Methodological Issues in L2 Perception Research and Vowel Spectral Cues in Spanish Listeners' Perception of Word-Final /t/ and /d/ in Spanish

Geoffrey Stewart Morrison
University of Alberta

1. Introduction

In this paper I would like to address some methodological issues in L2 speech perception research, and present some preliminary results from a study on L1 Spanish perception related to these methodological issues.

There is a consensus in our laboratory at the University of Alberta, that before one can understand L2 speech perception, it is necessary to have detailed models of native-listener perception of the L2 and the L1. A critical question in L2 speech perception research, is often whether an experimental result is due to L1 transfer or to some general L2 learning mechanism independent of the L1. Unless one has a detailed model of native L1 speech perception, it is impossible to answer this question. In order to avoid this problem a rigorous research paradigm would include the following:

1. Build a model of native perception of the L2 contrast of interest, *C2*, using the appropriate set of acoustic cues, *A2*.
2. Build a model of native L1 listeners' perception of the set of acoustic cues *A2* in terms of L1 categories, *C1′* (open category responses may be needed).
3. Build a model of native perception of any and all L1 contrasts, *C1*, that are related to *C2* (determined by *C1′* and by phonological theory), using the appropriate set of L1 acoustic cues, *A1*.
4. Build a model of L1 listeners' perception of the set of acoustic cues *A2* in terms of L2 categories, *C2′* (C2 plus any response categories suggested by *C1′*, by phonological theory, and by open category L2 responses).
5. Build a new model of native perception of the L2 contrast *C2* using the appropriate set of acoustic cues, *A2*, but including all the response categories of *C2′*.

Ideally, the models in 1 and 5 would be based on the perception of monolingual L2 listeners. The models in 2 and 3 would be based on the perception of monolingual L1 listeners, as well as on the bilingual listeners on whom the model in 4 is also based. Such a research paradigm may appear overly rigorous, and indeed one may legitimately abbreviate certain steps by relying on the results of earlier research, rather than conducting new experiments, and one may publish the results of different experiments in a series of reports; however, in the long term it is only by rigorous investigation that we will replace speculation, and vague theorising, with concrete hypotheses that stand up to repeatedly being tested.

Speech perception research is typically conducted using continua created from synthetic speech (or edited or resynthesised natural speech) in which the only acoustic properties that vary over the continua are those that are of theoretical interest. Typically participants must respond in a forced-choice manner, selecting from the set of possible response items provided by the experimental design. For example, the following describes a typical experimental design: Listeners are presented with a two-dimensional continuum, such as that in Figure 1, in which one dimension consists of differences in vowel duration, and the other dimension consists of differences in vowel spectral properties. At one corner of the continuum is a stimulus with typical production values for the vowel /ɪ/ in the English

 Selected Proceedings of the 2nd Conference on Laboratory Approaches to Spanish Phonetics and Phonology, ed. Manuel Díaz-Campos, 35-47. Somerville, MA: Cascadilla Proceedings Project.

word *bit* /bɪt/ (the stimulus with the shortest duration, and highest F1 / lowest F2), and at the opposite corner is a stimulus with typical production values for the vowel /i/ in the English word *beat* /bit/ (the stimulus with the longest duration, and lowest F1 / highest F2). Participants hear the stimuli in random order and must respond either *bit* or *beat* to each stimulus.

A potential problem arises because at another corner of the continuum is a stimulus with the spectral properties typical of productions of /ɪ/ in *bit* /bɪt/, but with the longer duration typical of productions of /i/ in *beat* /bit/, i.e., a stimulus with acoustic properties appropriate for the English word *bid* /bɪd/ (similar vowel durations for /bɪd/ and /bit/ have been observed in several English dialects, Peterson & Lehiste 1960, Elsendoorn 1985, Tsukada 1996, 1999:52, and Morrison 2002a:57). The listener hears the word *bid*, but is only allowed to answer either *bit* or *beat*. In this instance we may have missed important information about how the listener perceives the stimuli. If the extra category heard in an L2 perception experiment is due to the influence of the L1, then we may have missed important information about L2 speech perception.

Another potential problem arises if the listener does not hear a phonemic contrast (either L1 or L2) within the set of stimuli presented, but can hear some acoustic differences between the stimuli. Given two response options, the listener may decide to divide the stimuli between these options on the basis of whatever acoustic properties they can hear, even though they would not normally use these acoustic properties to distinguish speech sounds. The results could then be an artifact of the experimental design, and not reveal anything about what listeners normally do outside the laboratory. In order to avoid both these problems, experiments should be designed so that they include all response categories that would be appropriate for the stimuli, both those motivated by native perception of the L2 and by perception of the L2 in terms of the L1. Continua should also cover some contrast that the listeners can perceive. Such experiments cannot be designed without a detailed model of both L1 and native L2 speech perception. More complex crossed-response design and analysis are also needed, see for example Massaro & Oden (1980), Massaro & Cohen (1983), and Nearey (1990, 1997).

It was primarily on the basis of results of two-dimensional-continuum two-way-forced-choice experiments, that Bohn (1995) developed the *Desensitisation Hypothesis*. In one experiment (described in greater detail in Flege et al. 1997) English, Spanish, and Mandarin listeners identified members of a two-dimensional US English *bit* /bɪt/ – *beat* /bit/ continuum as either *bit* or *beat*. In distinguishing the stimuli, native English listeners relied almost exclusively on spectral properties, Mandarin listeners relied almost exclusively on duration properties, and Bohn (1995) claimed that Spanish listeners relied predominantly on duration properties.[1] Of key import was the fact that duration is not used to cue phonemic contrasts in either Mandarin or Spanish, and that Mandarin and Spanish each only have one vowel in the F1–F2 region where English has /i/ and /ɪ/. On the basis of this and similar experiments Bohn (1995) proposed the *Desensitisation Hypothesis*: "a perceptual principle that listeners apply independently of their native language background. This principle states that whenever spectral differences are insufficient to differentiate vowel contrasts *because previous linguistic experience did not sensitize listeners to these spectral differences*, duration differences will be used to differentiate the non-native vowel contrast" (Bohn 1995:294–295), emphasis mine.

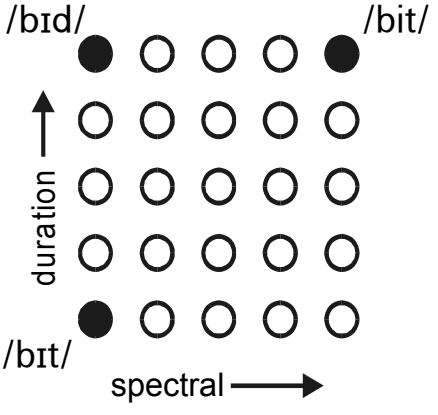

Figure 1. Typical two-dimensional continuum. Each dot represents a stimulus. The black dots have typical production values for English /bɪt/, /bit/, and /bɪd/.

[1] However, standard deviations and error bars reported in Bohn (1995) indicate that in fact there was no statistical difference between Spanish listeners' reliance on spectral and duration properties over the range of the stimuli. This was also the conclusion drawn in Flege et al. (1997).

What if Spanish listeners are in fact sensitised to spectral differences in the low F1 – high F2 part of the vowel space, but use these spectral differences to distinguish some other contrast in Spanish, e.g., a post vocalic consonant contrast? In that case, I could hypothesise that the Spanish listeners in Bohn's (1995) study avoided using spectral properties to distinguish English /i/ and /ɪ/, not because they were desensitised to the spectral differences, but because they had reserved those cues to distinguish a postvocalic consonant contrast. This possibility would not have been observable in Bohn's experiment, given that both stimuli had the same postvocalic consonant, /t/. Although there are other possible hypotheses as to why Spanish listeners prefer duration cues over spectral cues, in this paper I will focus on the one just outlined.

The experiments that I describe below illustrate that it may be necessary to know more about L1 perception in order to interpret the results of L2 perception research, and were motivated by an unusual result found in two L2 speech perception studies that I conducted (Morrison 2002c, and Morrison submitted). These studies investigated the perception of General Canadian English *bit* /bɪt/ – *beat* /bit/ – *bid* /bɪd/ – *bead* /bid/ continua by native English and L1 Spanish listeners.

Briefly, both studies used similar resynthesised-natural-speech continua covering production values for the four words in question. Continuum dimensions included vowel duration and spectral properties (the latter consisting of covarying F1 and F2, with formant tracks raised and lowered equally over the whole duration of the vowel).[2] Response categories included *bit – beat – bid – bead*, the first study also included an *other* response and the second study *bet – bed* responses.[3] The response options in the second experiment were an attempt to implement parts 4 and 5 of the rigorous research paradigm outlined above; the experiments presented in the current paper implement part of part 3. Participants in the first experiment were five Mexican university students whose perception was measured after having spent one month and six months in Canada. Participants in the second study were twenty native Spanish speakers from various Spanish-speaking countries who had lived in Canada for various lengths of time ranging from a few weeks to 18 years.

The unexpected result was that two of the five L1 Spanish participants in Morrison (2000c), and two of the twenty in Morrison (submitted), used vowel spectral properties to distinguish the phonemic voicing contrast on the postvocalic consonant, /t/–/d/. Although low F1 (especially at offset) has been found to correlate with postvocalic obstruent voicing in native English speakers' production and perception (e.g., Wolf 1978, Hillenbrand et al. 1984, Summers 1987, 1988, Fischer & Ohde 1990, Nearey 1990, 1997, Hillenbrand et al. 2001, Morrison submitted), the Spanish listeners' perception showed the opposite pattern: stimuli with high F1 (and low F2) were identified as ending in /d/. It is therefore unlikely that the Spanish listeners could have learnt this response pattern as the result of exposure to English.[4] Perhaps this unexpected response pattern is instead due to L1 transfer.

The research question addressed in the experiment reported below is therefore: Do native Spanish listeners use vowel spectral properties to identify a postvocalic /t/–/d/ contrast in Spanish, in a way that would be consistent with the finding for L2 English /t/–/d/ perception above?

Spanish has a /t/–/d/ contrast which, prevocalically or adjacent to certain other consonants, is typically realised as a prevoiced versus voiceless unaspirated plosive. Intervocalically it is typically a

---

[2] The first study also included consonant closure duration and carrier sentence speaking rate as continuum dimensions. In the first study the consonant closures were silent, and in the second they contained an ambiguous degree of voice bar.

[3] Álvarez González (1980), Flege (1991) and studies cited therein, Morrison (2002a), Imai et al. (2002), and Escudero & Boersma (2004) found that some English /ɪ/ tokens are identified by Spanish listeners as cases of English /ɛ/ or Spanish /e/.

[4] Moreton (2004) proposes that the relevant cue for phonemic voicelessness is peripherality in F1–F2 for both monophthongs and diphthongs, i.e., low F1 and high F2 would cue /t/, and high F1 and low F2 would cue /d/, the same pattern as obtained for the Spanish listeners. In the case of diphthongs, he presents evidence in support of his proposal; however, the proposal is contrary to published findings for monophthongs: F1 differences have often not been found to be significant for vowels with inherently low F1 (/i/, /u/), but the trend in the literature has always been for the F1 to be lower before voiced obstruents (Hillenbrand et al. 1984, Fischer & Ohde 1990, Nearey 1997, Hillenbrand et al. 2001).

contrast between a voiceless plosive and a voiced approximant. Lewis (2001) found substantial variability in the realisation of intervocalic /t/, e.g., it was often realised as partially voiced. He characterised the intervocalic /t/–/d/ contrast as a difference in degree of stricture and duration, rather than as a difference in voicing (Lewis 2001:35). Romero & Honorof (2004) characterised intervocalic /d/ as a deocclusivised or gesturally-reduced stop (see also Martínez Celdrán 1985, 2004).

It is normally assumed that there is not a word-final postvocalic /t/–/d/ contrast in Spanish; however, just because a contrast is assumed not to exist, does not mean that it cannot be perceived: In a recent paper, Warner et al. (2004) found that native Dutch listeners could perceive a difference between supposedly neutralised final /t/ and /d/ in Dutch, and reviewed similar findings for German, Polish, and Catalan. Spanish has a substantial number of native Spanish words that end orthographically in "d", and a small number of loan words that end orthographically in "t". Pronunciation of morphological alternations such as *pared–paredes* 'wall–walls' suggest that the word-final and intervocalic "d" represent the same phoneme, i.e., /paɾed/–/paɾedes/. The realisation of word-final postvocalic /d/ is variable: my informal impressionistic observations are that it can be pronounced as a rapid cessation of the final vowel, or as a voiced or voiceless fricative.[5] The latter realisations could be gesturally-reduced stops as in Romero & Honorof's (2004) analysis of intervocalic /d/. My informal impressionistic observations are that word-final postvocalic /t/ is never pronounced as a fricative, and may have a more pronounced release burst. I am not aware of any prior formal phonetic investigation into the pronunciation or perception of the postvocalic word-final /t/–/d/ contrast. Data on native Spanish speakers' pronunciation and perception of this contrast will be presented here, albeit only for a single phonetic context in a sentence reading protocol.

In the remainder of this paper, I will present two experiments investigating the production and perception of word-final postvocalic /t/ and /d/ in Spanish. Experiment 1 investigates whether native Spanish listeners are able to distinguish word-final postvocalic /t/ and /d/ produced by native Spanish speakers. Experiment 2 builds a normal a posteriori probability (NAPP) model (Assmann et al. 1982, Nearey & Hogan 1986, Nearey & Assmann 1986) to determine which acoustic properties the listeners may have used to distinguish the contrast, and the relative weighting of acoustic cues. The results suggest that native Spanish listeners may use F2 differences to distinguish word-final postvocalic Spanish /t/ and /d/, and that their identification of word-final postvocalic English /t/ and /d/ may therefore be due to L1 transfer.

2. Experiment 1 – Perceptibility

Experiment 1 investigated whether native Spanish speakers produced differences between words ending orthographically in "t" and "d" that were perceptible to native Spanish listeners.

2.1. Methodology

2.1.1. Participants

Speakers in the current study were the five listeners from Morrison (2002a,b,c) and twenty listeners from Morrison (submitted): 15 Mexicans, 3 Argentinians, 2 Colombians, 2 Spaniards, 1 Venezuelan, 1 Uruguayan, and 1 Chilean. The amount of time they had spent in Canada varied from 3 weeks to 18 years, age of arrival ranged from 8 to 52, and self estimates of percentage of daily communication conducted in English ranged from 0 to 95%. Although there were homogeneous subgroups, the speakers represented a wide variety of language backgrounds.

[5] See Harris (1969:37–45), Hualde (1989), and Quilis (1993:218–221) for other impressionistic descriptions of Spanish /t/ and /d/ including in word-final context. One of the reviewers noted that "in north-central Peninsular, *sed* may rhyme with *pez*, but not with *Bonet*, whereas both in Catalonia and in Paraguay, *sed* may rhyme with *Bonet*. For speakers from a number of different areas a voiced stop or approximant realization for /d/ is typical of 'reading style' but not of conversational speech."

Listeners were a Mexican (age 27, 1.5 years in Canada) and two Chileans (ages 40 and 53, 5 and 11 years in Canada). A small number of listeners were recruited in order to run a preliminary study, with a view to producing a more efficient experiment that could be presented to a larger group of listeners.

2.1.2. Procedures

The speakers were recorded reading randomised lists of the Spanish carrier sentence *Lo que llevan son ____ suyos* 'What they're wearing are ____ of theirs' in which the blank was filled with one of several nonsense words, or the words *bit* or *bid*. Each word appeared four times in the reading lists, in random order but never with the same word in succession, and the lists were padded with additional initial and final sentences in order to avoid initial and final intonation effects. *Bit* and *bid* were chosen because of their similarity to the words in the English *bit* /bɪt/ – *beat* /bit/ – *bid* /bɪd/ – *beat* /bid/ continua used in Morrison (2002c) and Morrison (submitted). Coincidentally, the Spanish words *bit* /bit/ (a loanword from English meaning 'binary digit), and *vid* /bid/ (a native Spanish word meaning 'vine')[6] are possibly the only word-final /t/–/d/ minimal pair in Spanish.[7]

The sentences were digitally recorded at a sampling rate of 44.1 kHz using a Sony ECM-MS907 microphone and a Sony MZS-R5ST Minidisc Recorder, and were transferred to computer via a Roland ED UA-30 soundcard.

Listeners heard the sentences containing *bit* and *bid*, and responded by pressing response buttons labelled *bit* and *bid*. Listeners controlled the rate at which sentences were presented and could listen to each sentence up to two times before responding. Sentences were blocked by speaker and each of the speaker's eight sentences was presented four times: four randomised blocks of eight sentences. The order in which speakers were presented was also randomised. To allow the listeners to get used to each speaker's voice, an additional randomly-selected sentence was added to the beginning of the block, responses to these sentences were not included in the analysis.

2.2. Results

An ANOVA was conducted to determine whether the native Spanish speakers produced differences between *bit* and *bid* that were perceptible to the native Spanish listeners. The word produced (*bit* or *bid*) was treated as a fixed factor, and both speakers (25 levels) and listeners (3 levels) were treated as random factors. The dependent variable was the rationalised arcsine transformation (Studebaker 1985) of the proportion of productions identified as *bit*, pooled across tokens and repetitions within speakers and listeners.[8]

The main effect for word was highly significant [$F(1, 18.198) = 27.142$, $p < .0001$] indicating that, overall, the listeners could distinguish, at above chance, whether the speakers had read *bit* or *bid*. Pooled across listeners, the proportion of tokens perceived as *bit* was 0.706 for words produced as *bit* and 0.305 for words produced as *bid*.

---

[6] "v" and "b" in modern Spanish are orthographic variants for the same phoneme /b/ and represent identical sets of allophones. They have not represented distinct phonemes since the 1500s (Penny 1991).

[7] The production data was collected as part of an experiment originally designed for other purposes. Hence there is no particular motivation for the choice of carrier sentence in this experiment. The words *bit* and *bid*, that the participants read, were mixed with other words which were all nonsense words in Spanish (that they happened to be real words, was, as stated above, coincidence). One of the reviewers pointed out that because of the lack of number and gender agreement between the target words and carrier sentence, the task may have been a test of Spanish speakers ability to read English words in Spanish sentences, rather than a purely L1 reading task.

[8] This resulted in a quasi-F test of the form: $\frac{MS_{word}}{(MS_{word\times speaker} + MS_{word\times listener} - MS_{word\times speaker\times listener})}$

The word by speaker interaction was also highly significant [$F(24, 48) = 13.485$, $p < .0001$]. The other interactions, word by listener, and speaker by listener, were significant: [$F(2, 48) = 3.377$, $p < .05$], [$F(48, 48) = 3.377$, $p < .005$]. Main effects for speaker and listener were not significant: [$F(24, 28.939) = .356$, $p > .05$], [$F(2, 4.010) = .227$, $p > .05$].

The significant word by speaker interaction was further examined: Figure 2 is a stacked bar graph for the proportion of *bit* and proportion of *bid* perceived correctly (i.e., perceived as the word read by the speaker) for each speaker. The proportion correct is pooled across all three listeners. Differences can be observed across speakers: For speaker number 7, the proportion perceived correctly approached 1 for both *bit* and *bid*, indicating that this speaker produced highly perceptible differences between the two words. For speaker number 25 the proportion perceived correctly was approximately 0.5 for both *bit* and *bid*, the expected proportion correct if there were no perceptible difference and listeners randomly guessed the identity of the word. In order to divide the speakers into two groups, one that produced a more-perceptible difference, and one that produced a less-perceptible or non-perceptible difference, an arbitrary criterion was selected: speakers were placed in the former group if their productions received correct identification rates of greater the 0.6 for each word. This resulted in 13 speakers in the more-perceptible group, and 12 speakers in the less-perceptible group. The mean correct-identification rate for speakers in the more-perceptible group was 0.789 for *bit* and 0.881 for *bid*, compared to 0.706 and 0.695 for all speakers, and 0.616 and 0.507 for the less-perceptible group.

Language background information was available for the speakers, including country or origin, age, length of time in Canada, age at which they began to learn English, number of years for which they had studies English, other languages spoken, and whether they had ever taken an English pronunciation class. Non-parametric Mann-Whitney U tests were conducted where appropriate: Only age differed significantly between the two groups [*Mann-Whitney U* = 37.5, $p < .05$], with the median age of the more-perceptible group being greater (34 versus 25). No trends were apparent in country of origin: Mexicans were evenly distributed between the perceptibility groups (7 in the more-perceptible group, 8 in the less-perceptible) and the remaining nationality groups were too small to establish a trend. Five of the participants in the more-perceptible group class, but only one in the less-perceptible group, reported having taken an English pronunciation course.

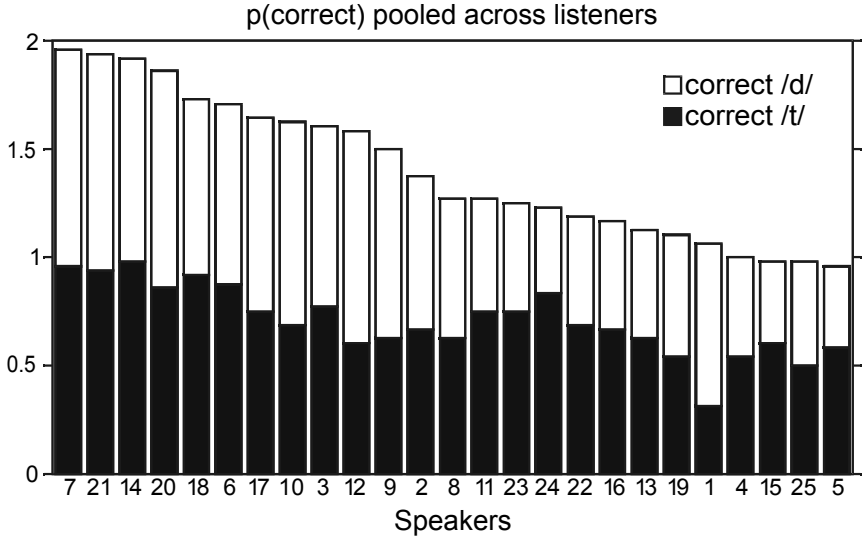

Figure 2. Stacked bar graph of the proportion (pooled across listeners) of each speaker's *bit* and *bid* productions perceived correctly (perceived as the word read by the speaker).

2.3. Discussion

The results of the ANOVA conducted on the perception data indicated that native Spanish listeners could correctly identify native Spanish speakers' *bit* and *bid* productions at rates significantly

above chance. Although the random factor design could be statistically extrapolated to the wider population, since only three listeners took part in this study, results should be treated as preliminary and all conclusions be regarded as tentative.

The significant word by speaker interaction indicated that some speakers produced differences between *bit* and *bid* that were more perceptually salient than those produced by other speakers. When speakers were divided into a group that produced more-perceptible *bit–bid* differences, and another group that produced less-perceptible / non-perceptible differences, it was found that speakers in the former group were older and that more of them had taken an English pronunciation course. One may speculate that older speakers tend to be more careful in their pronunciation (see for example, Labov's analysis of Cedergren's Panama City data, Labov 1994:94–97), and that speakers who chose to take an English pronunciation course may be more concerned with pronunciation in general.

3. Experiment 2 – Acoustic Analysis & NAPP Model

Experiment 2 investigated the acoustic properties that the native Spanish listeners may have used to distinguish *bit* and *bid*. First, acoustic measurements were made, a series of univariate ANOVA were conducted to determine which acoustic cues could potentially be used to distinguish the /t/–/d/ contrast. Significant cues were used in a linear discriminant analysis to build a normal a posterior probability (NAPP) model that could predict category membership on the basis of acoustic properties, and indicate the relative importance of each acoustic cue for prediction. The use of discriminant analysis to build a NAPP model here is similar to its application in a series of studies by Nearey and colleagues (Assmann, Nearey, & Hogan 1982, Nearey & Hogan 1986, Nearey & Assmann 1986, Nearey 1989, Andruski & Nearey 1992, Hillenbrand & Nearey 1999, Hillenbrand, Clark, & Nearey 2001).

3.1. Acoustic measurement procedures

The recordings of *bit* and *bid* produced by the speakers in Experiment 1 were downsampled to 22.05 kHz and their acoustic properties measured. The acoustic measurements made were the superset of those made by Lewis (2001) for intervocalic /p t k/, and those made by Warner et al. (2004) for Dutch word-final /t/ and /d/. These were:

vowel duration
- measured from the end of the release burst of /b/ to the drop in intensity and end of clear periodic format structure at the end of the vowel

consonant closure duration
- measured from the end of the vowel to the beginning of the /t/–/d/ release burst

duration of the consonant release burst
- this included any post release aspiration visible in the spectrogram

duration of the voice bar during the consonant closure
- measured from the end of the vowel to the first point at which the intensity fell 25 dB below the maximum intensity during the vowel, RMS intensity was measured over a 20 ms Hamming window

F1 and F2 measured at 25%, 75%, and (offset) 100% of the duration of the vowel
- measurement were made using the automatic tracking algorithm described in Nearey et al. (2002) (eight tracksets were displayed overlayed on spectrograms, and the best F1–F2 trackset was selected manually)

relative intensity of the consonant closure
- minimum RMS intensity during the consonant closure relative to the maximum intensity during the vowel, measured as the difference in dB using a 10 ms Hamming window

relative intensity of the consonant burst
- maximum peak intensity during the burst relative to the maximum intensity during the vowel, measured as the ratio of peak intensities

whether consonant is produced as a fricative
- a dichotomous variable measured on the basis of auditory perception, confirmed by spectrogram

3.2. Results
3.2.1. Univariate comparisons

For 19 of the 198[9] words measured, the consonant was produced as a fricative. These were all productions of *bid*, and were all perceived as *bid*. They were all produced by the group of more-perceptible speakers as determined in Experiment 1.

All the continuous acoustic measurements were used as dependent variables in a series of 17 univariate ANOVAs in which the word produced (bit or bid) was a fixed factor, and speaker was a random factor. F1 and F2 at 75% of vowel duration and at offset were analysed twice, once using their absolute values in Hertz and a second time as the change in Hertz relative to the previous measuring point, e.g., F2 at 75% minus F2 at 25%, and F2 at offset minus F2 at 75%. Voice bar duration was analysed as an absolute value and as a proportion of the duration of the consonant closure. Prior to analysis, all duration measurements and relative burst intensity were normalised using a natural log transform.

The main effect for *word* was significant at a nominal alpha level of .01 for the ANOVAs in which the dependent variables were F2 at 75% of the duration of the vowel, F2 at vowel offset, change in F2 from 75% to offset, vowel length, and burst length. All except F2 at 75% of the duration of the vowel were also significant at an alpha level of .05 following a Bonferroni correction for 17 tests. Of these, only burst duration also had a significant word by speaker interaction [$F(24,1\ 48) = 5.822$, $p < .0001$]. F values for significant main effects and marginal means are given in Table 1.

Table 1. *F* values and degrees of freedom from univariate ANOVA main effects significant at .01. (based on data from all speakers). *F* values and degrees of freedom corrected for missing data. Marginal means for dependent variables for all speakers, and for the group of speakers who produced more perceptible differences between *bit* and *bid*.

Dependent Variable	df	F	Marginal Means in original units			
			All speakers		More-perceptible speakers	
			bit	*bid*	*bit*	*bid*
F2 @ 75%	1, 24.094	8.929	2473 Hz	2442 Hz	2498 Hz	2452 Hz
F2 @ offset	1, 24.093	30.602	2377 Hz	2310 Hz	2393 Hz	2305 Hz
ΔF2 75% to offset	1, 24.139	23.085	-96 Hz	-131 Hz	-106 Hz	-147 Hz
vowel length	1, 24.129	39.472	96 ms	109 ms	99 ms	114 ms
burst length	1, 24.026	12.326	25 ms	13 ms	32 ms	12 ms

3.2.2. Multivariate NAPP model

In order to determine which of the acoustic measurements were the best predictors for word identity, the dependent variables from the ANOVAs above (those with a significant main effect for word) were included as predictor variables in a linear discriminant factor analysis (see Klecka 1980). Whether the consonant was produced as a fricative, was included as a binary variable. Speakers were also coded in a series of 24 dummy binary variables, so as to reduce the effect of inter-speaker variability on the other variables, e.g., to potentially separate formant variability due to differences in inter-speaker vocal tract length from formant variability due to differences between the words.[10] Two

[9] Data was missing for one *bit* and one *bid* production for one speaker.

[10] Such dummy coding is common in statistical techniques such as logistic regression, but not in linear discriminant analysis. This is in part due to the assumption that predictor variables be interval or ratio rather than categorical. However, including the dummy variables improved the models' classification rates (an indicator that the violations of assumptions are not very harmful, Klecka 1980:62), and had similar trends in the relative values of the standardised discriminant function coefficients of interest.

F2-at-offset coding options were tested: one with F2 at offset entered as an absolute value, and the other with the difference between F2 at offset and F2 at 75% of the duration of the vowel. A second set of discriminant analyses were conducted using only data from the speakers who produced more-perceptible differences between *bit* and *bid*, as determined in Experiment 1.

The values of the standardised canonical discriminant function coefficient are given in Table 2. The coefficients indicate that the F2 differences are the best predictors of whether the word was produced as *bit* or *bid*: The two largest coefficients in the absolute-F2-offset analysis for all speakers were *F2 at 75%* and *F2 at offset*. In the relative-F2-offset analysis, the correlation between the two F2 variables is removed, and *F2 at 75%* is now substantially larger than the next largest coefficient. The remaining coefficient values indicate that *vowel duration* is the next best predictor of whether the word was produced as *bit* or *bid*. Similar results are obtained for the more-perceptible speakers only, a notable difference being that the size of the *burst-length* and *whether-fricative* coefficients are larger relative to the F2 and vowel length coefficients.

Table 2. Standardised discriminant function coefficients for models based on all speakers, and on the speakers who produced more perceptible differences between *bit* and *bid*. *F2 at offset* coded as and absolute value (to left of slash) and as the difference between F2 at 75% and 100% of the duration of the vowel (to right of slash). Coefficients for speaker dummy coding variables not shown.

Predictor Variable	Coefficients	
	All speakers	More-perceptible speakers
F2 @ 75%	1.288 / 2.900	0.018 / 1.900
F2 @ offset	1.477 / 0.393	1.721 / 0.471
vowel length	-1.205	-0.923
burst length	0.529	0.651
whether fricative	-0.236	-0.255

The cross-validated by token correct-classification rates from the discriminant analysis on all speakers were 0.747 for *bit* and 0.778 for *bid*. These are similar to the overall correct identification rates obtained from the pooled listener responses in Experiment 1: 0.706 for *bit* and 0.695 for *bid*. The non-parametric Spearman rank order test was used to measure the correlation between the pooled-listeners' proportion-*bit* responses for each token, and the a posteriori probabilities from the discriminant analysis (a posteriori probabilities calculated using the replacement method, i.e., not via cross-validation). The correlation was highly significant [$p < 0.0001$] with a correlation coefficient of .656.

For the more-perceptible group the correct classification rates from the discriminant analysis were 0.863 for *bit* and 0.824 for *bit*, compared to 0.789 and 0.881 for the listeners. The Spearman rank correlation coefficient was .807 [$p < 0.0001$].[11]

3.3. Discussion

A series of univiariate ANOVAs and a follow-up discriminant analysis indicated that F2 values towards the end of the vowel were the best predictors of whether the word read by the speakers was *bit* or *bid*. The next best predictor was vowel duration. There were no significant word by speaker

[11] Since both the model and listeners had high correct identification rates, a high correlation would be expected because of good separation of categories, even if correlation within categories were poor. A more conservative test was conducted on the proportion perceived / predicted correctly (as intended by the speaker) for each token (effectively a combination of the two within-category correlations). Spearman's rho were .362 [$p < 0.0001$] for all speakers and .262 [$p < 0.01$] for the more-perceptible speakers. Correlations were lower, as expected, but still significant.

interactions for these variables, indicating that all speakers made F2 and vowel duration differences of similar magnitude.

The discriminant analysis indicated that burst length, and whether the consonant was a fricative, were poorer predictors; however, this may be due to the fact that such cues were only available in a subset of the speakers speech. Whenever a fricative was produced, the word was always identified as *bid*, making this potentially a very strong predictor. However, only seven speakers produced one or more fricatives (two speakers produced one fricative, two produced three, and three produced four). That fricatives are robust cues for *bid* identification, is supported by the fact that all the speakers who produced fricatives were included in the more-perceptible speaker group. Similarly, the significant word by speaker interaction in the burst duration ANOVA indicated that some speakers produced burst duration differences of greater magnitude than those produced by others. Examination of spectrograms suggests that most speakers produced relatively short bursts for both *bit* and *bid*, but that some speakers produced longer bursts for *bit*. In the discriminant analysis using only the more-perceptible speakers, the standardised coefficient values for burst-length and whether-fricative were relatively larger. This suggests that these cues may have been important factors in making these speakers more perceptible. Also note that although the *bit–bid* differences were greater for the more-perceptible speakers for all significant acoustic variables, the relative difference between the two groups was greatest for burst duration (a *bit*:*bid* burst duration ratio of 2.75 for the more-perceptible speakers, compared to 2 for all speakers).

The discriminant analysis is a statistical model based on measured differences in the acoustic properties of the speakers productions. It is not a direct model of the listeners' perception of these stimuli: It is possible that differences in measured acoustic properties exploited by the discriminant analysis could be ignored or given a different weighting by listeners. It is also possible that listeners could attend to some acoustic property that was not included in the measurements. However, comparing the discriminant analysis and the listeners, the overall correct identification rates were similar. There was also a moderately large correlation of .656 for the listeners' proportion of *bit* responses, and the model's a posteriori probabilities of *bit* responses, for the tokens produced by all speakers. Considering only the more-perceptible speakers, the correlation was a large .807. This suggests that the acoustic properties and weightings exploited by the listeners are not highly dissimilar from those exploited by the discriminant analysis. Future experiments are planned that will use synthetic speech and logistic regression to directly model listeners' perception patterns.

4. General Discussion & Conclusions

I began this paper by discussing methodological issues in L2 speech perception. The point which I wish to emphasise, is that, before one can interpret L2 speech perception results, it is important to have a full and accurate model of L1 speech perception. Indeed it would be advantageous to have the L1 model before conducting the L2 research, so as to arrive at an appropriate experimental design. I gave the example of an experiment in which Spanish listeners gave *bit* or *beat* (effectively /ɪ/ or /i/) responses to a continuum varying in spectral and duration dimensions; the sort of experiment that led Bohn (1995) to posit the Desensitisation Hypothesis. I raised the possibility that listeners had used duration differences to distinguish the vowels, not because they were desensitised to spectral differences, but because they had reserved spectral differences to distinguish a postvocalic obstruent voicing contrast. In Morrison (2002c) and Morrison (submitted) I found that when Spanish listeners were given *bit*, *beat*, *bid*, *bead* (effectively /ɪt/, /it/, /ɪd/, /id/) responses to the same sort of continua, a subset of them (four out of twenty five) used vowel spectral properties to identify the consonant; specifically, low F2 was correlated with voiced consonant identification. I hypothesised that they may use spectral cues to distinguish a putative Spanish word-final postvocalic stop voicing contrast, and may have transferred this cue use to their perception of the English stimuli. The whole proposition was predicated on their being a Spanish word-final postvocalic /t/–/d/ contrast, in which the voiced consonant is cued by low F2 in the preceding vowel. Although it is normally assumed that there is no word-final postvocalic stop voicing contrast in Spanish, I set out to examine whether native Spanish listeners could perceive whether Spanish words read by naive Spanish speakers had been *bit* /bit/ or *vid* /bid/ (with a spelling change to *bid*). Having determined that listeners did indeed identify the words

at rates significantly above chance (70% correctly identified overall, 83.5% for the half of the speakers who were identified as producing more-perceptible differences), I used discriminant analysis to build NAPP models predicting /t/–/d/ category membership on the basis of the acoustic properties of the speakers' productions. The NAPP models had moderately high, to high correlation with the listeners' responses (.656 and .807), indicating that they were good models of the listeners' perception. Standardised discriminant function coefficients indicated that F2 differences were the best predictors as to whether the word spoken had been *bit* or *bid*; low F2 towards the end of the vowel correlated with /d/ perception. The next best predictor was vowel duration. When the consonant was produced as a fricative it was perceived as a /d/, and when it was produced with a long release burst it was perceived as /t/; however, these cues were only produced in small subsets of tokens. The results were consistent with my hypothesis that Spanish speakers use low F2 in a vowel as a cue to phonemic voicing in a following word-final stop. Thus, it is possible that they transfer this cue when identifying English word-final postvocalic stops. The number of listeners in the present study was small, as was the number of repetitions of each token; It would therefore be wise at this stage to treat the findings with caution.

Even if Spanish word-final postvocalic /t/–/d/ differences are reliably produced by native Spanish speakers, one may question whether this contrast occurs frequently enough to allow robust perceptual representations to be established by native Spanish listeners.[12] I would like to speculate that the same sort of F2 differences seen in word-final position here, also occur prior to (and following) intervocalic /t/ and /d/ (see the acoustic data in Morrison 2002c). In this case, the word-final postvocalic context would be similar to one half of a intervocalic context, and infrequent word-final /t/ and /d/ could be identified by analogy with the frequently occurring intervocalic pattern. Future research is planned to investigate the perceptual cues to intervocalic /t/ and /d/.

Does this hypothesis, supported by the evidence presented here, account for why Spanish listeners use vowel duration and not vowel spectral properties to identify English /ɪ/ and /i/? First, it should be pointed out that the non-use of spectral cues may have been exaggerated. Some Spanish listeners make use of duration cues, some make use of spectral cues, and some use a combination of both, as can be seen in the results of Flege et al. (1997), Escudero & Boersma (2004), and Morrison (submitted). However, use of duration cues during the initial stages of English learning may be the norm for English dialects with spectral and duration production differences (see Morrison 2002a,b for the case of Canadian English). My hypothesis, that Spanish listeners use F2 differences as a cue to the word-final postvocalic /t/–/d/ contrast in Spanish, and transfer this to English, may well account for the subset of listeners in Morrison (2002c) and Morrison (submitted) who identified English /t/ and /d/ on the basis of F2 differences in the vowel. However, my hypothesis that Spanish listeners therefore reserve the F2 cue for consonant perception, and thus must rely for vowel identification on the only other available cue, vowel duration, is much more tenuous. My purpose in pursuing this hypothesis was not exclusively for its own sake, it was also to demonstrate the methodological principle that one should investigate L1 perception as thoroughly as possible in order to be able to understand the results of L2 perception research. Another hypothesis as to why Spanish listeners use duration rather than spectral properties to identify English /i/ and /ɪ/is presented by Escudero & Boersma (2004). They argue that it is easier to create a new category distinction in a previously unused (duration) dimension, than it is to split a category in an existing (spectral) dimension. Perhaps one of the most compelling, but linguistically uninteresting hypotheses for why Spanish speakers use duration to distinguish English /i/ and /ɪ/, is that, in classrooms all over the world, students are taught (misguidedly) that the difference between English /i/ and /ɪ/ is that one is long and the other short.

[12] The mean F2 offset differences for Spanish /bit/ and /bid/ in the present study were 67 Hz for all speakers, and 88 Hz for the more-perceptible speakers. Whether these exceed a just-noticeable difference should be taken into consideration: Kewley-Port (2001) found that trained native US English listeners could discriminate a mean F2 difference of 63 Hz in English /bɪd/ in a high-uncertainty condition. However, for untrained listeners listening to normal speech, her estimate of the minimal discriminable difference for a base formant value of 2400 Hz (close to the mean offset values for /it/ in the present study) was 111 Hz, somewhat higher than the mean offset differences in the present study.

Acknowledgements

This work was supported by a Social Sciences and Humanities Research Council of Canada (SSHRC) Doctoral Fellowship awarded to the Author and a SSHRC Standard Research Grant awarded to Terrance M. Nearey. My thanks to the volunteers who took part in this study, to Terrance M. Nearey, and to the anonymous reviewers for comments on earlier versions of this paper (any remaining defects are my own responsibility).

References

Álvarez González, Juan Antonio. 1980. *Vocalismo español y vocalismo inglés* [Spanish and English vowels]. Doctoral thesis, Universidad Computense de Madrid.

Andruski, Jean E. and Terrance M. Nearey. 1992. On the sufficiency of compound target specification of isolated vowels and vowels in /bVb/ syllables. *Journal of the Acoustical Society of America* 91.390–410.

Assmann, Peter F., Terrance M. Nearey, and John T. Hogan. 1982. Vowel identification: Orthographic, perceptual, and acoustic aspects. *Journal of the Acoustical Society of America* 71.975–989.

Bohn, Ocke-Schwen. 1995. Cross-language speech perception in adults: First language transfer doesn't tell it all. *Speech perception and linguistic experience: Issues in cross-language research*, ed. by Winifred Strange, 279–304. Timonium, MD: York Press.

Elsendoorn, Ben A. G. 1985. Production and perception of Dutch foreign vowel duration in English monosyllabic words. *Language and Speech* 28.231–254.

Escudero, Paola and Paul Boersma. 2004. Bridging the gap between L2 speech perception and phonological theory. *Studies in Second Language Acquisition* 26.551–585.

Fischer, Rebecca M. and Ralph N. Ohde. 1990. Spectral and duration properties of front vowels as cues to final stop-consonant voicing. *Journal of the Acoustical Society of America* 88.1250–1259.

Flege, James E. 1991. The interlingual identification of Spanish and English vowels: Orthographic evidence. *Quarterly Journal of Experimental Psychology* 43.701–731.

Flege, James E., Ocke-Schwen Bohn, and Sunyoung Jang. 1997. Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics* 25.437–470.

Harris, James W. 1969. *Spanish phonology*. Cambridge, MA: MIT Press.

Hillenbrand, James M., Michael J. Clark, and Terrance M. Nearey. 2001. Effects of consonant environment on vowel formant patterns. *Journal of the Acoustical Society of America* 109.748–763.

Hillenbrand, James, Dennis R. Ingrisano, Bruce L. Smith, and James E. Flege. 1984. Perception of the voiced-voiceless contrast in syllable-final stops. *Journal of the Acoustical Society of America* 76.18–26.

Hillenbrand, James M. and Terrance M. Nearey. 1999. Identification of resynthesized /hVd/ utterances: Effects of formant contour. *Journal of the Acoustical Society of America* 105.3509–3523.

Hualde, José Ignacio. 1989. Procesos consonánticos y estructuras geométricas en español [consonant precesses and geometric structures in Spanish]. *Lingüística (Revista de la Asociación de Lingüística y Filología de la América Latina)* 1.7–44. [Reprinted, 2000, in *Panorama de la fonología española actual*, ed. by Juana Gil, 395–431. Madrid: Arco Libros.]

Imai, Satomi, James E. Flege, and Ratree Wayland. 2002, June. *Perception of cross-language vowel differences: A longitudinal study of native Spanish learners of English*. Poster presented at the 143rd Meeting of the Acoustical Society of America, Pittsburgh, PA.

Kewley-Port, Diane. 2001. Vowel formant discrimination II: Effects of stimulus uncertainty, consonantal context, and training. *Journal of the Acoustical Society of America* 110.2141–2155.

Klecka, William R. 1980. *Discriminant analysis*. Beverly Hills, CA and London: Sage Publications.

Labov, William. 1994. *Principles of linguistic change*. Oxford, UK / Cambridge, MA: Blackwell.

Lewis, Anthony Murray. 2001. *Weakening of intervocalic /p, t, k/ in two Spanish dialects: Toward the quantification of lenition processes*. Unpublished Doctoral dissertation, University of Illinois at Urbana-Champaign.

Martínez Celdrán, Eugenio. 1985. Cantidad e intensidad en los sonidos obstruyentes del castellano: Hacia una caracterización acústica de los sonidos aproximantes [Quantity and intensity in Castilian Spanish obstruents: Towards an acoustic characterisation of approximants]. *Estudios de Fonética Experimental* 1.73–129.

Martínez Celdrán, Eugenio. 2004. Problems in the classification of approximants. *Journal of the International Phonetic Association* 34:201–210.

Massaro, Dominic W. and Michael M. Cohen. 1983. Phonological context in speech perception. *Perception & Psychophysics* 34.338–348.

Massaro, Dominic W. and Gregg C. Oden. 1980. Evaluation and integration of acoustic features in speech perception. *Journal of the Acoustical Society of America* 67.996–1013.

Moreton, Elliott. 2004. Realization of the English postvocalic [voice] contrast in F1 and F2. *Journal of Phonetics* 32.1–33.

Morrison, Geoffrey Stewart. 2002a. *Effects of L1 duration experience on Japanese and Spanish listeners' perception of English high front vowels*. Unpublished Masters thesis, Simon Fraser University, Burnaby, BC, Canada.

Morrison, Geoffrey Stewart. 2002b. Perception of English /i/ and /ɪ/ by Japanese and Spanish listeners: Longitudinal results. *Proceedings of the North West Linguistics Conference 2002*, ed. by Geoffrey Stewart Morrison and Les Zsoldos, 29–48. Burnaby, BC, Canada: Simon Fraser University Linguistics Graduate Student Association. [Available online: http://edocs.lib.sfu.ca/projects/NWLC2002]

Morrison, Geoffrey Stewart. 2002c. Spanish listeners' use of vowel spectral properties as cues to post-vocalic consonant voicing in English. *Collected Papers of the First Pan-American/Iberian Meeting on Acoustics* [CD-ROM]. Mexico, DF: Mexican Institute of Acoustics.

Morrison, Geoffrey Stewart. Submitted. *Development of L2 vowel perception and production: L1-Spanish speakers and the acquisition of the English /i/-/ɪ/ contrast.*

Nearey, Terrance M. 1989. Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustical Society of America* 85.2088–2113.

Nearey, Terrance M. 1990. The segment as a unit of speech perception. *Journal of Phonetics* 18.347–373.

Nearey, Terrance M. 1997. Speech perception as pattern recognition. *Journal of the Acoustical Society of America* 101.3241–3254.

Nearey, Terrance M. and Peter F. Assmann. 1986. Modeling the role of inherent spectral change in vowel identification. *Journal of the Acoustical Society of America* 80.1297–1308.

Nearey, Terrance M., Peter F. Assmann, and James M. Hillenbrand. 2002. Evaluation of a strategy for automatic formant tracking. Paper presented at the First Pan-American/Iberian Meeting on Acoustics, Cancún, Quintana Roo, Mexico, December 2002.

Nearey, Terrance M. and John T. Hogan. 1986. Phonological contrast in experimental phonetics: Relating distributions of production data to perceptual curves. *Experimental phonology*, ed. by John J. Ohala and Jeri J. Jaeger, 141–161. New York: Academic Press.

Penny, Ralph. 1991. *A history of the Spanish language*. Cambridge, UK: Cambridge University Press.

Peterson, Gordon E. and Ilse Lehiste. 1960. Duration of syllable nuclei in English. *Journal of the Acoustical Society of America* 32.693–703.

Quilis, Antonio. 1993. *Tratado de fonología y fonética españolas*. Madrid: Gredos.

Romero, Joaquín and Douglas N. Honorof. 2004. Spirantization revisited. Paper presented at the 2nd Conference on Laboratory Approaches to Spanish Phonetics & Phonology, Bloomington, IN, September 2004.

Studebaker, Gerald A. 1985. A "rationalized" arcsine transform. *Journal of Speech and Hearing Research* 28.455–462.

Summers, W. Van. 1987. Effects of stress and final-consonant voicing on vowel production: Articulatory and acoustic analyses. *Journal of the Acoustical Society of America* 82.847–863.

Summers, W. Van. 1988. F1 structure provides information for final-consonant voicing. *Journal of the Acoustical Society of America* 84.485–492.

Tsukada, Kimiko. 1996. Acoustic analysis of Japanese-accented vowels in English. *Proceedings of the 6th Australian International Conference on Speech Science and Technology*, ed. by P. McCormick and A. Russell, 373–378. Canberra: Australian Speech Science and Technology Association.

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

Warner, Natasha, Allard Jongman, Joan Sereno, and Rachèl Kemps. 2004. Incomplete neutralization and other sub-phonemic durational differences in production and perception: Evidence from Dutch. *Journal of Phonetics* 32.251–276.

Wolf, Catherine G. 1978. Voicing cues in English final stops. *Journal of Phonetics* 6.299–309.

Selected Proceedings of the 2nd Conference on Laboratory Approaches to Spanish Phonetics and Phonology

edited by Manuel Díaz-Campos

Cascadilla Proceedings Project Somerville, MA 2006

Copyright information

ISBN 1-57473-411-3 library binding

Ordering information

Orders for the library binding edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, e-mail: sales@cascadilla.com

Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Morrison, Geoffrey Stewart. 2006. Methodological Issues in L2 Perception Research and Vowel Spectral Cues in Spanish Listeners' Perception of Word-Final /t/ and /d/ in Spanish. In *Selected Proceedings of the 2nd Conference on Laboratory Approaches to Spanish Phonetics and Phonology*, ed. Manuel Díaz-Campos, 35-47. Somerville, MA: Cascadilla Proceedings Project.

or:

Morrison, Geoffrey Stewart. 2006. Methodological Issues in L2 Perception Research and Vowel Spectral Cues in Spanish Listeners' Perception of Word-Final /t/ and /d/ in Spanish. In *Selected Proceedings of the 2nd Conference on Laboratory Approaches to Spanish Phonetics and Phonology*, ed. Manuel Díaz-Campos, 35-47. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #1324.