Perception of English /i/ and /ɪ/ by Japanese and Spanish Listeners: Longitudinal Results[1]

Geoffrey Stewart Morrison
Department of Linguistics, Simon Fraser University[2]

Flege's Speech Learning Model predicts that if an L2 learner perceives an L2 speech sound as similar to an L1 speech sound, the two sounds will be combined as a diaphone category, the properties of which will eventually be intermediate between the properties of the L1 and L2 sound. In contrast if the L2 sound is perceived as new, then a new category will be established with properties which may eventually match the properties of the L2 sound. Canadian English has two high front vowels: tense /i/ and lax /ɪ/ differing in spectral and duration properties. Japanese has two high front vowels: long /iː/ and short /i/ differing in duration only. English /i/ and /ɪ/ are expected to be perceived as similar to Japanese /iː/ and /i/, and Japanese learners of English are predicted to establish diaphone categories. Their identification of English /i/ and /ɪ/ is predicted to initially match their perception of Japanese /iː/ and /i/, but eventually be intermediate between the native norms for the L1 and L2 categories. Spanish has one high front vowel. Spanish learners of English are predicted to perceive English /ɪ/ as less similar to Spanish /i/ than English /i/, and are predicted to eventually establish a new /ɪ/ category. Their identification of English /i/ and /ɪ/ is predicted to initially be poor but eventually match that of English listeners. These predictions were tested using a multidimensional edited-speech continuum covering the English words /bɪt bit bɪd bid/. Properties which varied in the continuum included vowel spectral properties and vowel duration. A longitudinal study was conducted testing Japanese and Spanish speaking learners of English one month and six months after their arrival in Canada. Japanese listeners were found to have a primarily duration-based categorical boundary between English /i/ and /ɪ/ which did not change between the initial and final tests. Spanish listeners did not have a categorical identification pattern in the initial test, but they did establish duration-based or spectrally-based categorical boundaries by the time of the final test. Results were therefore consistent with the theoretical predictions.

1 Introduction

Flege's (1995) Speech Learning Model (SLM) predicts that if an L2 learner perceives an L2 speech sound as "similar" to an L1 speech sound, then the listener will establish and perceive the two sounds via a single diaphone category. Since instances of both the L1 and the L2 sound will contribute to the development of the diaphone category, the SLM predicts that the properties of the diaphone category will eventually be intermediate between the properties of the L1 and L2 sound. In contrast if the L2 sound is perceived as "new", then a new category will be established. A new category can be established if the listener perceives at least some of the differences between the L1 and L2 sounds, and the greater the perceived dissimilarity, the greater the chance that a new category will be established.

[1]This paper is based on part of my forthcoming MA thesis. My thanks to Dr. Murray J. Munro for comments on an earlier version of the work, and to the anonymous participants and model speaker. All correspondence should be addressed to geoff@japan.co.jp

[2]Now at Department of Linguistics, University of Ottawa

Since only instances of the L2 sound will contribute to this new category, its properties will eventually match the native norms for the L2 sound (unless the category is deflected away from existing L1 sounds). The present study investigates the differences between Japanese and Spanish listeners' longitudinal perception of English /i/ and /ɪ/. The differences between the two groups are consistent with differences due to the Japanese listeners perceiving both English vowels as similar to Japanese vowels, and the Spanish listeners perceiving one of the English vowels as similar to a Spanish vowel, and perceiving the other as new.

With respect to vowel quality, Japanese and Spanish speakers have similar five-vowel systems; however, Japanese but not Spanish has phonemic contrasts between long and short vowels: Spanish has one high front vowel /i/; and Japanese has two high front vowels, short /i/ and long /iː/. Japanese long and short vowels are traditionally described as having identical spectral properties (Akamatsu, 1997); however, Fitzgerald (1996) found that Japanese speakers produced long vowels that were more peripheral in the vowel space than short vowels. English has two high front vowels: lax /ɪ/ and tense /i/. In General Canadian English (and several other varieties of English) the tense vowel is more peripheral in the vowel space and, ceteris paribus, longer than the lax vowel.

The SLM predicts that L2 sounds will be perceived as similar or new sounds at a context sensitive allophonic level, rather than at a phonemic level. Across languages, vowels preceding voiced consonants tend to be longer than vowels preceding voiceless consonants (Chen, 1970; Kluender, Diehl, & Wright, 1988). In English, the vowel duration difference is greater than in other languages and may (possibly in combination with spectral cues in the vowel) be sufficient to cue the voicing distinction (Halle, Hughes, & Radley, 1957; Raphael, 1972; Fox & Terbeek, 1977; Hogan & Rozsypal, 1980; Flege, 1989; Crowther & Mann, 1992). Vowel duration contrasts due to post-vocalic consonant voicing in Japanese and Spanish are reported to be much smaller than in English (Crowther & Mann, 1992; Takahashi, 1987; Tsukada, 1999; Navarro Tomás, 1916). Morrison (2002a) found that Japanese listeners identified English /i/ preceding a voiceless consonant primarily as /ɪ/; whereas, English /i/ preceding a voiced consonant and English /ɪ/ preceding either a voiced or voiceless consonant were correctly identified at rates of over 90%. He developed the hypothesis that English /i/ is assimilated to Japanese long /iː/ before a voiced consonant but assimilated to Japanese short /i/ before a voiceless consonant; that English /ɪ/ is assimilated to Japanese short /i/ before both a voiced and voiceless consonant; that English vowels assimilated to Japanese long /iː/ are identified as English /i/, and that English vowels assimilated to Japanese short /i/ are identified as English /ɪ/. See Figure 1 for a graphical representation of this hypothesis. Consistent with this hypothesis, Morrison (in press) found that Japanese listeners identified members of an English /i/-/ɪ/ continuum according to vowel duration, and that they had a categorical boundary in the same location as their categorical boundary for a Japanese /iː/-/i/ continuum with matching duration properties.

Experience with English is predicted to have little effect on Japanese listeners' perception of English /i/ and /ɪ/. The assimilation pattern proposed by Morrison (2002a) results in two diaphone categories: English /i/ before a voiced consonant plus Japanese /iː/; and English /i/ before a voiceless consonant, plus English /ɪ/, plus Japanese /i/ (see Figure 2). Since English /i/ and /ɪ/ allophones are assimilated to Japanese /iː/ and /i/ according to Japanese vowel duration criteria, English vowel allophones contributing to a diaphone have duration properties within the range for the Japanese vowel contributing to that diaphone. Hence, all vowel allophones contributing to diaphone categories have the same duration properties as Japanese /iː/ or /i/, and the duration properties of the diaphone categories will continue to be identical to Japanese /iː/ and /i/. One diaphone consists of Japanese /iː/ and an allophone of English /i/; since these vowels have very similar spectral properties (Strange et al., 1998; Kewley-Port, Akahane-Yamada, & Aikawa, 1996) there will be no change in the spectral properties of this diaphone. The other diaphone consists of Japanese /i/, an allophone of English /i/, and English /ɪ/; although English /ɪ/ differs spectrally from Japanese /i/, movement of the properties of the diaphone category towards English /ɪ/ will be tempered by English /i/ which is spectrally very similar to Japanese /i/. Hence the spectral properties of this diaphone category is predicted to remain close to Japanese /i/ and move only slightly in towards English /ɪ/.

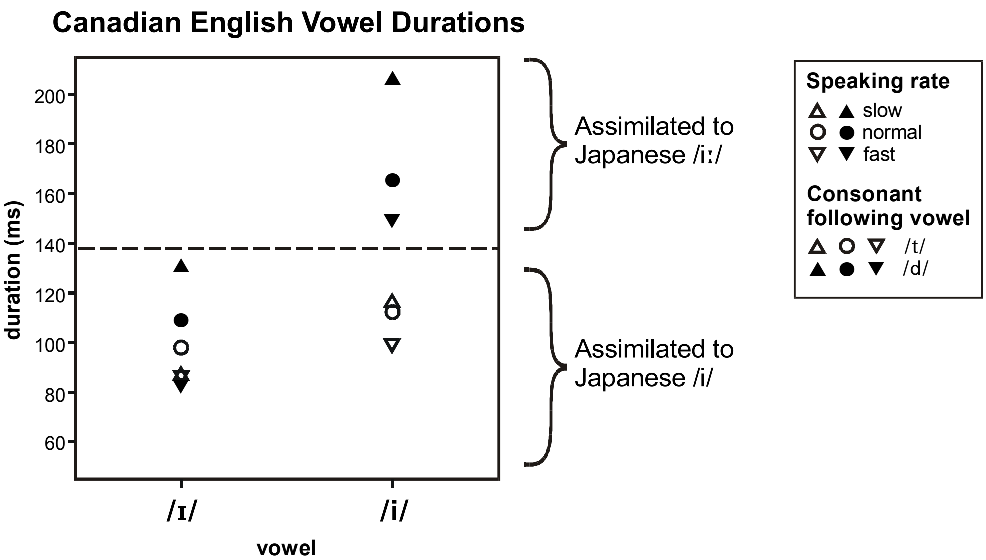

Figure 1 Durations of Canadian English high front vowels and hypothesised assimilation to Japanese long and short categories. The dashed line represents the theoretical location of the categorical boundary between Japanese /i/ and /iː/.

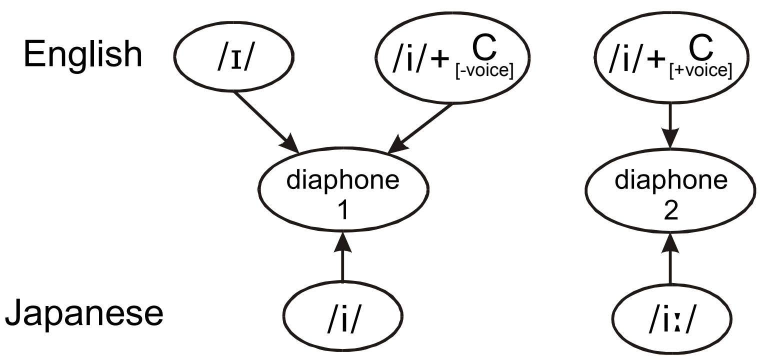

Figure 2 Hypothesised category formation for Japanese listeners.

Expressed in the terminology of Best's (1995) Perceptual Assimilation Model (PAM), Spanish listeners with limited exposure to English are predicted to assimilate English /i/ and /ɪ/ to Spanish /i/ via a category-goodness assimilation pattern, with English /i/ as a good match for Spanish /i/ and English /ɪ/ a poor match (see Figure 3). Spanish listeners' ability to distinguish English /i/ and /ɪ/ is predicted to be poor. This is consistent with the findings of Flege, Bohn, & Jang (1997), and Flege, Munro, & Fox (1994) for the perception of English vowels by Spanish listeners. Experience with English is predicted to improve Spanish listeners' ability to distinguish English /i/ and /ɪ/. Since English /i/ is very similar to Spanish /i/ (Flege, 1991; Flege, Munro, & Fox, 1994), these two vowels will form a diaphone category. Since English /ɪ/ is only a poor match for Spanish /i/, a new category will be established for English /ɪ/ (see Figure 3). The new category will develop first duration and then spectral properties that are close to the English norm for /ɪ/. The development from non-discrimination, to discrimination via duration properties, to discrimination via spectral properties was predicted by Escudero (2000, 2001a). The Spanish listeners are predicted to use duration differences to distinguish the vowels before using spectral differences. According to Bohn's (1995) Desensitisation Hypothesis, duration differences are easier to perceive than spectral differences and will be preferred if the listener's L1 has not exposed them to spectral contrasts in a particular part of the vowel space resulting in desensitisation to spectral differences.

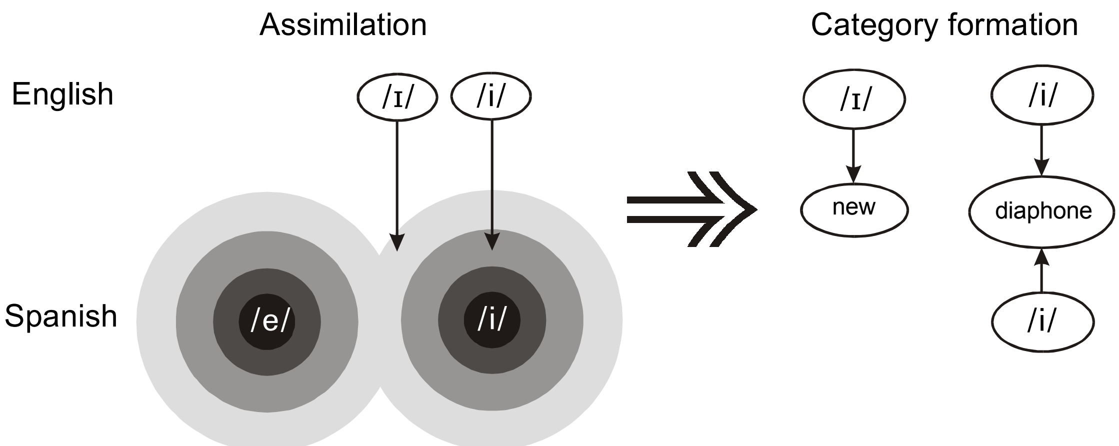

Figure 3 Hypothesised assimilation and category formation for Spanish listeners.

2 Methodology

2.1 Participants

The Canadian English participants (3 women, 4 men) had lived until age 16, and for most of the remainder of their lives, in an Anglophone region of Canada west of Quebec. The group had a mean age of 30 (range 25-38). The Japanese (4 women, 3 men) and Mexican Spanish (3 women, 2 men) participants were undergraduate students attending academic exchange programmes in Vancouver. Their programmes did not include any instruction in English pronunciation. They had lived in Vancouver for a period of less than 2 months prior to data collection. They had lived at least until age 16 in Japan/Mexico, and as children, had not been immersed in a language other than Japanese/Spanish. They had never lived outside Japan/Mexico for a continuous period of more than three months, and had not lived in an English speaking country during the year prior to their arrival in Canada. The Japanese group had a mean age of 20 (range 19-22). They had studied English in school for 7-9 years, starting at age 13 or 14. The Spanish group had a mean age of 20 (range 18-22). They had studied English in school for 12-15 years, starting at age 4-6.

2.2 Stimuli

2.2.1 Target segments and carrier sentences

The English target words ("bit, beat, bid, bead" /bɪt bit bɪd bid/) were chosen to exemplify the English vowels /i/ and /ɪ/ in a stressed position followed by both a voiced and voiceless plosive. The carrier sentences (which matched a Japanese carrier sentence used in another part of the wider study) was "What they're wearing are <u>beat</u> suits." Strange et al. (1998) found that the use of a carrier sentence, as compared to the presentation of target words in isolation, can have a significant impact on perception, such that long English vowels were assimilated to Japanese long vowels more often in a sentence condition than in an isolated word condition.

2.2.2 Acoustic properties of stimuli

Perceptual stimuli were constructed using edited natural speech based on the productions of a 34 year-old male monolingual English speaker who had lived most of his life in Vancouver. Recordings were made in a soundproofed room using a Sony MZS-R5ST Mini Disc recorder and a Sony ECM-MS907 microphone. The speaker read randomised lists of stimulus sentences at normal, slow, and fast speaking rates. The recordings were digitally transferred to computer, were analysed, and were used to create the stimuli using Praat computer software. Details of the procedure are given in Morrison (2002b).

A multidimensional continuum was created in which acoustic properties varied along the following dimensions: vowel spectra (5 points), vowel duration (7 points), plosive closure duration (5 points), and speaking rate (2 points). The dimensions are represented visually in the cuboids shown in Figure 4.

The F1, F2, and F3 values at the midpoints of the vowels are given in Table 1. The properties of spectral points1 and 5 were the mean spectral values from the model speaker's /i/ and /ɪ/. The five points were evenly spaced on the mel scale (which unlike the Hz scale has a linear relationship to human frequency perception). The mel - Hertz conversion formulae are:

$$f_{mel} = (1000 / \log 2) \log(f_{Hz} / 1000 + 1)$$

$$f_{Hz} = 1000 (10^{f_{mel} / (1000 / \log 2)} - 1)$$

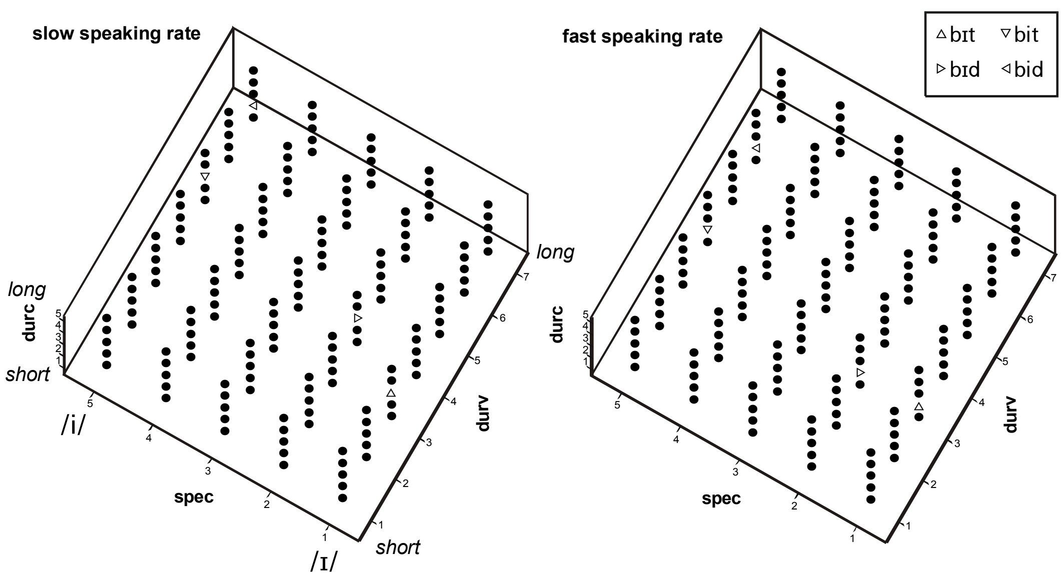

Figure 4 Cuboids representing the dimensions along which the perceptual stimuli vary. Vowel duration (durv) from front to back: 1 is shortest and 7 longest. Vowel spectral properties (spec) from right to left: 1 is most /ɪ/-like and 5 most /i/-like. Consonant closure duration (durc) from bottom to top: 1 is shortest and 5 longest. Speaking rate: left cube is slow and right cube is fast. Dots represent individual stimuli. The triangles represent the position of typical "bit, beat, bid, bead" produced by the model speaker.

The vowel and consonant durations of the stimuli are given in Table 2. The range of durations for the English stimuli was expanded to cover the same range as the durations for Japanese stimuli used in another part of the wider study. The duration points were evenly spaced on a logarithmic scale. This method of dividing the duration continuum was used by Escudero (2000, 2001a, 2001b) predicated on the concept that duration is perceived logarithmically (Turk, cited in Escudero, 2000). Conversions between stimulus point values (P) and time (t) in milliseconds can be made using the following formulae:

vowel duration $t = 60 \times 1.2^{(P-1)}$ $P = \log_{1.2}(t / 60) + 1$

consonant duration $t = 40 \times 1.45^{(P-1)}$ $P = \log_{1.45}(t / 40) + 1$

Based on mean values from the English and Japanese model speakers, sentence durations for the slow and fast speaking rates were 1.541 and 1.265 seconds respectively (excluding the duration of the target segments). Stimulus vowels had durations and spectral properties within 2.5% of the values given in Tables 2 and 3. Consonant closure consisted of periods of silence.

Table 1 Spectral values at the midpoints of the vowels in the perceptual stimuli.

Endpoint vowel	Dimension point	F1		F2		F3	
		Hz	mel	Hz	mel	Hz	mel
ɪ	1	410	496	1700	1433	2465	1793
	2	370	454	1834	1503	2569	1836
	3	330	412	1974	1572	2676	1878
	4	292	370	2121	1642	2786	1921
i	5	255	328	2275	1712	2900	1964

Table 2 Durations (in milliseconds) of the vowels and consonants in the perceptual stimuli.

Dimension point	1	2	3	4	5	6	7
Vowel duration	60	72	86	104	124	149	179
Consonant closure duration	40	58	84	122	177		

2.3 Data Collection

Canadian English participants were tested once. Japanese and Mexican-Spanish participants were tested approximately one month after their arrival in Canada and again approximately five months later. Participants were tested individually in a soundproofed room. They listened to the stimuli presented in random order via MEDS computer software over Optimus HP340 headphones, and responded to each English stimulus by clicking on one of five pictures on the computer screen. Four of the pictures represented the words “bit, beat, bid, bead.” The fifth picture, an “X,” was a null response which the participants were instructed to use if they heard a word other than one of the four target words. Pictures were used in order to avoid potential confusion from orthographic representations of the target words (The letter “i” in English “bit” and “bid” represents English /ɪ/ but in Spanish and romanised Japanese orthography “i” represents vowels which are closer to English /i/). Participants were trained to interpret the pictures before the perception test. The computer played a stimulus once and did not proceed until the participant had given a response. The inter-trial interval was 500 ms. The order of the pictures was assigned randomly for each trial. A subset of 24 stimuli was used as a warm-up before the full set of 350 stimuli, each of which was identified once.

3 Results and Discussion

3.1 Statistical analysis procedures

Discriminant analyses were applied to each groups’ data in order to determine which cues were used by the listeners, the relative weighing of the cues, whether perception of the two vowels was categorical, and, if so, the position of categorical boundaries between phonemes.

The procedure described below (based on information in SPSS, 1999; Brown & Wicker, 2000; and Stevens, 2002) was carried out using SPSS software. A discriminant analysis takes cases which are labelled as belonging to known groups and estimates the coefficients of predictor variables that best characterise the differences between the groups. The linear combination of variables is known as a discriminant function. A single discriminant function describes the difference between two groups. Given predictor variables x, y, ..., n a discriminant function value z is calculated as follows:

$$z = C_x x + C_y y + ... + C_n n + C_{constant} \quad (1)$$

where C_x, C_y, ..., C_n and $C_{constant}$ are the coefficients generated by the analysis

The analysis also calculates the within-group means which are the values of the discriminant function at the group centroids. Let z_α be the value of the discriminant function at the centroid of group α, and z_β be the value of the discriminant function at the centroid of group β. If, for given values of x, y, ..., n the value of z is closer to z_α than z_β, then the model derived from the discriminant analysis will classify a case with those x, y, ..., n values as a member of group α.

Of interest in the present study is the categorical boundary between the two groups, in the example above this may be defined as the line described by x, y, ..., n values such that the z value is exactly half way between z_α and z_β. For a model with two variables, this is the line[3] such that the probability that the case is a member of group α is 0.5 (and 0.5 for group β). To obtain an indication of the sharpness of the categorical boundary between the groups, lines can be calculated for x, and y combinations with 0.1 and 0.9 probability of being assigned to group α (0.9 and 0.1 for group β). A small distance between the 0.1 and 0.9 probability lines would indicate a sharp boundary (assuming that a linear model is appropriate), a large distance would indicate a fuzzy boundary. In the present study, listeners’ perception will be deemed categorical only if the distance between the 0.1 and 0.9 probability points/lines/planes is less than half the range of spectral and duration properties tested (i.e. less than 3 for *durv* [range 1 to 7], less than 2 for *spec* and *durc* [range 1 to 5]). The 0.1 and 0.9 probabilities were chosen as values which would clearly delineate robust identification of a given category whilst allowing for random errors due to non-linguistic factors such as fatigue. Posterior probabilities for assignment to group α can be calculated using the following formula:

[3]a point for a model with one variable, a plane for a model with three variables etc.

$$p = \frac{e^{-(z_\alpha - z)^2}}{e^{-(z_\alpha - z)^2} + e^{-(z_\beta - z)^2}} \tag{2}$$

where p is the probability of a case being assigned to group α

Formula 2 can be transformed into Formula 3 in order to obtain the z value for a given probability p.

$$z = \frac{\ln(\frac{1}{p} - 1) - z_\alpha^2 + z_\beta^2}{2(z_\beta - z_\alpha)} \tag{3}$$

Formula 1 can be transformed into Formula 4 so that for given values of $y, ..., n$ it is possible to calculate an x value:

$$x = (z - C_{constant} - C_y y - ... - C_n n) / C_x \tag{4}$$

An example of an application of the above process would be the derivation of a categorical boundary between English /ɪ/ and /i/ based primarily on spectral properties. The discriminant analysis is applied to the predictor variables *spec* (vowel spectral properties), *durv* (vowel duration), *durc* (consonant duration), and *spd* (speaking rate). *durv*, *durc*, and *spd* are systematically varied from 1-7, 1-5, and 1-2 respectively and substituted into Formula 5 to produce corresponding values for *spec*.

$$spec = (z - C_{constant} - C_{spd}{\cdot}spd - C_{durc}{\cdot}durc - C_{durv}{\cdot}durv) / C_{spec} \tag{5}$$

C_{spec}, C_{durv}, C_{durc}, C_{spd}, and $C_{constant}$ are coefficients generated by the discriminant analysis. z is calculated from Formula 3 in which z_α and z_β, the discriminant function values at the /ɪ/ and /i/ group centroids respectively, are generated by the discriminant analysis. The values 0.1, 0.5, and 0.9 are substituted for p (the probability of a case being assigned to the /ɪ/ group). The *spec*, *durv*, *durc*, and *spd* values are then graphed in the four dimensional space defined by the continuum dimensions (see Figure 4).

Other information supplied by the discriminant analysis includes Wilks's lambda, chi-square and classification results. Wilks's lambda (Λ) is the ratio of within-group variablility to total variability in the discriminator variables. Values close to 1 indicate that almost all the variability is due to within-group differences, values close to 0 indicate that almost all the variability is due to group differences. A chi-square (χ^2) test indicates whether the variability related to group differences is statistically significant. Classification results indicate the success of the model at assigning cases to the group to which they actually belong. High correct-classification scores would indicate that the boundary is sharp and linear. Low correct-classification scores may indicate that the boundary is fuzzy or that it is non-linear. Cross-validated correct-classification scores are derived by determining the classification of each case by a model including all cases except the case being tested (i.e. a jackknife, leave-one-out, or U-method proceedure).

The discriminant analysis also supplies standardised canonical discriminant function coefficients which indicate the relative weight of the contribution of each variable in the model. Standardised coefficients will compensate for the differences in the range of values of the predictor variables in the present study, eg. *durv* has values from 1 to 7 but *spd* only has values of 1 and 2.

In order to determine which variables were significant contributors to the listeners' vowel identification, variables were entered into the model in a stepwise manner. At each step the variable which maximised the *F* ratio between the groups was entered or removed from the model. Variables were only entered if they had an *F* ratio significance of .05 or less and removed if they had an *F* ratio significance of .10 or less. Stepwise methods may include too many variables, therefore the last variables entered into the model (corresponding to the variables with the lowest weighting in the results below) may in fact not be relevant factors in the listeners' perception.

3.2 Canadian English Participants

A discriminant analysis was conducted on the native English listeners' group results to determine their categorical boundary between English /i/ and /ɪ/. The variables entered into the model were spectral properties, vowel duration, and consonant duration, but not speaking rate. This suggests that speaking rate did not affect the English participants' vowel perception. The model was highly successful at categorising the English participants' responses, resulting in a cross-validated correct classification rate of 95.5%. $\Lambda = .234$ and $\chi^2(3, N = 2101) = 3045.365, p < .0005$.

Figure 4 shows the planes describing the categorical boundary derived from the discriminant analysis of the English participants' /i/ and /ɪ/ perception. The spectral values of the planes were calculated by substituting the coefficients and function values in Table 3 into Formulae 3 and 4. The 0.5 probability of /ɪ/ identification plane indicates the position of the categorical boundary. The distance between the 0.9 and 0.1 probability of /ɪ/ identification planes provides a measure of the sharpness of the categorical boundary. The planes are almost perpendicular to the *spec* axis indicating that the native English participants based their vowel identification overwhelmingly on the spectral properties of the vowel. The standardised canonical discriminant function coefficients, see Table 4, indicate that the relative weighting given by the English participants to the different variables were 84% for vowel spectral properties, 11% for vowel duration, and 5% for consonant duration. The spectral value at the centre of the planes, i.e. the mean spectral value of the boundary between /i/ and /ɪ/, was 3.07 (F1, F2, F3 = 415, 1568, 1875 mel; 333, 1964, 2668 Hz). The spectral distance between 0.1 and 0.9 probability of /ɪ/ identification was 0.44 (ΔF1, ΔF2, ΔF3 = 18, -30, -19 mel; 17, -62, -47 Hz[4]) indicating a sharp boundary.

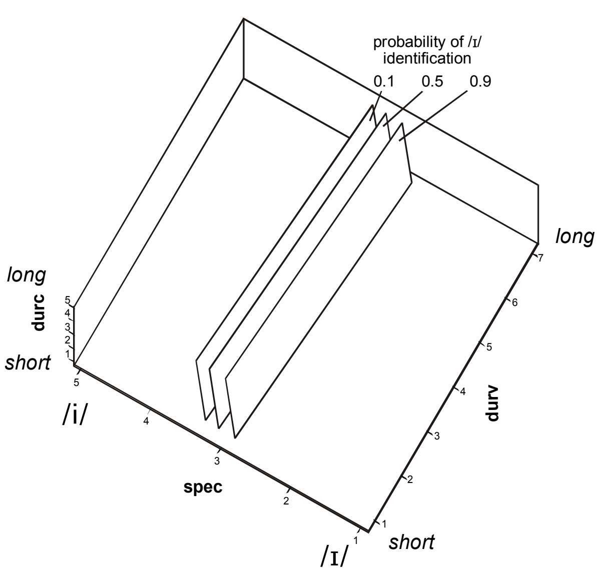

Figure 4 Planes describing the categorical boundary between /i/ and /ɪ/ derived from the linear discriminant function carried out on the identification responses of native English participants. Axes: *spec* = spectral properties, *durv* = vowel duration, *durc* = consonant duration.

[4]Hertz values were calculated at the centre of the planes.

Table 4 Discriminant function coefficients and centroid values from the discriminant analysis carried out on the identification of /i/ and /ɪ/ by native English listeners.

	Coefficients	
	Unstandardised	Standardised
spec	1.395	1.002
durv	0.064	0.128
durc	-0.046	-0.065
constant	-4.431	

Centroid Values	
ɪ	i
-1.84	1.775

3.3 Japanese Participants

3.3.1 Initial test

A discriminant analysis was conducted on the Japanese listeners' group results to determine their categorical boundary between English /i/ and /ɪ/. The results of one participant, JP09, were radically different to the results of the remainder of the Japanese participants (see Morrison, 2002b, for details) and were therefore not included in the general analysis of the Japanese participants' responses. The variables entered into the model were spectral properties, vowel duration, and consonant duration, but not speaking rate. This suggests that speaking rate did not affect the Japanese listeners' vowel perception, and is in accordance with the results of other studies (Strange et al., 1998; Guion et al., 2002, Ingram & Park, 1997) that found little effect for speaking rate on Japanese listeners' identification of English lax versus tense vowels. The discriminant-analysis model was moderately successful at categorising the responses, resulting in a cross-validated correct classification rate of 85.9% (see Table 4.3). $\Lambda = .468$ and $\chi^2(3, N = 1907) = 1446.863$, $p < .0005$.

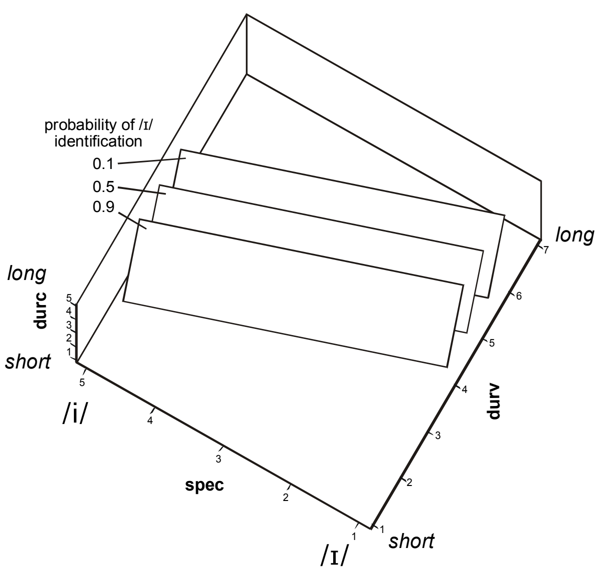

Figure 5 Planes describing the categorical boundary between English /i/ and /ɪ/ derived from the linear discriminant function carried out on the identification responses of Japanese participants in the initial test.

Figure 5 shows the planes describing the categorical boundary derived from the discriminant analysis of the Japanese participants' /i/ and /ɪ/ perception. The duration values of the planes were calculated from the coefficients and function values in Table 5. The orientation of the planes indicates that the Japanese participants based their vowel identification primarily on the duration properties of the vowel. The standardised canonical discriminant function coefficients, see Table 5, indicate that the relative weighting given by the Japanese participants to the different variables were 63% for vowel duration, 28% for vowel spectral properties, and 9% for consonant duration. The duration value at the centre of the planes, i.e. the mean duration of the boundary between /i/ and /ɪ/, was 4.22 (108 ms). The vowel duration distance between 0.1 and 0.9 probability of /ɪ/ identification was 1.49 (29 ms at the centre of the planes) indicating a relatively sharp boundary. The slopes due to the use of secondary cues resulted in a considerable range for the duration value of the boundary: The duration range for the 0.5 planes was 2.98 to 5.46 (86 to 135 ms), and for the 0.1 and 0.9 planes 2.24 to 6.20 (75 to 155 ms).

The Japanese participants' use of spectral properties may be a result of learning English. Although still relying primarily on transfer of L1 duration criteria, they may have realised that English /i/ and /ɪ/ differ spectrally and have begun to shift their cue weighting towards greater use of spectral properties. L1 experience, however, may have caused sensitivity to spectral properties: Fitzgerald (1996) found that Japanese speakers produced long vowels that were more peripheral in the vowel space than short vowels. He found that, with only F1 and F2 as independent variables, a discriminant analysis on all vowels correctly classified 77% of /iː/ and 69% of /i/ vowels. The existence of spectral differences leads to the possibility that spectral properties may in fact be a secondary cue differentiating long-short vowel pairs in Japanese. Further research is planned to determine whether this is the case.

Table 5 Discriminant function coefficients and centroid values from the discriminant analysis carried out on the identification of /i/and /ɪ/ by Japanese listeners.

	Coefficients	
	Unstandardised	Standardised
spec	0.324	0.436
durv	0.688	0.996
durc	-0.102	-0.143
constant	-3.444	

Centroid Values	
ɪ	i
-0.949	1.198

3.3.2 Final test

A discriminant analysis was conducted on the Japanese listeners' group results (excluding the results of JP09) to determine their categorical boundary between English /i/ and /ɪ/ in the final test. The variables entered into the model were vowel duration, spectral properties, consonant duration, and speaking rate). The discriminant-analysis model was moderately successful at categorising the responses resulting in a cross-validated correct classification rate of 87.6%; see Table 6. $\Lambda = .408$ and $\chi^2(4, N = 1983) = 1772.409$, $p < .0005$.

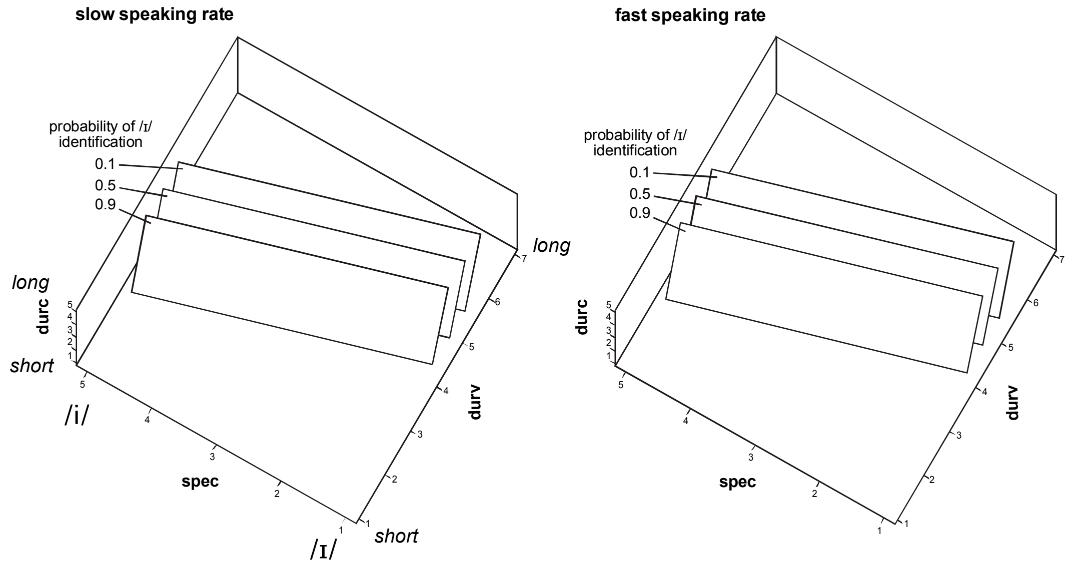

Figure 6 Planes describing the categorical boundary between /i/ and /ɪ/ derived from the linear discriminant function carried out on the identification responses of Japanese participants in the final test.

Figure 6 shows the planes describing the categorical boundary derived from the discriminant analysis of the Japanese participants' /i/ and /ɪ/ perception. The duration values of the planes were calculated from the coefficients and function values in Table 6. The slope of the planes indicates that the Japanese participants based their vowel identification primarily on the duration properties of the vowel. The standardised canonical discriminant function coefficients, see Table 6, indicate that the relative weighting given by the Japanese participants to the different variables were 62% for vowel duration, 26% for vowel spectral properties, 9% for consonant duration, and 4% for speaking rate. The duration value at the centre of the planes, i.e. the mean duration of the boundary between /i/ and /ɪ/, was 4.23 (108 ms). For a single speaking rate, the vowel duration distance between 0.1 and 0.9 probability of /ɪ/ identification was 1.61 (24 ms at the centre of the planes) indicating a relatively sharp boundary. Across speaking rates, the difference was 1.69 (27 ms). The slope due to the use of secondary cues resulted in a vowel duration range of 3.03 to 5.44 (87 to 135 ms) for the 0.5 line, and of 2.42 to 6.05 (78 to 151 ms) for the 0.1 and 0.9 planes.

A comparison between the Japanese participants' perception patterns for the initial and final English tests reveals little change other than the inclusion of speaking rate in the model for the final test but not the initial test. The contribution due to speaking rate, was, however, very small in the final test. The weighting of this cue was only 4%, and the weighting of other cues was very similar between tests (63, 28, and 9% for vowel duration, vowel spectral properties, and consonant duration respectively in the initial test, compared to 62, 26, and 9% in the final test). Measured to the nearest millisecond, the position of the mean value of the categorical boundary remained unchanged at 108 ms. The distance from the 0.1 to 0.9 probability boundaries was also similar, 29 ms in the initial test and 27 ms in the final test. Six months (compared to one month) living in an English speaking society had no effect on the Japanese listeners' perception of English /i/ and /ɪ/.

Table 6 Discriminant function coefficients and centroid values from the discriminant analysis carried out on the identification of /i/and /ɪ/ by Japanese listeners in the final test.

	Coefficients	
	Unstandardised	Standardised
spec	0.309	0.418
durv	0.741	1.009
durc	-0.106	-0.150
spd	0.122	0.061
constant	-3.766	

Centroid Values	
ɪ	i
-1.053	1.375

3.4 Spanish Participants

3.4.1 Initial test

A discriminant analysis was conducted on the Spanish listeners' group results to determine their categorical boundary between English /i/ and /ɪ/. The only variable entered into the model was vowel duration resulting in a cross-validated correct classification rate of 63.7%. This suggests that the Spanish participants based their identification of /i/ and /ɪ/ on vowel duration alone irrespective of other potential cues. The high Wilks's lambda indicates a high within-group variability: $\Lambda = .911$ and $\chi^2(1, N = 1577) = 147.407$, $p < .0005$. Visual inspection of graphs of individual participants' raw data (see example in Figure 7) did not reveal large contiguous subsets of stimuli that were clearly identified as one or other of the vowels. This suggests that the low correct-classification rate may in part be accounted for by a lack of consistency in intra-listener responses.[5]

The 0.1, 0.5, and 0.9 probability of /ɪ/ identification points were calculated from the coefficients and function values in Table 7. The Spanish participants' 0.5 probability of /ɪ/ identification was 3.95 (103 ms). The vowel durations for 0.1 and 0.9 probability of /ɪ/ identification were 0.59 (56 ms) and 7.30 (189 ms). These were greater than the range of vowel durations tested, indicating that there was no categorical boundary.[6] The Spanish participants' use of vowel duration in the perception of /i/ and /ɪ/ is shown graphically in Figure 8. The line representing the mean proportion of stimuli identified as /ɪ/ has a shallow slope, again indicating that the Spanish group's perception was not categorical. Separate discriminant analyses were also conducted on individual Spanish listeners' responses. No individual Spanish listener was found to have categorical perception of English /ɪ/ and /i/.

[5] Since each stimulus was identified only once, it was not possible to calculate an intra-speaker consistency rate. Given the large number of stimuli, it was not possible to obtain multiple responses without fatiguing the listeners.

[6] The criterion for categorical perception was that the distance from the 0.1 to 0.9 probability boundaries was less than half the range of stimulus values tested, in this case less than 3, half the vowel duration range (1 to 7).

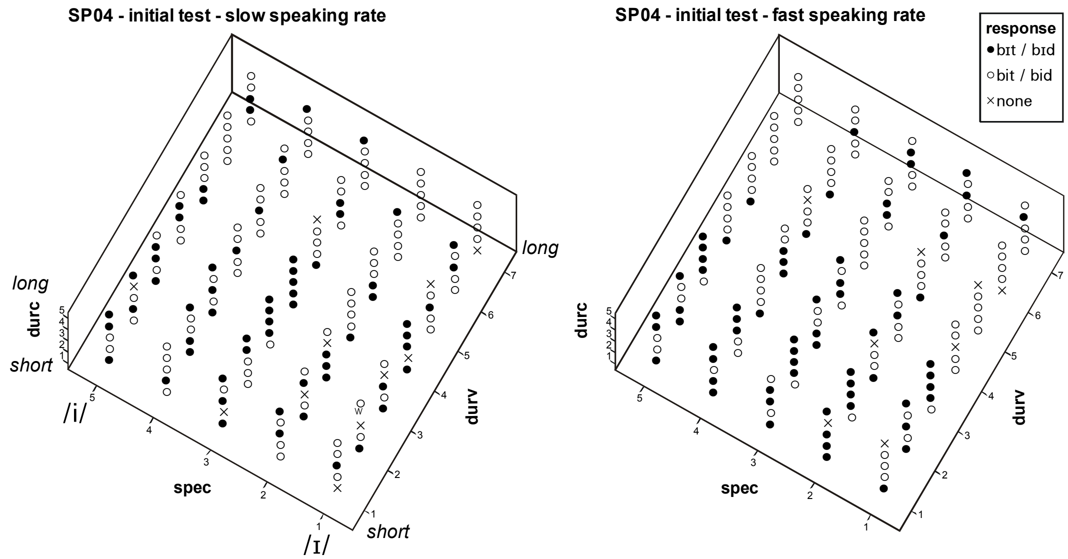

Figure 7 A typical English vowel identification pattern for a Spanish participant in the initial test.

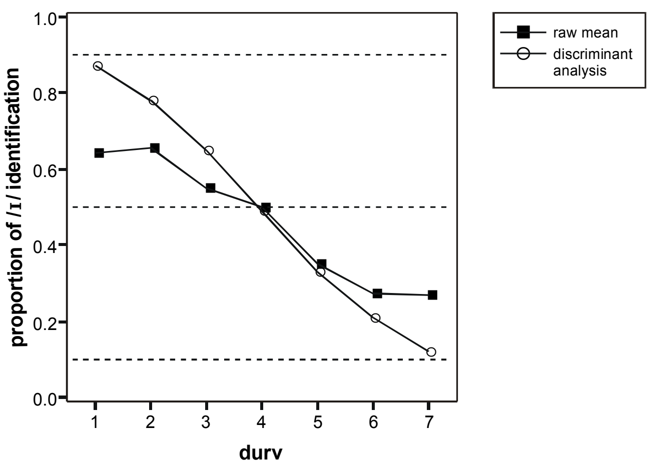

Figure 8 Proportion of stimuli identified as /ɪ/ by Spanish participants. Squares represent mean values for given durations. Circles represent probabilities derived from the discriminant analysis. Dashed lines represent 0.1, 0.5 and 0.9 probability thresholds.

Table 7 Discriminant function coefficients and centroid values from the discriminant analysis carried out on the identification of /i/ and /ɪ/ by Spanish listeners.

	Coefficients		Centroid Values	
	Unstandardised	Standardised	ɪ	i
durv	0.522	1	-0.336	0.291
constant	-2.079			

3.4.2 Final test

In the final test, four of the five Spanish participants showed categorical perception of English /i/ and /ɪ/. One had a categorical boundary based on vowel spectral properties and three had a categorical boundary based on vowel duration.

Participant SP02 had a perception pattern which was very similar to that of the native English participants. As can be seen in Figure 9, she had a categorical boundary between English /i/ and /ɪ/ based on the spectral properties of the stimuli. Variables entered into the discriminant analysis were vowel spectral properties and vowel duration. The standardised canonical discriminant function coefficients, see Table 8, indicate that the relative weighting given by the Japanese participants to the different variables were 83% for vowel spectral properties and 17% for vowel duration. The model had a moderate cross-validated correct classification rate of 78.3%. $\Lambda = .592$ and $\chi^2(2, N = 350) = 182.133$, $p < .0005$. Spectral values were calculated from the coefficients and function values in Table 8. The mean spectral value of the boundary between /i/ and /ɪ/, was 3.05 (F1, F2, F3 = 414, 1569, 1886 mel; 332, 1966, 2670 Hz), very close to the mean value obtained for the English participants of 3.07 (F1, F2, F3 = 415, 1567, 1875 mel; 333, 1964, 2668 Hz). The spectral distance between 0.1 and 0.9 probability

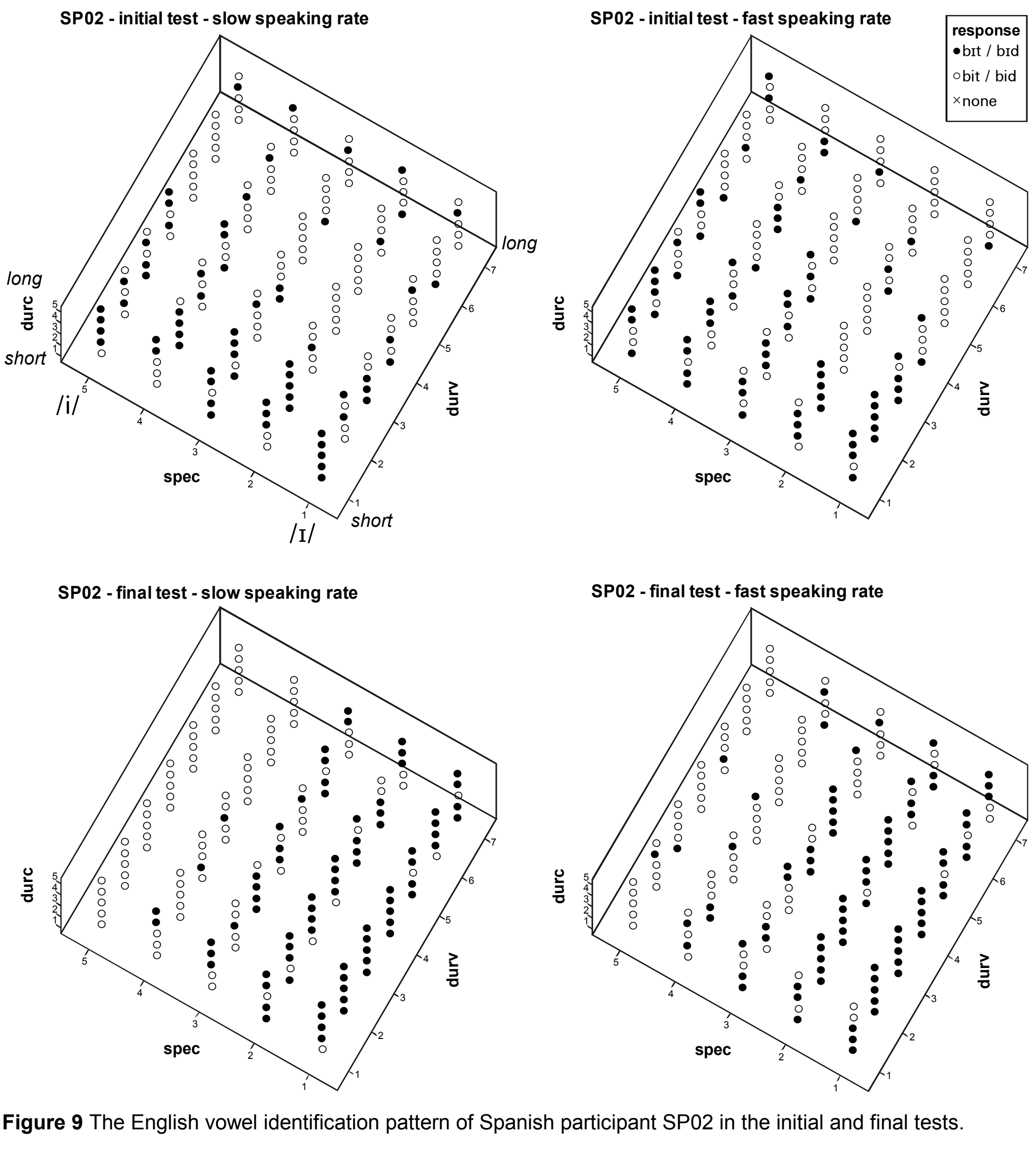

Figure 9 The English vowel identification pattern of Spanish participant SP02 in the initial and final tests.

of /ɪ/ identification was 1.46 (ΔF1, ΔF2, ΔF3 = 61, -102, -62 mel; 57, -209, -159 Hz), approximately three times greater than that obtained for the English participants. The 0.1 to 0.9 distance was, however, sufficiently small to be within the criterion for categorical perception (i.e it was less then 2, half the range of spectral values tested).

In summary, apart from the fact that SP02 had a fuzzier categorical boundary than that of the native English participants, she had developed native-like perception of English /i/ and /ɪ/.

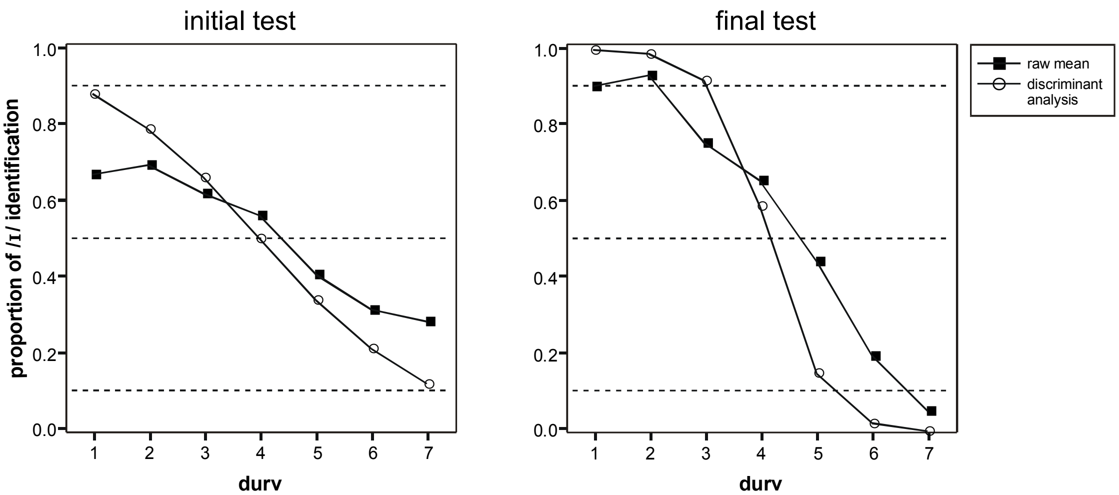

Figure 10 Proportion of stimuli identified as /ɪ/ by Spanish participants SP03, SP04, and SP06 in the initial test (left) and the final test (right). Squares represent mean values for given durations. Circles represent probabilities derived from the discriminant analysis. Dashed lines represent 0.1, 0.5 and 0.9 probability thresholds.

In the final test, participants SP03, SP04, and SP06 had a categorical boundary between English /i/ and /ɪ/ based on vowel duration. A discriminant analysis was conducted on their results to determine the location of their categorical boundary. The only variable entered into the model was vowel duration and the cross-validated correct classification rate was 79.4%. $\Lambda = .609$ and $\chi^2(1, N = 1037) = 512.362$, $p < .0005$. The 0.1, 0.5, and 0.9 probability of /ɪ/ identification points were calculated from the coefficients and function values in Table 9. The categorical boundary point between these participants' perception of /i/ and /ɪ/ was 4.20 (107 ms). The vowel duration distance between 0.1 and 0.9 probability of /ɪ/ identification was 2.13 (42 ms) indicating a relatively fuzzy boundary. The distance between the 0.1 and 0.9 probability points was much less in the final test than that calculated for the same participants in the initial test, the latter being 6.59 (133 ms) centred on a 0.5 probability point of 4.02 (104 ms). The model included only vowel duration as an independent variable. (Discriminant function coefficients and centroid values for this model are given in Table 10. $\Lambda = .908$ and $\chi^2(1, N = 1028) = 98.911$, $p < .0005$.) Their boundary in the final test was, therefore, much sharper than it had been in the initial test and met the criterion for categorical perception. Figure 10 shows a visual comparison of the three participants' non-categorical perception in the initial test and categorical perception in the final test. Figure 11 shows an example of one listener's change in perception from the initial to the final test.

In summary, although the categorical boundary was somewhat fuzzy, three Spanish participants had developed categorical perception of English /i/ and /ɪ/ based on vowel duration.

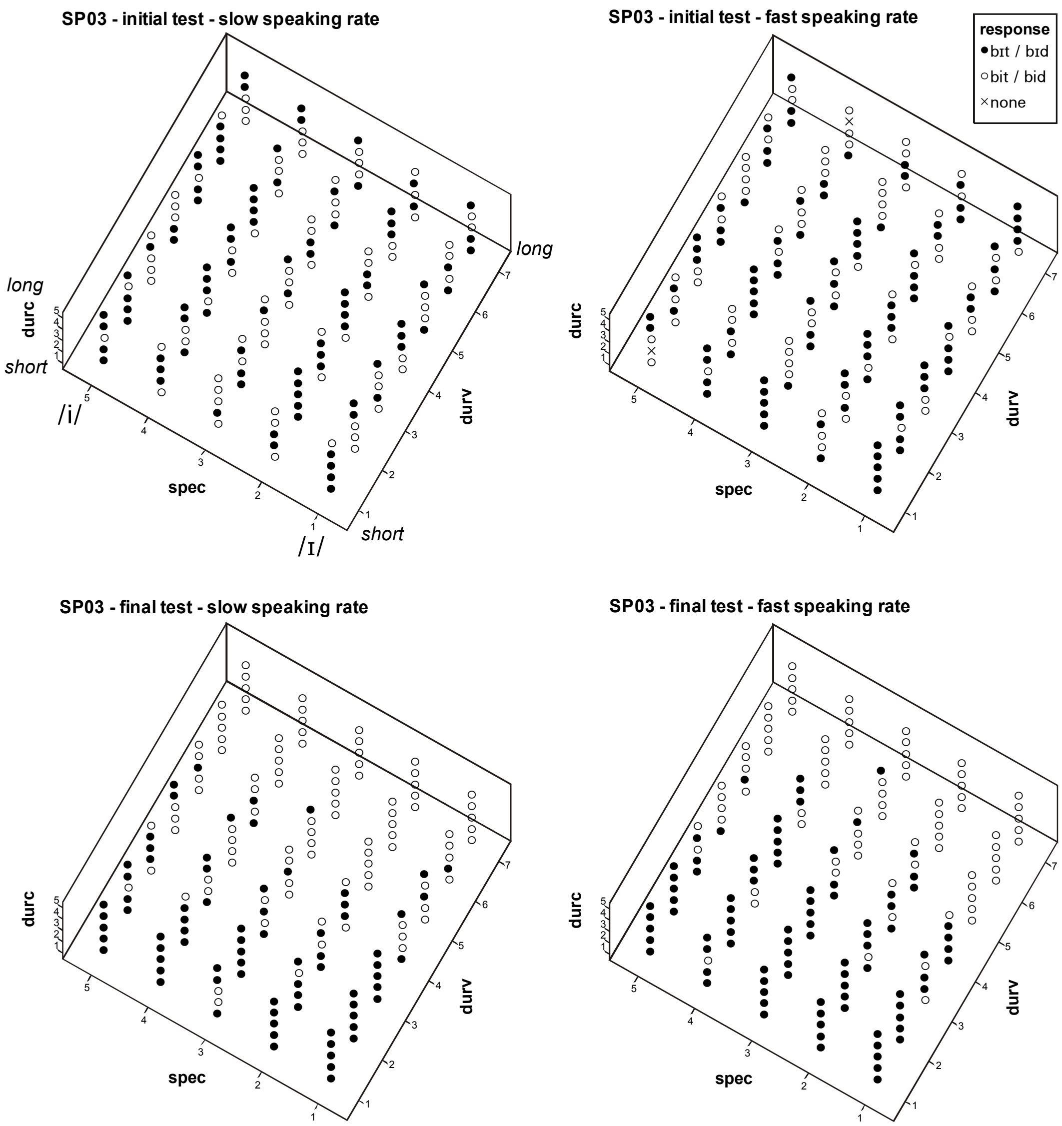

Figure 11 The English vowel identification pattern of Spanish participant SP03 in the initial and final English tests.

Table 9 Discriminant function coefficients and centroid values from the discriminant analysis carried out on the identification of /i/ and /ɪ/ by participant SP03, SP04, and SP06 in the final test.

	Coefficients		Centroid Values	
	Unstandardised	Standardised	ɪ	i
durv	0.64	1	-0.706	0.906
constant	-2.568			

Table 10 Discriminant function coefficients and centroid values from the discriminant analysis carried out on the identification of English /ɪ/ and /i/ by Spanish listeners SP03, SP04, and SP06 in the initial test.

	Coefficients		Centroid Values	
	Unstandardised	Standardised	ɪ	i
durv	0.524	1	-0.314	0.322
constant	-2.104			

In summary, as predicted, by the final English test the majority of Spanish participants had developed categorical perception of English /i/ and /ɪ/. Three identified the English vowels using vowel duration. One had a native-English-like boundary based on vowel spectral properties. Their categorical boundaries were somewhat fuzzy which may be indicative of an intermediate stage in development. It may be that with additional exposure to English the categorical boundary will become sharper.

4 General Discussion and Conclusions

Increased exposure to English did not change the Japanese listeners' perception of English /i/ and /ɪ/. This is consistent with the hypothesis that Japanese listeners assimilate allophones of English /i/ and /ɪ/ to Japanese /iː/ and /i/ resulting in diaphone categories. Although some spectral change had been predicted, the predicted change was slight, so the fact that no change was detected does not falsify the hypothesis. Also, as predicted, increased exposure to English changed the Spanish listeners' perception of English /i/ and /ɪ/. This is consistent with the hypothesis that Spanish listeners perceive English /i/ and /ɪ/ via a category-goodness assimilation and establish a new category for English /ɪ/.

Since Japanese and Spanish have similar five-vowel-quality systems but differ in that Japanese but not Spanish has a long-short vowel contrast, the difference in perception patterns between the Japanese and Spanish listeners suggests that the Japanese listeners' categorical perception of the English vowels was due to transfer of the Japanese long-short contrast. Since three of the five Spanish listeners developed a duration-based categorical boundary between English /i/ and /ɪ/ by the time of the final test, it could be argued that the difference in perceptual patterns between the Japanese and Spanish listeners in the initial test was due to the Japanese listeners being further along a developmental process than the Spanish listeners, i.e. that the Japanese listeners had earlier had non-categorical duration-based perception of English /i/ and /ɪ/ but had developed categorical perception by the time of the initial test. Given the difference in English experience between the two groups it seems highly unlikely that this could be the case. Such an explanation would be highly tenable if the Japanese listeners had more English experience than the Spanish listeners; however, the opposite was the case: The Spanish listeners had studied English for almost twice as many years as the Japanese listeners (12-15 compared to 7-9 years) and had started studying English at a much younger age (4-6 compared to 13-14). The weight of evidence therefore supports the hypothesis that Japanese listeners identify English /i/ and /ɪ/ via assimilation to Japanese /iː/ and /i/ according to the same long-short vowel duration criteria used to identify Japanese /iː/ and /i/.

References

Akamatsu, T. (1997). *Japanese phonetics: Theory and practice*. München & Newcastle: Lincom Europa.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171-204). Baltimore: York Press.

Brown, M. T., & Wicker, L. R. (2000). Discriminant analysis. In H. Tinsley & S. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modelling* (pp. 209-235). San Diego: Academic Press.

Chen, M. (1970). Vowel length variation as a function of the voicing of the following consonant environment. *Phonetica, 22*, 129-159.

Crowther, C., & Mann, V. (1992). Native language factors affecting use of vocalic cues to final consonant voicing in English. *Journal of the Acoustical Society of America, 92*, 711-722.

Escudero, P. (2000): *Developmental patterns in the adult L2 acquisition of new contrasts: The acoustic cue weighting in the perception of Scottish tense/lax vowels in Spanish speakers*. Unpublished M.Sc. thesis, University of Edinburgh.

Escudero, P. (2001a). The perception of English vowel contrasts: acoustic cue reliance in the development of new contrasts. In J. Leather & A. James (Eds.), *Proceedings of the 4th International Symposium on the Acquisition of Second-Language Speech, New Sounds 2000, University of Amsterdam*.

Escudero, P. (2001b). The role of the input in the development of L1 and L2 sound contrasts: language-specific cue weighting for vowels. In *Proceedings of the 25th Annual Boston University Conference on Language Development, November 2000*. Cascadilla.

Fitzgerald, B. H. (1996). *Acoustic analysis of Japanese vowels produced in multiple consonantal contexts*. Unpublished Master's thesis, University of South Florida.

Flege, J. E. (1989). Chinese subjects' perception of the word-final English /t/-/d/ contrast: Performance before and after training. *Journal of the Acoustical Society of America, 86*, 1684-1697.

Flege, J. E. (1991). The interlingual identification of Spanish and English vowels: Orthographic evidence. *Quarterly Journal of Experimental Psychology, 43*, 701-731.

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

Flege, J. E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics, 25*, 437-470.

Flege, J. E., Munro, M. J., & Fox, R. A. (1994). Auditory and categorical effects on cross-language vowel perception. *Journal of The Acoustical Society of America, 95*, 3623-3641.

Fox, R. A., & Terbeek, D. (1977). Dental flaps, vowel duration and rule ordering in American English. *Journal of Phonetics, 5*, 27-34.

Guion, S., Flege, J., Akahane-Yamada, R., & Pruitt, J. C. (2002). *Predicting Japanese adults' discrimination of English and Japanese vowels*. Manuscript in preparation.

Halle, M., Hughes, G., & Radley, J.-P. (1957). Acoustic properties of stop consonants. *Journal of the Acoustical Society of America, 29*, 107-116.

Hogan, J. T., & Rozsypal, A. J. (1980). Evaluation of vowel duration as a cue for the voicing distinction in the following word-final consonant. *Journal of the Acoustical Society of America, 67*, 1764-1771.

Ingram, J. C. L., & Park, S.-G. (1997). Cross-language vowel perception and production by Japanese and Korean learners of English. *Journal of Phonetics, 25*, 343-370.

Kewley-Port, D., Akahane-Yamada, R., & Aikawa, K. (1996). Intelligibility and acoustic correlates of Japanese accented English vowels. *Proceedings of ICSLP 96,*Philadelphia, PA. 450-453.

Kluender, K. R., Diel, R. L., & Wright, B. A. (1988). Vowel-length differences before voiced and voiceless consonants: An auditory explanation. *Journal of Phonetics, 16*, 153-169.

Morrison, G. S. (2002a). Perception of English /i/ and /ɪ/ by Japanese listeners. In S. Oh, N. Sawai, K. Shiobara, & R. Wojak (Eds.), *University of British Columbia Working Papers in Linguistics: Vol. 8.* Vancouver, BC: University of British Columbia, Department of Linguistics.

Morrison, G.S. (2002b). *Perception of Canadian English /i/ and /ɪ/ by Japanese and Mexican Spanish listners*. MA thesis in preparation.

Morrison, G. S. (in press). Japanese Listeners' Use of Duration Cues in the Identification of English High Front Vowels. *Proceedings of the 28th annual meeting of the Berkeley Linguistics Society*.

Navarro Tomás, T. (1916). Cantidad de las vocales accentuadas [The quantity of stressed vowels]. *Revista de Filología Española, 3*, 387-408.

Raphael, J. L. (1972). Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. *Journal of the Acoustical Society of America, 51*, 1296-1303.

SPSS Inc. (1999). SPSS base 10.0 applications guide [Computer software manual]. Chicago: Author.

Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences* (4th ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S. A., Nishi, K., & Jenkins, J. J. (1998). Perceptual assimilation of American English vowels by Japanese listeners. *Journal of Phonetics, 26*, 311-344.

Takahashi, Y. (1987). *Language universals and transfer effects: Acquisition of English vowel durations by Japanese learners*. Unpublished doctoral dissertation, Stanford University. [UMI: AAT 8801045]

Tsukada, K. (1999). *An acoustic phonetic analysis of Japanese-accented English*. Unpublished Doctoral dissertation, Macquarie University, Sydney, Australia.