category in the extent of the uncertainty region. However, unlike the stimuli in the condition with English listeners and English stimuli, none of the stimuli in this condition 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.

With respect to the /u/-/o/ perceptual boundary region that we observed in the /u/-/o/ condition with English subjects and English stimuli, these cross-language data indicate that the boundary persists under this cross-language condition. To see this effect, we consider the left half of the uncertainty region in figure 12 to correspond with the /u/-/o/ perceptual boundary region, and the right half to correspond to the stimuli that are within the /u/ region. As expected, we find that the /u/-/o/ perceptual boundary is primarily an F1 boundary centered around an F2 value of between 450 and 500 mels. Indeed, this corresponds to the natural /u/-/o/ boundary that we observed in experiment 1.

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 entire vowel space defined by this stimulus set. 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/ condition 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 gave relatively high
ranks to the stimuli that lie at the periphery of
the general vowel space implied by the stimuli.
Finally, these data provide further evi-
dence of vowel space to match non-native stimuli.

Fig. 13. Spanish listeners' identifications (a) and ratings (b) of the English /u/-/o/ stimuli.
Spanish Subjects with English /u/-/o/ Stims

Figure 13 shows the results of the identification and rating tasks for the Spanish listeners with the English /u/-/o/ stimuli. In this case the total number of responses per stimulus is 25, and the ellipses correspond to the native Spanish /u/ and /o/ acoustic categories (fig. 1d).

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 mean /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 figure 13 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.

With respect to the persistence of the natural, universal boundary that we find in the /u/-/o/ condition with Spanish subjects and Spanish stimuli, we find that in this case, as expected, the boundary is primarily an F1 boundary, which is centered around an F1 value of between 450 and 500 mels (300–414 Hz). Thus, in all of the /u/-/o/ conditions we find that the boundary is located in approximately the same region of the vowel space.

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 this cross-language experiment provide additional 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. The data also provide further evidence for the presence of cross-linguistically preferred category boundary locations in the /i/-/e/ and /u/-/o/ regions of the vowel space. Finally, these data reinforce the notion that listeners show clear preferences in their ratings of vowels that are at the extremes of the general vowel space implied by the stimuli they are presented with.

General Summary and Conclusions

The results of the within-language comparison of the acoustic and perceptual vowel spaces (experiment 1) show a close link between the production and perception of the /i/-/e/ and /u/-/o/ vowel contrasts in both English and Spanish. In all cases the data reveal that the location of the perceptual boundary regions between the vowel pairs coincides with the region in the acoustic space between the acoustic categories for these vowels. This re-
sult indicates an effect of language background on vowel perception, in that listeners identify vowel stimuli with reference to the acoustic characteristics of the vowel categories that speakers of the same dialect produce. The results of the cross-language experiment (experiment 2) provide further evidence for an effect of language background on vowel perception. In this experiment, there were clear differences between the response patterns of the English and Spanish listeners to both the Spanish and English stimuli. Furthermore, in each case, the different response patterns can be directly related to differences in the native vowel spaces of the listeners. When responding to the Spanish /i/-/e/ stimuli which lie within the bounds of the English listeners' native vowel space, English listeners treat the vowel stimuli as points in their native vowel space. Consequently, we observe a strong effect of the neighboring English /t/ category. Similarly, when responding to the English /al/-/o/ stimuli which lie within the bounds of the Spanish listeners' native vowel space, Spanish listeners treated the stimuli as points in their native vowel space and located the /al/-/o/ boundary between their native /al/ and /o/ acoustic categories. In contrast, when responding to stimuli which extend beyond the bounds of their native acoustic vowel space, listeners shift their perceptual vowel space to match the vowel category of the stimuli they are presented with. This effect is evidenced in the case where the Spanish listeners respond to the English /il/-/e/ stimuli which are higher in F2 that the upper F2 edge of the native Spanish acoustic vowel space, and where the English listeners respond to the Spanish /al/-/o/ stimuli which are lower in F2 than the lower F2 edge of the native English acoustic vowel space. In both of those cases, the listeners' response patterns can be directly related to the general arrangement of vowel categories in their native vowel space, and thus indicate that the listeners shifted their perceptual vowel space to match the vowel space suggested by the stimuli.

The results of the rating tasks indicate that there is a general tendency for listeners to give higher ratings to stimuli that are at the outer edge of the vowel space suggested by the stimulus range. Accordingly, listeners generally give higher goodness ratings to front vowels with relatively high F2 values, and to back vowels with relatively low F2 values. Furthermore, this effect holds true regardless of the absolute locations of the stimuli in the acoustic space. Thus, for example, the English listeners give high goodness ratings for the English back vowel stimuli with low F2 values relative to the English /al/-/o/ stimulus range, and they give low goodness ratings for the English back vowel stimuli with relatively high F2 values. However, when presented with the Spanish /al/-/o/ stimuli which are generally lower in F2 than the English /al/ and /o/ acoustic categories, the English listeners still give high goodness ratings to the stimuli with low F2 values and lower ratings to the stimuli with relatively high F2 values. Thus for the stimuli that lie in the region of overlap between the two stimulus sets, their goodness ratings vary according to the overall range of the stimulus set. This result suggests that listeners rate vowels relative to the general range of the stimuli they are presented with, rather than with reference to vowel categories that are defined in terms of absolute formant values. This result agrees with results reported by Lively [1993], who found that /i/ goodness ratings were highest at the high F2, low F1 edge of the stimulus range. This general result is also in accordance with the 'Hyperspace' effect found by Johnson et al. [1993] whereby subjects consistently matched 'extreme' versions of synthetic vowels to target vowels in English words.

These results address the issue of how vowel categories are defined in the perceptual domain. In work by Grieser and Kuhl [1989]
and Kuhl [1991] the notion of phonetic prototypes for vowels is proposed as an account of how individual vowel tokens are categorized into phonemically meaningful units. According to this account, phonetic vowel categories have internal structure and are defined in terms of a category prototype that incorporates all the features of the category. This category prototype acts as a 'perceptual magnet' to which incoming stimuli are either 'attracted' or 'repelled'. If the stimulus shares enough features with the prototype in question and thus belongs to the same category, then it will assimilate to the prototype and be perceptually indistinguishable from it. Thus, in this view of the perceptual structure of vowel categories, fixed phonetic prototypes serve as the 'standards' against which incoming stimuli are judged. However, as described above, the present data do not support this view; rather, they suggest that perceptual vowel categories are defined in a relative sense and are thus adjustable depending on the acoustic characteristics of the stimuli.

The present perceptual data also provide evidence of a universal aspect to vowel perception in that the orientation and location of the /a/-/o/ and /u/-/e/ boundaries across the two languages are relatively consistent. The effect of these language-independent boundaries is witnessed in the /a/-/e/ conditions by the fact that in neither language is the boundary centered around the mid-point between the mean /a/ and /e/. For English listeners responding to English stimuli, this boundary is shifted slightly towards the /e/ mean, and for Spanish listeners responding to Spanish stimuli the boundary is shifted slightly towards the /a/ mean. These shifts result in an /a/-/e/ boundary that is centered around an F1 value of 400 to 475 mels (320-390 Hz) in both languages. In the cross-language /a/-/e/ conditions this boundary persists, indicating that this is indeed a robust perceptual boundary. In the /a/-/o/ conditions we observe an equally robust, language-independent boundary. In both the within-language (experiment 1) and the cross-language (experiment 2) experiments with both English and Spanish listeners, the /a/-/o/ boundary is primarily an F1 boundary that is centered around an F1 value of between 450 and 500 mels (366-414 Hz). Thus, these data indicate a language-independent component to vowel perception in the form of a preferred boundary location at certain regions in the F1-by-F2 space.

The existence of universal perceptual boundary regions is consistent with the principle of Quanial Theory [Stevens, 1972, 1989] that posits a nonmonotonic relationship between certain changes in the acoustic domain and the perceptual response of the listener. Specifically, we observe that as the first formant frequency changes in a monotonic fashion, the phonemic identification of the vowel stimulus changes in a quantal fashion that is universally determined. Based on the spectral center of gravity effect demonstrated by Chrstovich and Lublinskaya [1979] we might expect that the specific /a/-/e/ and /u/-/o/ boundaries obtained in the present experiments can be directly related to the 3- to 3.5-bark critical distance found by Chrstovich and Lublinskaya [1979]. In particular, in the /a/-/e/ and /u/-/o/ conditions, we expect that as the F1-F0 bark values of the stimuli cross this critical distance, listeners switch their identification responses from the high to the non high vowel. In order to test this expectation, the F1-F0 bark difference levels were calculated for the stimuli that fall within the universal /a/-/e/ and /u/-/o/ boundary regions. For the /a/-/e/ boundary, the F1-F0 bark difference of the combined boundary region across all conditions (that is, for both English and Spanish stimuli and across all subject groups) ranges from 2.0 to 2.9 bark. For the /u/-/o/ boundary, the F1-F0 bark difference ranges from 2.5 to 3.2 bark.
These results differ in two ways from the expected results based on the spectral center of gravity effect demonstrated by Chistovich and Lublinskaya [1979]. First, these results indicate a different threshold for front and back vowels. Specifically, these results indicate that for front vowels, the F₁–F₀ difference level above which subjects switch their identification response from the high to the nonhigh vowel is lower than for back vowels. In other words, for front vowels subjects expect smaller distances between the spectral peaks than they do for back vowels; and thus, the critical distance between the two spectral peaks appears to vary across front and back vowels. Furthermore, the specific critical bark difference levels found in the present study are generally smaller than the 3- to 3.5-bark critical distance found by Chistovich and Lublinskaya [1979].

A possible source of this discrepancy is in the different bark difference dimensions used in the present analysis and in the work by Chistovich and Lublinskaya [1979]. In the present analysis of the /i/-/e/ and /u/-/o/ perceptual boundaries the F₁–F₀ bark difference dimension is the relevant dimension, whereas the work by Chistovich and Lublinskaya [1979] focuses on the F₂–F₁ dimension. Thus, it is possible that the precise value of the critical distance between two spectral peaks is not entirely independent of the frequency values of the two peaks, but rather that the critical distance can vary across different frequencies. In other words, the present results suggest that the notion of a constant critical distance may not be applicable across the entire frequency range, and thus may vary across the different bark difference dimensions. However, Stevens [1989] cites Stevens et al. [1969], Traumüller [1981], and Di Benedetto [1987] for perceptual data that indicate a boundary along the F₁–F₀ dimension of about 3.0–3.2 bark. In Traumüller's [1981] study of vowel height perception in eastern Central-Bavarian, he finds a change in response patterns for F₀ between 350 and 400 Hz which he attributes to an integration band with a width of 3 bark that operates along the F₁–F₀ dimensions as well along the dimensions involving higher partials [see also Fiomeke and Diehl, 1994; Fahey, 1994]. Evidently more work is needed in order to assign specific auditory-acoustic characteristics to the natural boundaries, nevertheless both the present data and the spectral center-of-gravity data suggest the presence of natural boundaries in certain regions of the vowel space.

The general conclusion that can be drawn from this perceptual study of the /i/-/e/ and /u/-/o/ contrasts in English and Spanish is that both language-specific factors, such as inventory size and structure, and universal factors, such as the presence of natural boundaries in certain regions of the acoustic-perceptual vowel space, interact in determining the response patterns of listeners from different language backgrounds. These data also suggest that perceptual vowel categories are defined in a relative sense, rather than in an absolute sense which makes reference to fixed phonetic characteristics. Finally, the stimuli in these experiments were highly “impoverished” since they varied only in F₁ and F₂ frequencies and did not incorporate all spectral and temporal properties that characterize natural vowels. Nevertheless, using an adaptation of Kuhl's prototype paradigm with such stimuli proved effective in revealing universal and languagespecific aspects of two common vowel contrasts.
Acknowledgements

The work presented in this article appeared as chapter 4 of the author’s 1993 doctoral dissertation. Many thanks to my dissertation committee members: Allard Jongman, John Kingston, Abigail Cohn, and Carol Krumhansl. I also thank Joan Sereno, Linda Polka, Randy Diehl and an anonymous reviewer for many helpful comments, and Scott Gargash and Luke Kavin for technical assistance in the Cornell Phonetics Lab. This work was supported in part by a Grant in Aid of Research from Sigma Xi, The Scientific Research Society.

References

Werker, J. F.; Polka, L.: The ontogeny and developmental significance of

Vowel Contrast Perception in English and Spanish

Phonetica 1996;53: 55–85