Perception of VOT and First Formant Onset by Spanish and English Speakers

José R. Benkí
University of Michigan

1. Introduction and background

Many languages exhibit a voicing distinction with two series of stops, voiced and voiceless, which differ primarily in the relative timing between the consonantal release and the onset of laryngeal vibration, known as voice onset time (VOT) (Lisker and Abramson, 1964). While the exact implementation of VOT for the voicing contrast differs among languages, long lag VOTs, with a long delay between release and onset of laryngeal vibration, generally signal a voiceless stop. Short lag VOTs, with a short delay between release and onset of laryngeal vibration, signal a voiced stop.

In Spanish, the voiceless stops /ptk/ are produced with a near-simultaneous release and onset of laryngeal vibration, resulting in VOT values that are approximately zero or a few milliseconds. Spanish voiced stops /bdg/ in utterance-initial position are prevoiced such that the onset of laryngeal vibration precedes the release by 40 ms or more, resulting in a negative VOT of less than -40 ms. Thus, for Spanish, the short lag VOTs of voiced stops are negative, while the voiceless stops have with a long lag VOTs between zero and 10 ms (Lisker and Abramson, 1964; Cho and Ladefoged, 2000).

In English, voiceless stops /ptk/ have a substantial delay between the release and the onset of laryngeal vibration, resulting in a VOT of 30 ms or longer, corresponding to the aspiration interval. English voiced stops /bdg/ in utterance-initial position are generally not prevoiced but released simultaneously with the onset of voicing, for a VOT of approximately zero (Caisse, 1982; Docherty, 1992). In many other languages,

As Lisker (1986) notes, the timing relation between oral release and onset of laryngeal vibration produces a number of acoustic cues, primarily manifested in differences in formant transitions. In English voiced stops, voicing begins simultaneously with the release. As a result, acoustic energy from vocal fold vibration excites the first formant (F1) during the entire consonant-vowel (CV) transition. Low frequency periodic acoustic energy from laryngeal vibration makes F1 audible during its rise from the consonantal release to the vowel steady state frequency.

The long VOTs for English voiceless stops result in F1 transitions that differ radically from corresponding transitions of voiced stops. Since voicing onset occurs much later than the release in voiceless stops, F1 is not excited until very late in the CV transition, at which time the vocal tract is close to the vowel steady-state configuration. The frequency of F1 at voicing onset frequency is much higher for English voiceless stops than for voiced stops. This delay in the excitation of F1 is also known as F1 cutback (Liberman et al., 1958).

Early work in speech perception has verified the role of VOT in the perception of voicing in utterance-initial position for speakers of English and other languages (Liberman et al., 1958; Lisker and Abramson, 1970). In these studies, listeners have been found to classify stops as voiced or voiceless depending on the VOT value, consistent with the observed VOTs of voiced and voiceless stops for the appropriate language. Stops with VOT values longer than some boundary value are classified as voiceless, while stops with VOTs shorter than the boundary value are classified as voiced.

In addition to the effects of VOT, the effects of F1 transition and frequency at voicing onset have also been shown to be important in the perception of the voicing contrast in English (Stevens and Klatt, 1974; Lisker et al., 1977; Summerfield and Haggard, 1977; Kluender, 1991; Pind, 1999 for the Icelandic aspiration contrast). These studies show that stops with a longer F1 transition and/or lower F1 frequency at voicing onset, all other things being equal (such as VOT), will be classified as voiced relative to stops with shorter F1 transitions and/or higher F1 frequency at voicing onset. Thus, VOT and the F1 transition pattern have a trading relation (Repp, 1982) in the perception of the voicing contrast. The direction of the effect of F1 transition pattern on voicing classification is thus consistent with the observed production data for English and other languages with aspiration in the voiceless stops.

In a series of experiments aimed at showing the cross-species generality of sensitivity to VOT, Kuhl and Miller (1975, 1978) obtained chinchilla and human identification functions for voicing judgements on a VOT continuum for different places of articulation. Kuhl and Miller found similar boundary values for the human and nonhuman categorization results. Kluender and colleagues (Kluender, 1991; Kluender and Lotto, 1994; Kluender et al., 1995) present data that show that humans and Japanese quail are similar in voicing categorization of synthetic stimuli varying in VOT, F1 transition pattern, and place of articulation (but for place of articulation cf. Miller, 1977; Benkí, 2001).

Despite the cross-species results for F1 transition pattern, the possibility remains that the effect of F1 transition pattern on voicing perception arises from speakers being sensitive to the covariation in production between VOT and F1 transition pattern. A cross-linguistic investigation could offer additional insight into the basis for the effect of F1 transition pattern on voicing perception. Languages such as Spanish that make a voicing distinction contrast between truly prevoiced /bdg/ and voiceless unaspirated /ptk/ lack the F1 transition cues to the voiceless category (or at least have substantially reduced F1 transition cues relative to languages such as English). Because laryngeal vibration begins nearly simultaneously with oral release for the voiceless stops, Spanish voiceless stops will have F1 transition patterns very similar to those of Spanish voiced stops, unlike the situation in English. Presumably, Spanish speakers would not be able to learn the covariation between F1 transition pattern and VOT like English listeners do.

One goal of the present study is an understanding of the relative importance of both VOT and F1 transition manipulations on the perception of voicing in Spanish, a language whose voiceless consonants exhibit much less aspiration than those of English. A second goal is a deeper understanding of the perception of voicing by comparing results from speakers of both languages. If the sensitivity to F1 transition pattern in voicing perception by English speakers arises from learning a pattern of covariation, then Spanish speakers should exhibit much less sensitivity to the F1 transition pattern. On the other hand, sensitivity to F1 transition patterns on the part of Spanish speakers would be consistent with a general auditory explanation for the effect of F1 on voicing categorization.

2. Method

Synthetic Spanish and English stimuli, varying in both VOT and F1 onset frequency, were presented to Spanish and English speaking participants respectively for categorization. The stimuli were resynthesized versions of natural productions of a male bilingual speaker of Midwestern American English and Carribean Spanish (the author) of the English words *bossy* [ˈb̥asi] and *posse* [ˈp^{h}asi], and the Spanish words *vaso* [ˈbaso] ("drinking glass") and *paso* [ˈpaso] ("step"). Aside from differences in VOT and release, and consequent differences in F1 cutback, the f0 and F1 trajectories of the initial syllables of all four items were found to be highly similar. Therefore, the F1 and fundamental frequency (f0) contours of the initial syllable were held constant across all stimuli. Differences are detailed below.

2.1. English stimuli

The Synthworks implementation of the cascade branch of the Klatt and Klatt (1990) synthesizer was used to create the stimuli at 12-bit resolution and 10 kHz sampling rate on a Windows NT computer. VOT was varied from 0 ms to 40 ms in 5 ms steps by manipulating the AF, AV and AH synthesis parameters. At oral release, AF and AH were set to 60 dB for 5 ms. For stimuli with VOT=0 ms, AV was also set to 60 dB. For stimuli with longer VOTs, AH remained at 60 dB for the duration of the aspiration interval, and AV was set to 60 dB at the end of the aspiration interval. The VOT of each stimulus was confirmed using spectrogram and time series displays from the Praat software package.

The F1 transition pattern was controlled by setting the F1 parameter to five different values at oral release: 200, 300, 400, 500, and 600 Hz. The F1 parameter was then linearly ramped to the steady-state value of 700 Hz over 30 ms, with B1 set to 90 Hz during the transition and steady-state interval of the initial vowel. The initial vowel had a duration of 150 ms, with F2 ramped from 1000 at release to 1300 Hz at vowel offset; F3 set to a constant 2200; F4 set to a constant 2900; and F5 ramped from 3700 at release to 4000 at offset. The orthogonal F1 and VOT manipulations yielded 45 stimuli. Each stimulus was 390 ms long. Since the cascade branch of the synthesizer was used, the amplitude of F1 was controlled automatically during the aspiration interval. However, the AH source during the aspiration interval has little energy in the F1 region.

The parameters for the unstressed second syllable [si], modeled on the above natural productions, were held constant for all stimuli.

2.2. Spanish stimuli

The Spanish stimuli were created using the same synthesizer as the English stimuli. The initial syllable VOT and F1 transition pattern manipulations were also entirely parallel to the English stimuli, except that the VOT range covered –20 ms to 20 ms. For prevoicing, low-frequency energy was enhanced by setting F1=200 Hz, B1=40 Hz, TL=65, OQ=24, AV=55 dB during the prevoicing interval. The high-frequency harmonics were further attenuated by setting B2, B3, B4, and B5 to 1000 Hz during the same interval. Following release, these values were matched to those used for the English stimuli. The initial vowel had a duration of 135 ms, with F2 at a constant 1200; F3 ramped from 2250 at release to 2350 at offset; F4 set to a constant 2900; and F5 ramped from 3700 at release to 4000 at offset. The orthogonal F1 and VOT manipulations yielded 45 stimuli. Each stimulus was 430 ms long plus the prevoicing interval, if any. The parameters for the unstressed second syllable [so], modeled on the above natural productions, were held constant for all stimuli.

2.3. Participants

Nine English-speaking participants were recruited from undergraduate linguistics courses at the University of Michigan and participated for course extra credit. Six native speakers of Spanish were recruited from the University of Michigan community, and were paid a nominal fee for participation. None reported any hearing impairments.

2.4. Procedure

The experiment was run using the Superlab software package running on a Windows NT notebook computer. The stimuli were presented binaurally over AKG headphones at a comfortable level, and a response box was used to collect judgments. Participants were run 1 or 2 at a time in an anechoic chamber, each on a single computer station in sessions lasting 20 minutes.

Each trial consisted of a stimulus presented to the participant and the identification by the English-speaking participants as one of *bossy*, *posse*, or *other* using the response box, or as one of *vaso*, *paso*, or *otro* by the Spanish-speaking participants. English or Spanish instructions appeared on the computer screen asking the participant to press the appropriate color-coded key to register a response. An 800 ms ISI intervened between each keypress and the subsequent trial. Ten responses per stimulus per participant were collected in two blocks of 5 presentations of each stimulus in random order. A pause separated the two blocks.

Following the experiment, participants were interviewed briefly concerning their impressions of the stimuli. In particular, they were asked if they clearly heard the words *bossy* or *posse* (*vaso* or *paso* for the Spanish-speaking participants), or any other words.

3. Results

Trials with response times less than 100 ms or greater than 3000 ms relative to stimulus onset were excluded from the analysis. Trials with the short response times were excluded because the response occurred before the participant heard the entire first syllable. Trials with the excessively long response times were excluded to avoid responses resulting from second guesses. These latency criteria excluded 1.1% of total English responses and 3.7% of total Spanish responses. Of the trials that met the latency criterion, those with *other* responses were also excluded, numbering 9.4% of the English responses and 13.7% of the Spanish responses. The English-speaking participants reported hearing mostly the words *bossy* and *posse*, as expected, but also on occasion hearing [asi] (no initial consonant), and less frequently *mossy* [masi] and *lossy* [lasi]. Reports of initial consonant by the English speakers probably correspond to the *other* responses for the stimuli with F1 onset frequency of 600 Hz. Reports of initial [m] or [l] reports probably correspond to the *other* responses for stimuli with F1 onset frequency of 200 Hz and short VOT values. A similar pattern holds for the Spanish speakers, who reported hearing mostly *vaso* and *paso* as intended, but also infrequently [aso], [maso], and [laso].

The remaining responses are plotted in Figure 1 for the English-speaking participants and Figure 2 for the Spanish-speaking participants. The proportion of /p/ responses, averaged across participants is displayed for each combination of F1 onset frequency and VOT value.

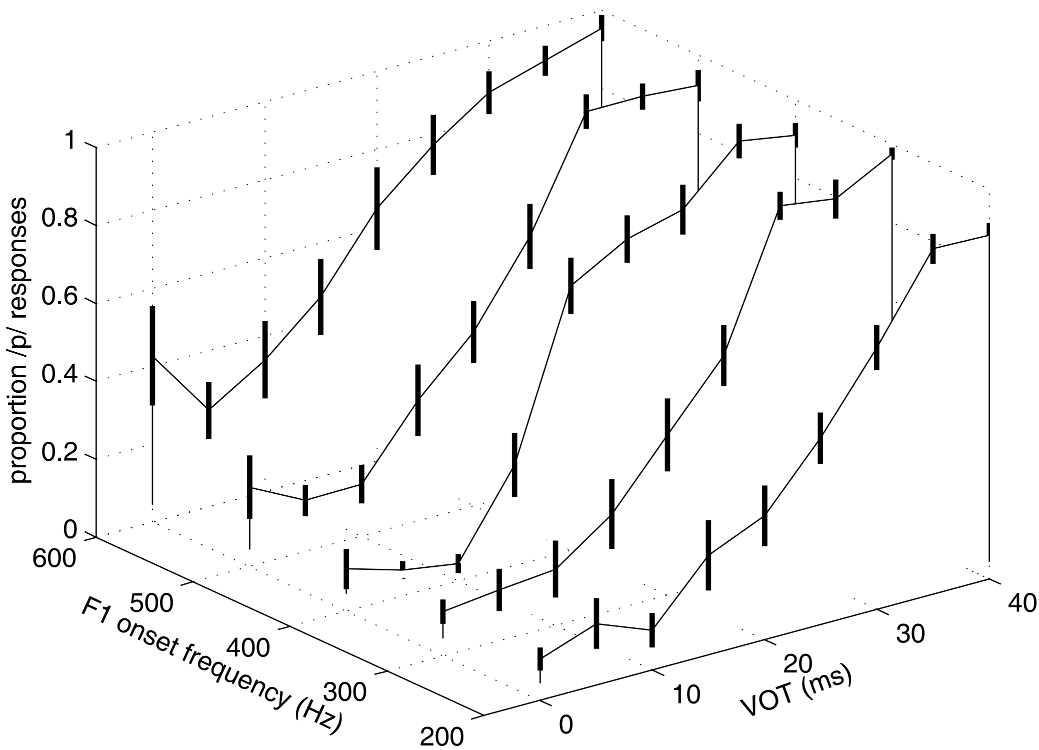

Figure 1. Proportion of /p/ responses averaged across English-speaking participants as a function of VOT value (0 to 40 ms) and F1 onset frequency (200 to 600 Hz). Error bars represent ±1 standard error.

For both groups of participants, increasing VOT and F1 onset frequency are both correlated with increased rates of /p/ responses. These effects are evaluated preliminarily in the pooled analysis below, and then in a more definitive subjects analysis.

The data plotted in Figures 1 and 2 were analyzed using a binary logistic regression analysis for each language, with VOT and F1 onset frequency as the independent variables and probability of voiceless (/p/) response as the dependent variable (Hosmer and Lemeshow, 1989; cf. Nearey 1997, Benkí, 2001 for examples of loglinear analyses of phonological categorization). The pooled analyses are presented in Table 1 and the resulting models for each language are plotted in Figure 3 for F1 onset = 200 Hz and F1 onset = 600 Hz.

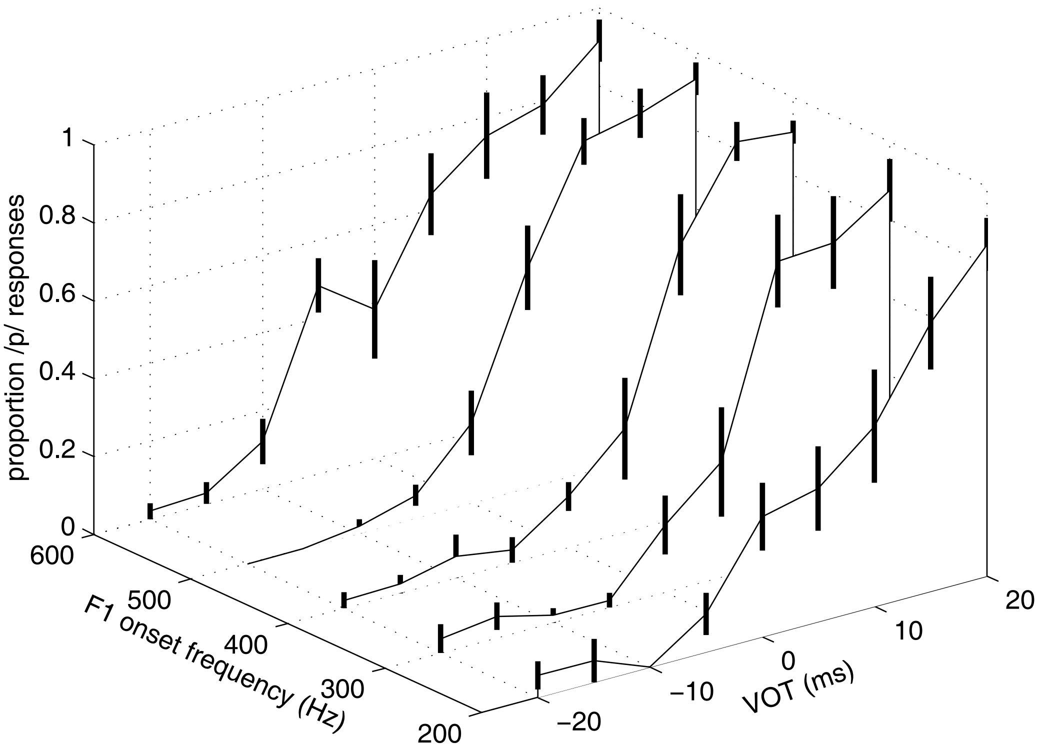

Figure 2. Proportion of /p/ responses averaged across Spanish-speaking participants as a function of VOT value (–20 to +20 ms) and F1 onset frequency (200 to 600 Hz). Error bars represent ±1 standard error.

Table 1. Binary logistic regression models of /p/ responses as a function of VOT and F1 onset frequency for the pooled English and Spanish-speaking participants.

English model: 3630 cases, Deviance = 3392.4

	β (95% C.I.)	G	Sig.
Constant	–3.55		
VOT (ms)	0.132 ±0.0083	1577.2	$p < 0.001$
F1 onset (Hz)	0.00244 ±0.00062	61.8	$p < 0.001$

Spanish model: 2253 cases, Deviance = 1625.8

	β (95% C.I.)	G	Sig.
Constant	–2.28		
VOT (ms)	0.180 ±0.014	1379.8	$p < 0.001$
F1 onset (Hz)	0.00285 ±0.00086	39.5	$p < 0.001$

While the models for each language differ in the constant β_0, otherwise the models are remarkably similar. Both the VOT and F1 onset parameters are significant for each language as measured by *G*, the reduction in deviance when each parameter is added to the model. A more rigorous evaluation of the effects of each variable is provided by the subjects analysis. Logistic regression models were calculated for each participant and the resulting parameters are displayed in Table 2, with Bonferroni-corrected 95% confidence intervals for the mean parameter estimates. Identification functions using the mean parameter estimates in the logistic regression model are plotted in Figure 3.

Table 2. Logistic regression parameters for each participant and averages with 95% Bonferroni-corrected confidence intervals (family of 3 comparisons).

Participant	Constant (β_0)	VOT	F1 onset
English 1	-7.88	0.285	0.00525
2	-2.57	0.104	0.00215
3	-6.70	0.210	0.00453
4	-3.41	0.110	0.00233
5	-6.24	0.197	0.00392
6	-7.90	0.300	0.00547
7	-3.65	0.143	0.00109
8	-1.40	0.050	0.00114
9	-3.36	0.214	0.00307
Mean	-4.79 ±2.47	0.179 ± 0.086	0.0032 ± 0.0017
Spanish 1	-1.87	0.292	0.00295
2	-3.15	0.216	0.00301
3	-1.58	0.084	0.00179
4	-1.27	0.235	0.00136
5	-3.91	0.186	0.00458
6	-2.99	0.244	0.00499
Mean	-2.46 ±1.59	0.210 ± 0.108	0.0031 ± 0.0022

The mean parameter estimates are consistent with the estimates from the pooled analysis. The confidence intervals are large, reflecting individual differences but also the difficulties of interpreting parameters from models being based on slightly fewer than 10 judgments per stimulus.

The major difference between Spanish and English speakers is the location of the VOT boundary, 5.8 ms for Spanish speakers and 19.6 ms for English speakers, as calculated using the mean parameter estimates with F1 onset = 400 Hz. The location of the VOT boundary in the model is controlled by the β_0 parameter, which is significantly larger in magnitude for the English speakers than for the Spanish speakers. The mean Spanish model shows a slightly steeper slope for VOT than the mean English model. The F1 onset parameters are very similar and statistically no different according to the confidence intervals. The VOT boundary value shifts by 5.9 ms for Spanish and 7.1 ms for English for a change in F1 onset frequency from 200 to 600 Hz.

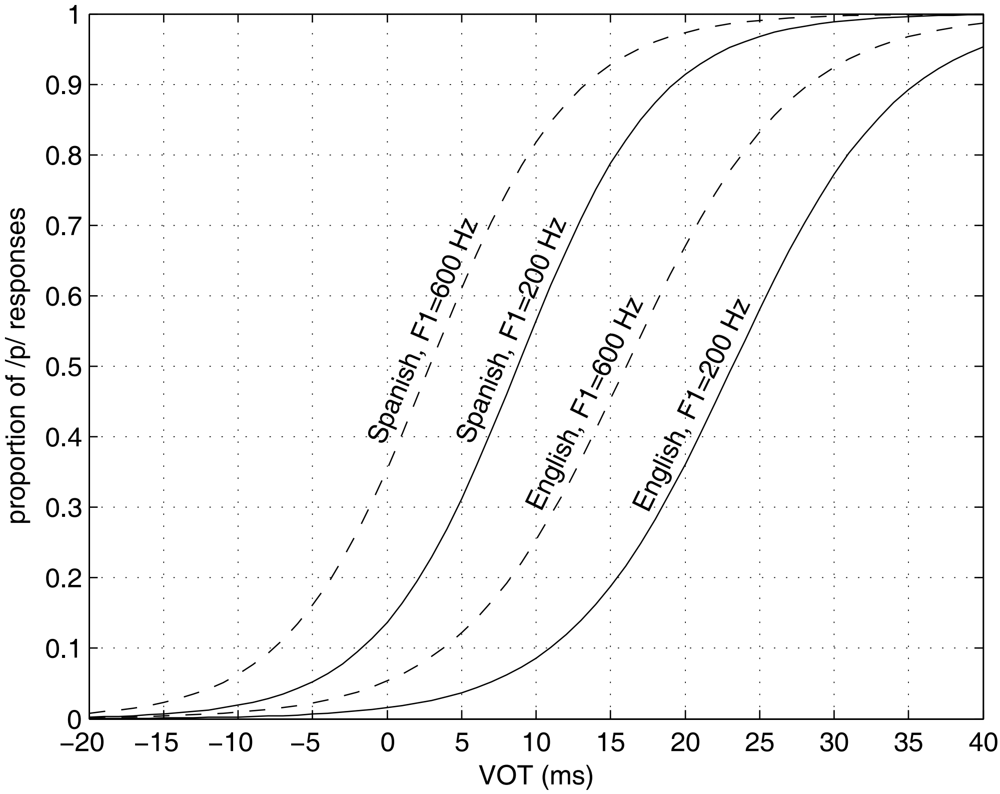

Figure 3. Binary logistic regression models of English and Spanish /p/ categorization using the mean parameter estimates from Table 2, for F1 onset frequencies of 200 and 600 Hz.

4. Discussion and conclusion

Both sets of speakers exhibit similar sensitivities to the F1 transition pattern in the perception of the voicing contrast. The results are therefore consistent with a general auditory (and therefore unlearned) mechanism for the effect of F1 transition pattern in voicing perception.

The perceptual VOT boundary value reported here for Spanish speakers, 6.3 ms, is consistent with reports of native speakers of Spanish showing VOT boundary values of 14 ms or less for word-initial bilabial stops (Abramson and Lisker, 1973; Williams, 1977a, b; López-Bascuas et al., 1998 as cited by Rosner et al., 2000). While the Spanish speakers in the present study have considerable exposure to English, the boundary value is still significantly shorter than for English speakers. Data from Canadian French speakers (Caramazza et al., 1973) and Spanish-English bilinguals (Elman et al., 1977; Williams, 1977b) indicate that perceptual VOT boundary values from a /ba/–/pa/ continuum may become longer for speakers with considerable English exposure or proficiency. The Spanish speakers in the present study may have failed to show this shift because the continuum endpoints were natural-sounding actual Spanish words and the intermediate stimuli also were natural-sounding, though ambiguous.

The magnitude of the effect of F1 onset frequency manipulation on voicing compares well to previously reported data for English. Kluender (1991) reports a difference of approximately 7 ms in VOT boundary value for VOT continua with 30 ms transitions and a 400 Hz difference in F1 onset frequency, averaged across bilabial, alveolar, and velar places of articulation. For the present study, the boundary value shifts for the best fitting models of the pooled data over a 400 Hz manipulation of F1 onset frequency are 6.3 ms for Spanish and 7.4 ms for English.

The finding of an effect of F1 onset frequency for perception of voicing for the Spanish speakers conflicts with negative results for F1 in an early comparative developmental study of voicing perception by Simon and Fourcin (1973). In that report, English-speaking children showed F1 transition pattern sensitivity by age 4, while French-speaking children never showed any sensitivity to F1. However, a number of issues make generalizations on the basis of those results difficult. For many participants, only 3 judgements per stimulus per participant were collected. The English children were tested on two continua, *goat–coat* and *ball–Paul*, while the French children were tested on a bisyllabic *dodo–Toto* continuum, which may have had a lower vowel target F1 frequency than the English *ball–Paul* continuum. A further confound with assessing sensitivity to F1 transition patterns is that the high frequency F1 onset continuum had *no* F1 transition, acknowledged as unnatural in that report. The present study has a more careful sampling of F1 onset frequency, natural-sounding stimuli, matched F1 transitions and targets, and omission of trials when a consonant other than /p/ or /b/ is perceived.

Both English and Spanish-speaking listeners showed similar sensitivity to F1 transition, despite different patterns of covariation of F1 transition with the obstruent voicing categories in those languages. The present findings are consistent with F1 transition pattern not being a learned acoustic-phonetic cue for voicing. Instead, sensitivity to F1 transition patterns may instead arise from a more general property of the auditory system, in line with previous cross-species results. The effect of the F1 transition pattern may be mediated by the low frequency property, proposed by Stevens and Blumstein, (1981) to be an invariant acoustic correlate for voiced obstruents. On the basis of a number of speech and nonspeech experiments, Kingston and Diehl (1995) argue that the various acoustic-phonetic cues for voicing, such as F1 transition patterns, closure duration, and fundamental frequency (f0) cohere as a result of properties of the auditory system not specific to speech.

Although f0 also covaries with VOT, a recent animal study indicates that sensitivity to f0 in voicing perception may be learned, unlike sensitivity to the F1 transition pattern. Voiced obstruents tend to depress f0 in the following vowel relative to voiceless obstruents (House and Fairbanks, 1953), and human listeners are sensitive to this covariation (Haggard et al., 1970). Holt et al. (2001) trained Japanese quail to discriminate elements of a VOT continuum with three different training regimens that corresponded to the observed pattern of covariation of f0 and voicing, the reverse, and no covariation. After training, the birds responded to novel stimuli according to their training, suggesting that f0 effects on voicing are learned. The F1 transition pattern appears to be different than that of f0 with regard to the perception of voicing.

It may be possible that the Spanish speakers of the present study have acquired sensitivity to F1 transition covariation with voicing through their exposure to English, though it is surprising that their perceptual VOT boundary values have not shifted much relative to reported values for monolingual Spanish speakers. Further data from monolingual Spanish speakers with little or no exposure to English (or other languages with aspirated plosives) would bear on this issue.

Acknowledgments

Special thanks to Justin Fitzpatrick for helping to design the stimuli, and to Helen Hanson for advice on the Klatt synthesizer. A preliminary version of this work was presented at the Speech perception as pattern classification workshop at the Max Plank Institute for Psycholinguistics in Nijmegen, July 2001, organized by John Kingston, Terry Nearey, and Roel Smits. I am grateful to audience members for useful comments and discussion. All errors are mine.

References

Abramson, A. S. and Lisker, L. (1973). Voice-timing perception in Spanish word-initial stops. Journal of Phonetics **1**, 1-8.

Benkí, J. R. (2001) Place of articulation and first formant transition pattern both affect perception of voicing in English. Journal of Phonetics **29**, 1–22.

Caisse, M. (1982) Cross-linguistic differences in fundamental frequency perturbation induced by voiceless unaspirated stops, M.A. thesis, University of California.

Caramazza, A., Yeni-Komshian, G. H., Zurif, E. B., and Carbone, E. (1973) The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals. Journal of the Acoustical Society of America **54**, 421–428.

Cho, T. and Ladefoged, P. (1999) Variation and universals in VOT: evidence from 18 languages. Journal of Phonetics **27**, 207–229.

Docherty, G. J. (1992) The timing of voicing in English obstruents. Berlin: Foris.

Elman, J. L., Diehl, R. L., and Buchwald, S. E. (1977) Perceptual switching in bilinguals. Journal of the Acoustical Society of America **62**, 971–974.

Haggard, M., Ambler, S., and Callow, M. (1970) Pitch as a voicing cue. Journal of the Acoustical Society of America **47**, 613–617.

Holt, L. L., Lotto, A. J., and Kluender, K. R. (2001) Influence of fundamental frequency on stop-consonant voicing perception: A case of learned covariation or auditory enhancement? Journal of the Acoustical Society of America **109**, 764–774.

Hosmer, D. W. and Lemeshow, S. (1989) Applied logistic regression. New York: John Wiley and Sons.

House, A. S., and Fairbanks, G. (1953) The influence of consonant environment on the secondary acoustical characteristics of vowels. Journal of the Acoustical Society of America **25**, 105–135.

Kingston, J. and Diehl, R. L. (1994) Phonetic knowledge, Language **70**, 419–454.

Klatt, D. H. and Klatt, L. C. (1990) Analysis, synthesis, and perception of voice quality variations among male and female talkers, Journal of the Acoustical Society of America **87**, 820–857.

Kluender, K. R. (1991). Effects of first formant onset properties on voicing judgments result from processes not specific to humans. Journal of the Acoustical Society of America **90**, 83–96.

Kluender, K. R. and Lotto, A. J. (1994) Effects of first formant onset frequency on [-voice] judgments result from auditory processes not specific to humans. Journal of the Acoustical Society of America **95**, 1044–1052.

Kluender, K. R., Lotto, A. J., and Jenison, R. L. (1995) Perception of voicing for syllable-initial stops at different intensities: Does synchrony capture signal voiceless stop consonants? Journal of the Acoustical Society of America **97**, 2552–2566.

Kuhl, P. K. and Miller, J. D. (1975) Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants, Science **190**, 69–72.

Kuhl, P. K. and Miller, J. D. (1978) Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli, Journal of the Acoustical Society of America **63**, 905–917.

Liberman, A.M., Delattre, P.C., and Cooper, F. S. (1958). Some cues for the distinction between voiced and voiceless stops in initial position. Language and Speech **1**, 153–167.

Lisker, L. (1986) “Voicing” in English: a catalog of acoustic features signaling /b/ versus /p/ in trochees. Language and Speech **29**, 3-11.

Lisker, L. and Abramson, A.S. (1964) A cross-language study of voicing in initial stops: acoustical measurements. Word **20**, 384-422.

Lisker, L. and Abramson, A.S. (1970). The voicing dimension: Some experiments in comparative phonetics. Proceedings of the 6th International Conference of Phonetic Sciences, (B. Halá, M. Romportl and P. Janota, editors). pp. 563–567. Prague: Academia.

Lisker, L., Liberman, A. M., Erickson, D. M., Dechovitz, D. and Mandler, R. (1977). On pushing the voice onset time (VOT) boundary about. Language and Speech **20**, 209–216.

López-Bascuas, L. E., Fahey, R. P., García-Albea, J. E., and Rosner, B. S. (1998). Identificación del orden temporal en sonidos de habla y de no-habla. Cognitiva **10**, 195-210.

Miller, J. L. (1977) Nonindependence of feature processing in initial consonants, Journal of Speech and Hearing Research **20**, 519–528.

Nearey, T. M. (1997) Speech perception as pattern recognition, Journal of the Acoustical Society of America **101**, 3241–3254.

Pind, J. (1999). The role of *F*1 in the perception of voice onset time and voice offset time. Journal of the Acoustical Society of America **106**, 434–437.

Repp, B. (1982). Phonetic trading relations and context effects: New evidence for a phonetic mode of perception. Psychological Bulletin **92**, 81–110.

Rosner, B. S., López-Bascuas, García-Albea, J. E., and Fahey, R. P. (2000) Voice-onset times for Castilian Spanish initial stops. Journal of Phonetics **28**, 217–224.

Simon, C. and Fourcin, A. J. (1978) Cross-language study of speech-pattern learning. Journal of the Acoustical Society of America **63**, 925-935.

Stevens, K. N. and Blumstein, S. E. (1981) The search for invariant acoustic correlates of phonetic features. Perspectives on the Study of Speech (P. D. Eimas and J. L. Miller, editors), pp.1–38. Hillsdale: Erlbaum.

Stevens, K. N. and Klatt, D. H. (1974). Role of formant transitions in the voiced-voiceless distinction of stops. Journal of the Acoustical Society of America **55**, 653–659.

Summerfield, A. Q. and Haggard, M. P. (1977). On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. Journal of the Acoustical Society of America **62**, 435–448.

Williams, L. (1977a). The voicing contrast in Spanish. Journal of Phonetics **5**, 169-184.

Williams, L. (1977b). The perception of stop consonant voicing by Spanish-English bilinguals. Perception and Psychophysics **21**, 289-297.

ISB4: Proceedings of the 4th International Symposium on Bilingualism

edited by James Cohen, Kara T. McAlister, Kellie Rolstad, and Jeff MacSwan

Cascadilla Press Somerville, MA 2005

Copyright information

ISBN 978-1-57473-210-8 CD-ROM
ISBN 978-1-57473-107-1 library binding (5-volume set)

Ordering information

To order a copy of the proceedings, contact:

Cascadilla Press
P.O. Box 440355
Somerville, MA 02144, USA

phone: 1-617-776-2370
fax: 1-617-776-2271
sales@cascadilla.com
www.cascadilla.com

Web access and citation information

This paper is available from www.cascadilla.com/isb4.html and is identical
to the version published by Cascadilla Press on CD-ROM and in library binding.