Spanish listeners' use of vowel spectral properties as cues to post-vocalic consonant voicing in English

Geoffrey Stewart Morrison, University of Ottawa

1. Introduction

In English, phonemically voiced post-vocalic obstruents may be phonetically devoiced (i.e., lack glottal pulsing during the obstruent closure); however, since multiple cues signal the voicing contrast, the contrast is maintained and native English listeners are able to reliably identify underlyingly voiced obstruents. Keating (1985) argued that although physiological constraints (e.g., difficulty of maintaining sufficient airflow) may result in a tendency for utterance final obstruents to be phonetically devoiced and preceded by longer vowels, the occurrence of the same pattern in word final, but not utterance final, position is due to a phonological process. In perception tests, Halle, Hughes, & Radley (1957); Raphael (1972); Flege (1989); and Crowther & Mann (1992) found that vowel duration was a reliable cue to post-vocalic obstruent voicing: obstruents following longer vowels were perceived as voiced. Suen & Beddoes (1974) and Lisker (1957) claimed that consonant closure duration was a relevant cue, and Denes (1955), Port (1981), and Port & Dalby (1982) proposed that the relevant cue was the ratio of vowel to consonant duration. However, Crystal & House (1982) and Luce & Charles-Luce (1985) did not find either consonant closure duration or the vowel to consonant duration ratio to be reliable cues to consonant voicing. Wolf (1978); Revoile et al. (1982); Walsh & Parker (1983); Hillenbrand et al. (1984); and Fischer & Ohde (1990) found that the F1 offset frequency of the vowel was a cue to post-vocalic consonant voicing: vowels with low frequency F1 offsets were identified as voiced. Lower F1 steady states in vowels have also been found to cue perception of voicing (Summers, 1987, 1988). Kluender, Diehl, & Wright (1988) and Kingston (in press) argued that listeners do not hear individual acoustic cues, but perceive a higher order feature, e.g., phonemic voicing, as a result of perceptual integration of multiple cues. Kingston (in press) also argued that speakers control their production in order to produce a combination of cues which optimises the phonemic contrast for a given context. Vowel duration and vowel spectral properties have been found to have a trading relationship where F1 offset is a more pertinent cue to consonant voicing in low vowels (Walsh & Parker, 1983; Halle, Hughes, & Radley, 1957; Hillenbrand et al., 1984) and vowel duration is more pertinent in high vowels where steady-state F1 is already low (Hogan & Rozsypal, 1980). In summary, the English post-vocalic obstruent voicing contrast may be maintained via cues in the duration and spectral properties of the preceding vowel rather than via a difference in phonetic voicing, e.g., /bɪt/ - /bɪd/ → [bɪt] - [bɪ̞ːt].

Navarro Tomás (1916), Zimmerman & Sapon (1957), Quilis & Esgueva (1983), and Marín Gálvez (1995) reported that Spanish vowel productions were longer preceding voiced as opposed to voiceless consonants; however, the differences in duration were small compared to the duration differences reported for English, and no known study has reported that vowel duration affects Spanish listeners' perception of consonant voicing. Spanish has a phonemic voicing contrast for stops but not for fricatives. In intervocalic position, the phonemic voicing contrast for stops is maintained as a phonetic contrast between voiceless plosives and voiced approximants, e.g., *bota (boot) - boda (wedding)*, /bota/ - /boda/ → [bot̪a] - [boð̞a] (see Martínez Celdrán, 1984). In word final position, the contrast is maintained as a phonetic contrast between a voiceless plosive[1] and a fricative which may be voiced or partially or fully devoiced, e.g., *bit (binary*

[1]The word final voiceless stop is restricted to recent loanwords.

digit) - vid (vine), /bit/ - /bid/ → [bit̪] - [bið] or [biθ]. Preconsonantally, the post-vocalic voicing contrast is generally not maintained, e.g., *atmósfera (atmosphere) - admirar (admire)*, /atˈmosfera/ - /admiˈrar/→ [at̪m...] or [ad̪m...] or [aðm...] - [ad̪m...] or [aðm...] (see D'Introno, Del Teso, & Weston, 1995).

The present study was designed to determine how Spanish-speaking learners of English perceive the English post-vocalic stop voicing contrast (specifically for /t/ and /d/), and whether exposure to English would lead to more English-like perception. Reference will also be made to acoustic production data as it relates to findings for perception (a full analysis of production data will be reported elsewhere). Some studies have investigated vowel perception whilst varying consonant voicing (e.g., Morrison, in press), and others have investigated consonant voicing perception whilst varying vowel duration (e.g., Hillenbrand et al., 1984; and Crowther & Mann, 1992), yet others have used a word identification task, reflecting the real-life situation where listeners must simultaneously identify both vowel and consonant (e.g., Mermelstein, 1978). The present study will make use of a word identification task which also tested the participants perception and production of the English /i/-/ɪ/ contrast (Spanish only has one high front vowel). Vowel perception results were reported in Morrison (2002a, 2002b) and vowel production results will be reported elsewhere.

2. Methodology

2.1 Participants

Spanish speaking participants were three women and two men with a mean age of 20 (range 18-22). They had been raised in Mexico,[2] as children, they had not been immersed in any language other than Spanish, and they had never lived outside Mexico for a period of more than three months. They had studied English in school for 12-15 years, starting at ages 4-6. They had completed some university training in Mexico, and were university students at Anglophone universities in western Canada taking regular undergraduate courses and no language courses. They had lived in western Canada for approximately one month (range 3-6 weeks) prior to data collection.

Native English speaking participants were three women and four men with a mean age of 30 (range 25-38). They had been raised in Anglophone regions of Canada west of Quebec, were graduates of Anglophone universities in western Canada, and had been living in western Canada for at least one month prior to data collection.

2.2 Stimuli

English production and perception stimuli consisted of the words *bit - bid - beat - bead*, /bɪt/ - /bɪd/ - /bit/ - /bid/ embedded in the carrier sentence "*What they're wearing are ___ suits.*" The choice of /s/ immediately following the target words was designed to ensure that the stop would be released (an affricated release) so that stop closure duration would be measurable. Spanish production stimuli consisted of the words *bit - biti - bití - bid - bidi - bidí*, /bit/ - /ˈbiti/ - /biˈti/ - /bid/ - /bˈidi/ - /biˈdi/ in the carrier sentence *"Lo que llevan son ___ suyos" (What they're wearing are ___ of theirs)*.

For both the English and Spanish production stimuli, the carrier sentences were written four times each in random order, with additional sentences added at the beginning and end to prevent list effects. In the English stimuli, the target words were represented by pictures: a bit of pie, someone making a bid at an auction, a musical beat, and a bead. Pictures were used to avoid orthographically based errors: The letter *i* in English *bit* and *bid* represents English /ɪ/ but in Spanish orthography it represents a vowel which is closer to English /i/.

English perception stimuli consisted of edited natural speech based on the productions of a 34 year-old male monolingual English speaker who had lived all his life in western Canada. A multidimensional continuum was created in which acoustic properties varied along the following dimensions: vowel spectra (5 points, equally spaced in mel), vowel duration (7 points, equal logarithmic steps), stop closure duration (5 points, equal logarithmic steps), and speaking rate (2 points, sentence durations of 1.541 and 1.265 s excluding the duration of the target VC). The dimensions are represented visually in the cuboids shown in Figure 1, and spectral and duration values are given in Table 1. Stop closures consisted of periods of silence with no glottal pulsing. For details of the stimuli production procedures see Morrison (2002a).

[2]Several studies (Flege, 1991b; Flege, Munro, & Skelton, 1992; Flege, Bohn, & Jang, 1997; Escudero, 2000, 2001a, 2001b) used Spanish speakers from a number of countries; however, Godínez (1978) and P. Escudero (personal communication, 11 June 2001) found considerable differences in the production and perception of Spanish vowels by speakers of different Spanish dialects. Only participants from Mexico were therefore recruited in order to minimise dialectal variation.

Table 1 Spectral values at the midpoints of the vowels, and durations of vowels and stop closures in the perceptual stimuli.

Dimension point	Spectral properties						Stop duration	Vowel duration
	F1		F2		F3			
	Hz	mel	Hz	mel	Hz	mel	ms	ms
1	410	496	1700	1433	2465	1793	40	60
2	370	454	1834	1503	2569	1836	58	72
3	330	412	1974	1572	2676	1878	84	86
4	292	370	2121	1642	2786	1921	122	104
5	255	328	2275	1712	2900	1964	177	124
6								149
7								179

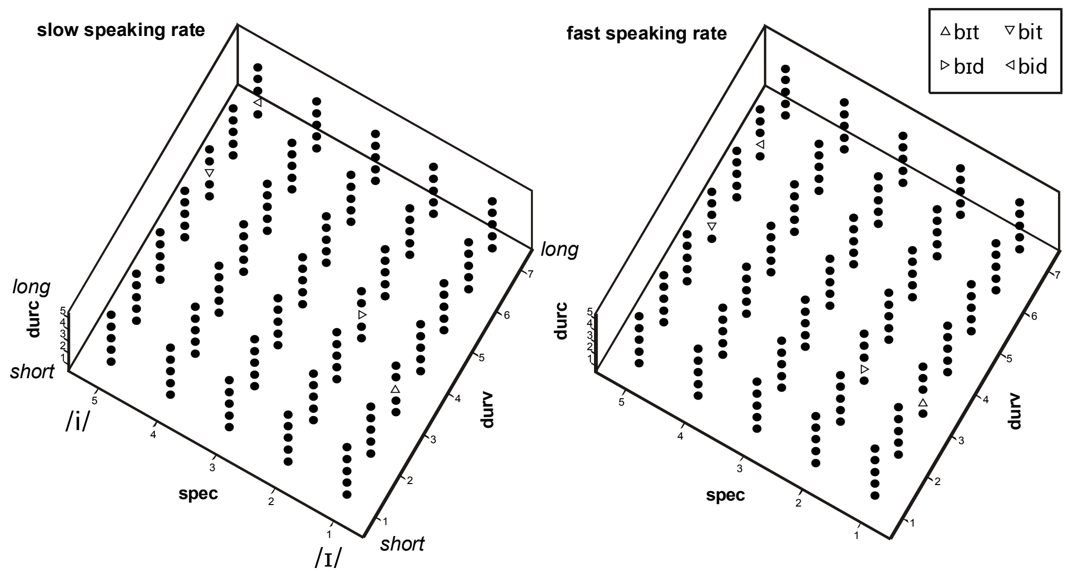

Figure 1 Cuboids representing the dimensions along which the perceptual stimuli vary. Vowel duration (durv) from front to back: 1 is shortest and 7 longest. Vowel spectral properties (spec) from right to left: 1 is most /ɪ/-like and 5 most /i/-like. Stop closure duration (durc) from bottom to top: 1 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.

2.3 Data collection

Participants were trained to interpret the pictures: The investigator elicited the words using picture cards and without pronouncing the words himself. The participants then practised reading the English production stimuli until the investigator was satisfied that they could do so fluently. The participants were then recorded reading the English stimuli. The Spanish participants also read the Spanish stimuli. Recordings were made in a soundproofed room using a Sony MZS-R5ST Mini Disc recorder and a Sony ECM-MS907 microphone.

Participants listened to the perception stimuli presented in random order via MEDS computer software (Kendall, 2001) over Optimus HP340 headphones. They responded to each stimulus by clicking on one of five pictures on the computer screen. Four of the pictures were the same as those used to represent *bit*, *beat*, *bid*, and *bead* in the production stimuli. 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.

English production and perception data were collected from the Spanish participants on two occasions: approximately one month and six months after their arrival in Canada. Spanish production data were collected once: immediately following collection of the first set of English production data. Production and perception data were collected from the English participants once.

3. Results and Discussion

Although it should be remembered that the identification task required word identification, for sake of simplicity of presentation, reference will be made to vowel perception and stop voicing perception.

English listeners identified the vowels according to vowel spectral properties with a categorical boundary at spectral continuum point 3. In the initial identification test, Spanish listeners did not have categorical perception of the vowels but did have a tendency to identify longer vowels as /i/. In the final identification test, four of the five Spanish participants had categorical perception of the vowels, one (participant SP02) had an English-like spectral boundary, and three (SP03, SP04, and SP06) had duration boundaries at vowel duration point 4. For details of vowel perception results see Morrison (2002a, 2002b).

3.1 English listeners' consonant perception

Two of the Canadian English participants identified almost all of the stops as voiceless. For these participants it seems that the lack of glottal pulsing was the only cue they attended to, and vowel and consonant duration cues alone were insufficient to give the perception of phonemic voicing. The remaining five English participants tended to identify stimuli with long vowels as having voiced consonants, Figure 2 shows the modal stop identification pattern for these listeners. Boundaries tended to be fuzzy and were not consistent across listeners. One listener had a boundary around vowel duration points 2 and 3, and one around duration points 4 and 5. The other three listeners' stop voicing perception interacted with their vowel perception: One listener had a stop voicing boundary at vowel duration point 5 when the vowel was identified as /i/, and at point 3 when the vowel was identified as /ɪ/; for another listener the boundaries were at points 6 and 3 respectively; and for the third at points 6 and 4. No speaking rate effects were observed.

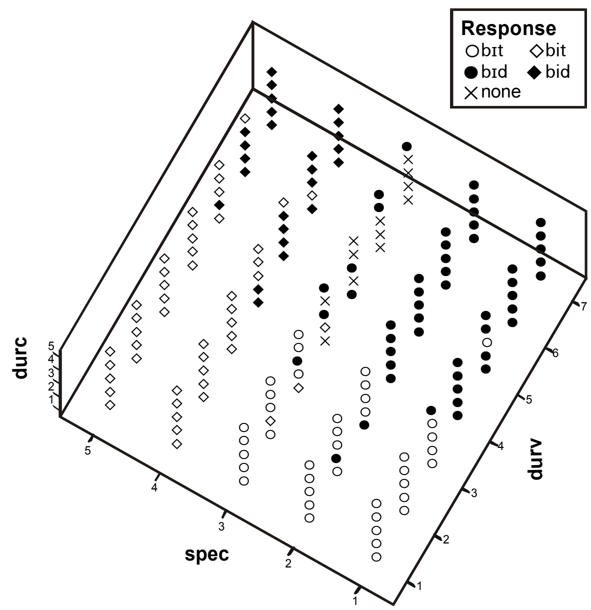

Figure 2 Modal responses to perception test from all English participants who gave both voiced and voiceless consonant responses (speaking rates combined).

These results are consistent with Hogan & Rozsypal's (1980) finding that, for Canadian English listeners, vowel duration is a perceptual cue to post-vocalic consonant voicing when the vowel is a high vowel.

3.2 Spanish listeners' consonant perception

In the initial perception test, three Spanish listeners' stop voicing identification was random. The other two listeners (participants SP04 and SP06) had categorical perception of voicing. Figure 3 shows the stop identification pattern for these listeners. Unlike the English listeners, they based their consonant identification primarily on the spectral properties of the vowels; they had a categorical boundary around spectral point 3. These listeners also made secondary use of vowel duration: vowels with spectral values 3 and 2 were more likely to lead to voiced stop identifications if the vowels were long. No speaking rate effects were observed. In the final test, the first three listeners continued to have random identification patterns for stop voicing, and the other two continued to have categorical perception based on vowel spectral properties. SP04 had a greater separation between voiced and voiceless identifications in the final test (fewer voiceless identifications for stimuli with spectral values 1 and 2, and fewer voiced identifications for stimuli with spectral values 4 and 5), and SP06 had less separation. Clearly no Spanish participant had learnt to identify consonant voicing via the same vowel duration cues that the majority of English participants used.

The perception results from SP04 and SP06 are surprising: they used vowel spectral cues to identify the voicing of the stops in the stimuli, but did not use spectral cues to identify the vowels themselves. Although as reported above, a low F1 has been reported to cue the perception of voicing for native English listeners, these Spanish listeners gave voiced responses to stimuli containing vowels with a high F1. Acoustic data from the Spanish participants' Spanish and English productions and from the English participants' productions were examined in a search for correlates which might

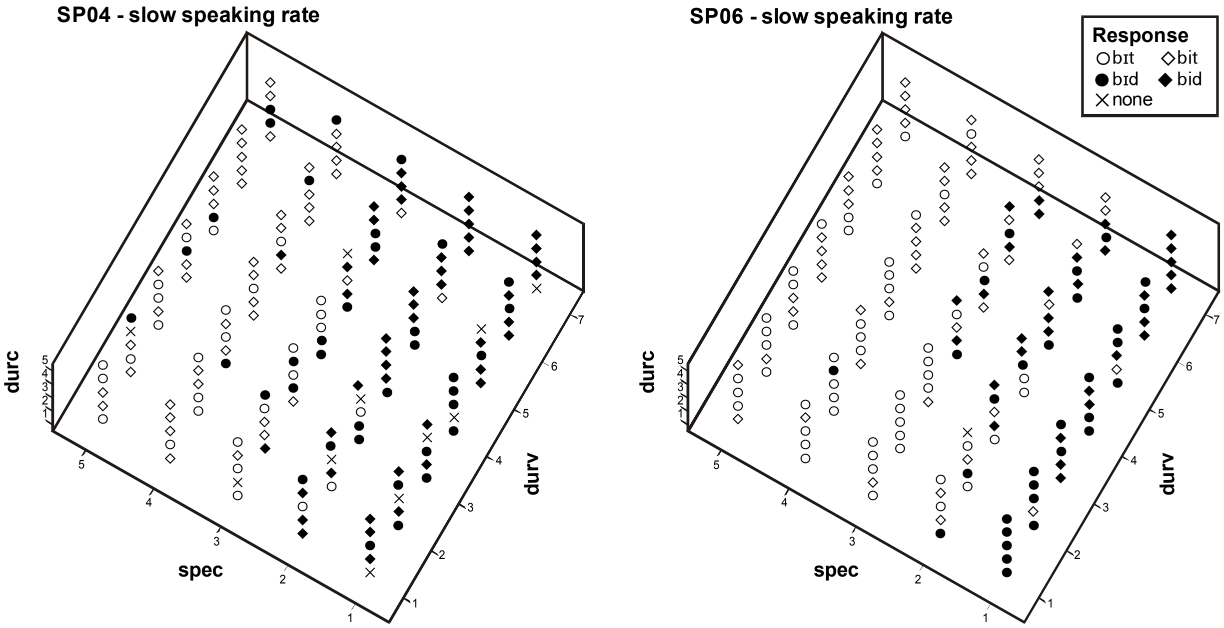

Figure 3 Responses to slow speaking rate stimuli by Spanish participants who had categorical perception of consonant voicing. Responses to the fast speaking rate stimuli were similar.

help explain the perception results. The only potentially relevant correlation found was that there was a slight tendency for the F2 offset in the Spanish speakers' /i/ to be lower before /d/ than before /t/, see Figure 4. The differences in F2 offset may be due to differences in tongue position necessary to articulate the dental plosive realisation of /t/ versus the approximant and fricative realisations of /d/. When data from all speakers were included in a mixed-design ANOVA with word as a fixed factor, participant as a random within group factor, and F2 offset frequency in Hertz as the dependent variable, the difference in F2 offset across words did not reach statistical significance at an α level of .05: $F(5, 24) = .132$, $p = .984$. The difference was more pronounced for participants SP03 and SP06, and when only data from these participants were included in the ANOVA, the difference was significant: $F(5, 6) = 20.274$, $p = .001$. Post hoc Tukey tests found no significant difference in F2 offset between *bid*, *bidi*, and *bidí*, which were all significantly different from *biti*, and *bití* with no significant difference between the latter two. The F2 offset for *bit* was significantly different from all other words except for *bidí*; however, this difference just failed to reach significance: $p = .052$. It is therefore possible that the higher F2 offset in vowels preceding Spanish /t/ compared to vowels preceding Spanish /d/ caused Spanish listeners SP04 and SP06 to identify English vowel stimuli with /i/-like spectral properties as preceding /t/ and stimuli with /ɪ/-like spectral properties as preceding /d/.

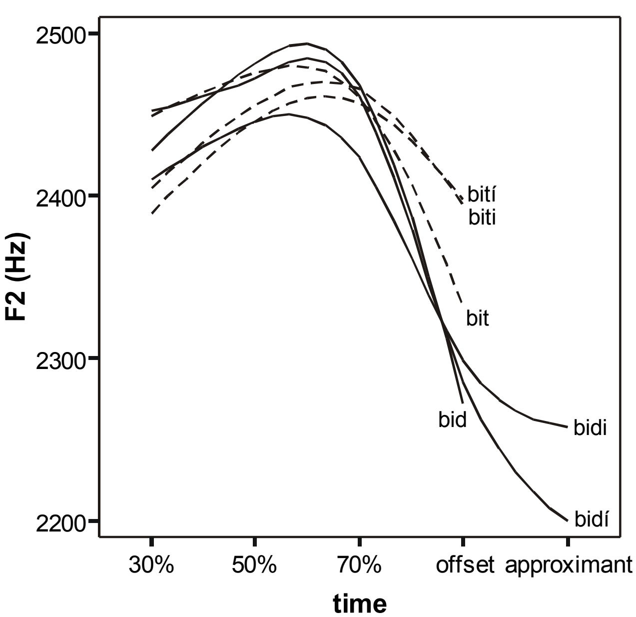

Figure 4 Mean F2 contours pooled across all Spanish speakers' productions of Spanish words. Contours interpolated from measurements taken at 30, 50, and 70% of the duration of the vowel, at the vowel offset, and, when applicable, at the centre of the following approximant.

4. Conclusion

The majority of native English listeners tested identified silent-closure stops as voiced or voiceless /d/ or /t/ according to the duration of the preceding vowel: stops preceding a longer vowel where more likely to be identified as voiced. For some English listeners, /i/ had to be longer than /ɪ/ in order to elicit voiced stop responses.

The majority of Spanish listeners had a random identification pattern for stop voicing. Two Spanish listeners used the spectral properties of the vowel to identify the voicing of the following stop, but failed to use vowel spectral properties to identify the vowels themselves. Vowels with /i/-like spectral properties were perceived as being followed by /t/, and vowels with /ɪ/-like spectral properties were perceived as being followed by /d/. The identification pattern may be a result of differences in Spanish: the F2 offsets of two participants' Spanish /i/ productions preceding Spanish /d/ were found to be lower than the F2 offsets preceding Spanish /t/. Further research including a larger number of participants is needed to determine whether the observed pattern for Spanish listeners' identification of English stop voicing is widespread, whether the F2 offset difference before productions of Spanish /t/ and /d/ is widespread, and whether the F2 offset difference is a relevant perceptual cue in Spanish.

References

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.

Crystal, T. H., & House, A. S. (1982). Segmental duration in connected speech: A preliminary analysis. *Journal of the Acoustical Society of America, 72*, 705-716.

D'Introno, F., Del Teso, E., & Weston, R. (1995). *Fonética y fonología actual del español* [Current phonetics and phonology of Spanish]. Madrid: Cátedra.

Denes, P. (1955). Effect of duration on the perception of voicing. *Journal of the Acoustical Society of America, 27*, 761-764.

Fischer, R. M., & Ohde, R. N. (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, 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.

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

Hillenbrand, J., Ingrisano, D. R., Smith, B. L., & Flege, J. E. (1984). Perception of the voiced-voiceless contrast in syllable-final stops. *Journal of the Acoustical Society of America, 76*, 18-26.

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.

Keating, P. (1985). Universal phonetics and the organization of grammars. In V. A. Fromkin (Ed.), *Phonetic linguistics: Essays in honor of Peter Ladefoged* (pp. 115-132). New York: Academic Press.

Kendall, R. A. (2001). Music experiment development system: MEDS (version 2001-B) [Computer software]. Los Angeles. [On-line]. Available: http://www.ethnomusic.ucla.edu/systematic/ Faculty/Kendall/meds.htm

Kingston, J. (in press). Keeping and losing contrasts. *Proceedings of the 28th annual meeting of the Berkeley Linguistics Society*.

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.

Lisker, L. (1957). Closure duration and the intervocalic voiced-voiceless distinctions in English. *Language, 33*, 42-49.

Luce, P. A., & Charles-Luce, J. (1985). Contextual effects on vowel duration, closure duration, and the consonant/vowel ratio in speech production. *Journal of the Acoustical Society of America, 78*, 1949-1957.

Marín Gálvez, R. (1995). La duración vocálica [Vowel duration]. *Estudios de Lingüística Universidad de Alicante, 10*, 213-226.

Martínez Celdrán, E. (1984). Cantidad e intensidad en los sonidos obstruyentes del castellano: Hacia una caracterización acústica de los sonidos aproximantes [Quantity and intensity in Spanish obstruents: Towards an acoustic characterisation of the approximants]. In Eugenio Martínez Celdrán (ed.), *Estudios de fonética experimental, I* (pp. 73–129). Barcelona: PPU.

Mermelstein, P. (1978). On the relationship between vowel and consonant identification when cued by the same acoustic information. *Perception and Psychophysics, 23*, 331-336.

Morrison, G. S. (2002a). *Effects of L1 duration experience on Japanese and Spanish listeners' perception of English high front vowels*. Unpublished master's thesis, Simon Fraser University, Burnaby, British Columbia, Canada.

Morrison, G. S. (2002b). Perception of English /i/ and /ɪ/ by Japanese and Spanish listeners: Longitudinal results. In G. S. Morrison, & L. Zsoldos (Eds.), *Proceedings of the North West Linguistics Conference 2002* (pp. 29-48). Burnaby, BC, Canada: Simon Fraser University Linguistics Graduate Student Association.

Morrison, G. S. (In Press). Perception of English /i/ and /I/ by Japanese listeners. In S. Oh, N. Sawai, K. Shiobara, & R. Wojak (Eds.), *University of British Columbia Working Papers in Linguistics Volume 8: Proceedings of NWLC 2001: Northwest Linguistics Conference.* Vancouver, BC: University of British Columbia, Department of Linguistics.

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

Port, R. (1981). Linguistics timing factors in combination. *Journal of the Acoustical Society of America, 69*, 262-274.

Port, R. F., & Dalby, J. (1982). Consonant/vowel ratio as a cue for voicing in English. *Perception and Psychophysics, 32*, 141-152.

Quilis, A., & Esgueva, M. (1983). Realización de los fonemas vocálicos españoles en posición fonética normal [Realization of Spanish vowel phonemes in normal phonetic position]. In *Estudios de fonética I, Colectanea Phonetica VII* (pp. 159-252). Madrid: Consejo Superior de Investigaciones Científicas.

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.

Revoile, S., Picket, J. M., Holden, L. D., & Talkin, D. (1982). Acoustic cues to final stop voicing for impaired- and normal-hearing listeners. *Journal of the Acoustical Society of America, 72*, 1145-1154.

Suen, C., & Beddoes, M. (1974). The silent interval of stop consonants. *Language and Speech, 17*, 126-134.

Summers, W. V. (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. V. (1988). F1 structure provides information for final-consonant voicing. *Journal of the Acoustical Society of America, 84*, 485-492.

Walsh, T., & Parker, F. (1983). Vowel length and vowel transition cues to [±voice] in post-vocalic stops. *Journal of Phonetics, 11*, 407-412.

Wolf, C. (1978). Voicing cues in English final stops. *Journal of Phonetics, 6*, 299-309.

Zimmerman, S. A., & Sapon, S. M. (1958). Note on vowel duration seen cross-linguistically. *Journal of the Acoustical Society of America, 30*, 152-153.