# Speaker Ethnicity as a Perceptual Shift Trigger

Daniel W. Bruhn

*North Carolina State University*

#### Abstract

The purpose of this investigation was to determine whether a change in speaker ethnicity can trigger a shift in the perceptual *voiced*/*voiceless* boundary of Spanish-English bilingual listeners, given that Spanish and English exhibit different values of VOT for this category boundary. Two groups, Spanish-English bilinguals and English monolinguals, were presented with auditory-visual stimuli of the utterance [bo] and were asked to identify each stimulus as "bo" or "po." The stimuli varied independently in the VOT of the bilabial stop and the speaker ethnicity presented. The experiment showed that the English monolinguals shifted the VOT value of their /b/-/p/ category boundaries in response to the change in speaker ethnicity, while the bilinguals failed to exhibit a significant shift. These results suggest that 1) speaker ethnicity lies under the level of consciousness required to trigger a perceptual shift, 2) English listeners utilize stereotypes of Hispanic English in perception of utterances, and 3) Latin American Spanish-English bilinguals operate under no comparable stereotypes of American English. (There is some preliminary data to support the notion that speaker ethnicity may indeed trigger a perceptual shift in bilinguals with extensive L2 experience, but more research is needed to verify this finding.)

## Introduction

The primary cue for perception of voicing[1] in prevocalic stops in both Spanish and English is Voice Onset Time (VOT). These languages differ in their values of VOT corresponding to the *voiced* and *voiceless* categories. In the case of Spanish, the voiced bilabial stop [b] is produced with large negative (*long lead*) VOT around $-40$ ms (Lisker and Abramson, 1964). The articulatory gesture needed to accomplish this is *prevoicing*, which involves initiating vocal fold vibration prior to release of the closure (in this case, $40$ ms prior). The voiceless bilabial stop [p] in Spanish is produced with small positive (*short lag*) VOT around $+4$ ms (Cho and Ladefoged, 1999). For English, on the other hand, $0$ ms and short lag VOT correspond not to [p], but to [b] (Docherty, 1992), often causing English listeners to perceive the Spanish [p] as their /b/. The prevocalic English /p/ exhibits a *long lag* VOT of approximately $+30$ ms (Caisse, 1982) and is articulated as the aspirated [$\mathrm{p^h}$].

Thus, prevocalic bilabial stops generally range in VOT values from $-40$ ms (voiced) to $+4$ ms (voiceless) for Spanish, and from $0$ ms (voiced) to $+30$ ms (voiceless) for English. As a consequence, the boundary of *categorical perception* between prevocalic /b/ and /p/ for Spanish corresponds to a VOT value somewhere between $-40$ and $+4$ ms. This is less than the VOT value for the same perceptual category boundary in English, which lies somewhere between $0$ and $+30$ ms.

In view of this disparity in category boundaries between the two languages, an investigation was conducted to determine whether Spanish-English bilinguals shift their perceptual boundary in response to the ethnicity of a speaker. It was hypothesized that a bilingual encountering a Hispanic individual producing a bilabial stop with

[1] See Benkí (2005) for a discussion of the importance of F1 transitions.

a VOT around 0 ms will perceive /p/, while this listener encountering an Anglo speaker producing the same phone will perceive /b/. The ethnicity of the speaker will act as cue to placement of the category boundary. Elman, Diehl, and Buchwald (1977) have already shown that Spanish-English bilinguals "vary their placement of category boundaries as a function of language set."[2] However, what is not known is whether such individuals also shift their voicing boundary based on perceived ethnicity of the speaker in the presence of language-independent phonetic stimuli. The purpose of this investigation was therefore to determine if this hypothesized perceptual shift does indeed take place.

## Previous Literature

There are several previous studies pertinent to the current investigation.

### VOT Definitions

As cited above, both Lisker and Abramson (1964) and Cho and Ladefoged (1999) present measurements of Spanish VOT. It is important to note the definitions of VOT utilized in these two studies:

> "the time interval between the burst. . . and the onset of quasi-periodicity that reflects laryngeal vibration"
> (Lisker and Abramson, 1964, p. 422)

> "the interval between cursors placed at the onset of release (the final release, if there was more than one) and the onset of the first complete vibration of the vocal folds as indicated on the waveform"
> (Cho and Ladefoged, 1999, p. 215)

These appear slightly different, although one could certainly argue that the "onset of quasi-periodicity" is basically equivalent to the "onset of the first complete vibration." Whatever the case may be, this research avoids the conflict by employing one definition of VOT in all measurements, namely that of Lisker and Abramson.

[2] See discussion in the next section.

### Acoustic Cues

Benkí (2005) notes that the voiced/voiceless distinction in stops involves a number of other acoustic cues, one of these being a difference in F1 transitions. Obstruents produced with short lag VOT exhibit a lower F1 frequency at voicing onset than those produced with long lag VOT, since the vocal tract is still early in the process of moving to a steady-state configuration when its resonances are activated. Benkí also describes the results of a perception experiment that demonstrated the sensitivity of both English and Spanish speakers to F1 transitions, independent of VOT. He determined that the listeners shifted their category boundaries to smaller VOT values in response to higher F1 onset frequencies.

Benkí's results, however, do not invalidate the exclusive use of VOT for determining category boundaries. His experiment utilized unnatural stimuli in which F1 varied independently of VOT. In reality, the distinct F1 transitions are caused by the VOT differences, and thus are not independent of VOT. The same is true of aspiration in the case of [$p^h$]: The duration of aspiration corresponds to the value of VOT (Benkí, 2005). Calculation of category boundaries as a function of VOT therefore accounts for these additional acoustic cues.

### Category Boundaries

Abramson and Lisker (1970), Wood (1976), Lisker, Liberman, Erickson, Dechovitz, and Mandler (1977), and Kuhl and Miller (1978) describe the conventional method for determining the VOT value of a category boundary: Several monosyllabic samples containing stops are generated, with VOT of those stops ranging from long lead to long lag values (e.g. long lead [ba] to long lag [$p^h$a]). These stimuli are presented to the subjects, who are asked to identify each stimulus as "pa" or "ba" (assuming they are humans, rather than

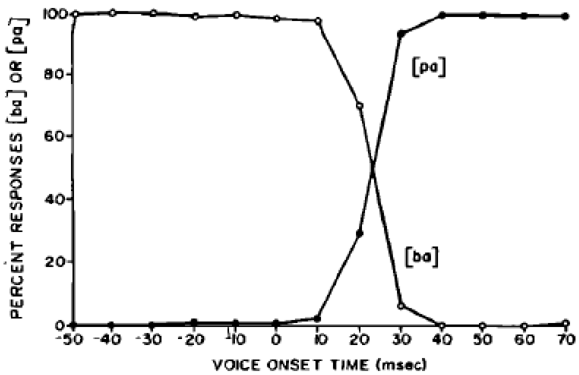


Fig. 1: Response percentage vs. VOT of stimuli (Wood, 1976)

chinchillas; cf. Kuhl and Miller, 1978). The percentage of "pa" and "ba" responses are then plotted against the VOT values for each stimulus, resulting in two curves that transition drastically around the category boundary, as in the example from Wood (1976) shown in **Figure 1**. The value on the horizontal axis at their point of intersection (and, consequently, at a response value of 50% on the vertical axis) is the VOT value corresponding to the /b/-/p/ category boundary. It can be seen that the category boundary for the graph shown in **Figure 1** has a VOT value of approximately 24 ms. The method as described was selected for use in this study.

### Bilingual VOT Perception

Research has also been specifically conducted on Spanish-English bilingual perception of VOT by Ju and Luce (2004) and Elman, Diehl, and Buchwald (1977). Ju and Luce investigated bilingual responses to manipulated VOT values in spoken Spanish words, while Elman and his colleagues studied the effect of language set on bilingual perception. Elman et al. conducted experiments in which bilinguals listened to instructions in either Spanish or English, then participated in perception tests involving VOT-varying stimuli (as previously described) to determine their category boundaries. The researchers found that the bilinguals shifted their category boundaries along the VOT continuum as a function of the language used in the experiment, with the magnitude of the shift positively correlated to the degree of bilingualism.

### Speaker Characteristics

Two final studies relevant to this research are Johnson, Strand, and D'Imperio (1999) and Williams, Whitehead, and Miller (1971). Both groups investigated the influence of speaker characteristics on listener perception. Johnson et al. conducted experiments in presenting videos of female and male faces with manipulated audio to subjects, discovering that listeners modified their perception of vowels in response to speaker gender. Williams et al. tested subjects' responses to videos of ethnically diverse children combined with identical, standard English audio tracks, revealing the existence of ethnic stereotypes in language perception.

## Experiment Overview

The experiment conducted in this research utilized elements from the aforementioned studies. The purpose of the investigation was to obtain data on any shift in the perceptual category boundary of prevocalic stops exhibited by listeners in the presence of speakers with different ethnicities. Two groups of distinct listeners (6 Spanish-English bilinguals and 6 English monolinguals) were presented with stimuli consisting of a series of short, randomly-ordered video clips showing either a Hispanic or an Anglo speaker. The audio sample for each video clip was an ethnically-neutral, synthesized monosyllabic utterance containing a prevocalic bilabial stop. The bilabial stops in the samples varied in VOT values from $-40$ ms (long lead [bo]) to $+40$ ms (long lag [$p^h$o]). Each stimulus thus belonged to one of two sets:

1. A set of audio samples swept in VOT from $-40$

to $+40$ ms with the visual parameter *ethnicity* = Hispanic

2. A set of the same audio samples with the visual parameter *ethnicity* = Anglo

The monolingual English and bilingual listeners performed perception tests by identifying each stimulus as "po" or "bo." The VOT values of their category boundaries were then extracted from their responses and analyzed to ascertain any difference that correlated with the presented ethnicity of each stimuli set. It was hypothesized that the bilingual listeners would exhibit a significant negative shift in their category boundaries from viewing an Anglo speaker to viewing a Hispanic speaker. The monolingual English speakers, lacking experience with Spanish category boundaries, were expected to exhibit no significant shift.

## Method

### Participants

The 12 volunteer participants for this study were all students at North Carolina State University, with ages ranging from 19 to 27 years. Of the 6 Spanish-English bilinguals, 3 were male and 3 were female. 4 bilinguals were ethnically Mexican, 1 was Guatemalan, and 1 was El Salvadorian. Of the 6 English monolinguals, 2 were male and 4 were female, and all were native Anglos. All but one of the English monolinguals had some prior experience with Spanish, but none were at the level of bilingualism.

### Auditory Stimuli

Monosyllabic utterances of the type [bo] were synthesized with 16-bit encoding and 22.5 kHz sampling rate using a Windows implementation of the GnuSpeech Tube Resonance Model.[3] 12 samples were generated, each with a different VOT value for the bilabial stop: $-40$, $-30$, $-20$, $-10$, $-5$, 0, $+5$, $+10$, $+15$, $+20$, $+30$, and $+40$ ms.[4] The pitch of each utterance ranged from 95 to 125 Hz, with the intonation and amplitude contours patterned after [po] as spoken by a bilingual male. The duration of samples with positive VOT was approximately 320 ms (prevoiced utterances were up to 40 ms longer). **Figure 2** shows the waveform and spectrogram of the prevoiced [bo] stimulus with VOT$\approx -40$ ms. The waveform and spectrogram of the stimulus on the opposite end of the VOT spectrum ([$\mathrm{p^h o}$] with long lag VOT$\approx +40$ ms) are shown in **Figure 3**.

### Visual Stimuli

Two video recordings were made of the head and shoulders of two individuals: a 22-year-old Anglo male producing [bo] and a 27-year-old Hispanic male producing [po]. The videos were digitized at 320 x 240 pixels and 24-bit color. The Hispanic video was encoded at 29 frames per second and the Anglo video at 20 frames per second to sync with the audio stimuli. Both videos showed the speaker for approximately 0.5 seconds before the articulators visibly moved, and the final frame of each video was repeated to yield total durations around 4.6 seconds.

### Combined Stimuli

The 12 audio samples were then aligned with the 2 video clips to yield 24 combined stimuli: 12 videos of a Hispanic male producing utterances varying in VOT, and 12 videos of an Anglo male producing the same utterances. This set of 24 stimuli was doubled to 48 to increase response accuracy, and these 48 stimuli were given a random presentation order by the online Research Randomizer.[5]

[3] c.f. Hill, Manzara, and Taube-Schock (1995)

[4] Due to the limitations of this tube model implementation, round values of VOT exactly 5 to 10 ms apart could not be attained. The VOT values were therefore synthesized to be as close as possible, and the exact values were obtained to one-tenth of a millisecond for accuracy in subsequent calculations: $-38.7$, $-30.5$, $-22.2$, $-13.5$, $-6.3$, 0, $+6.2$, $+12.8$, $+14.5$, $+20.9$, $+29.7$, and $+38.3$ ms.

[5] http://www.randomizer.org/

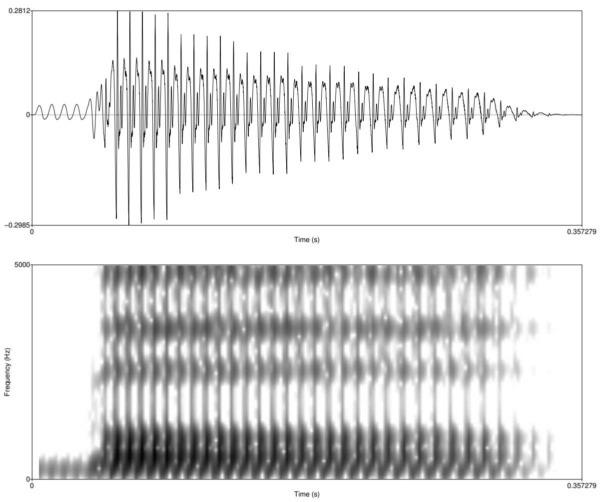


Fig. 2: Long lead [bo] stimulus (VOT = $-40$ ms)

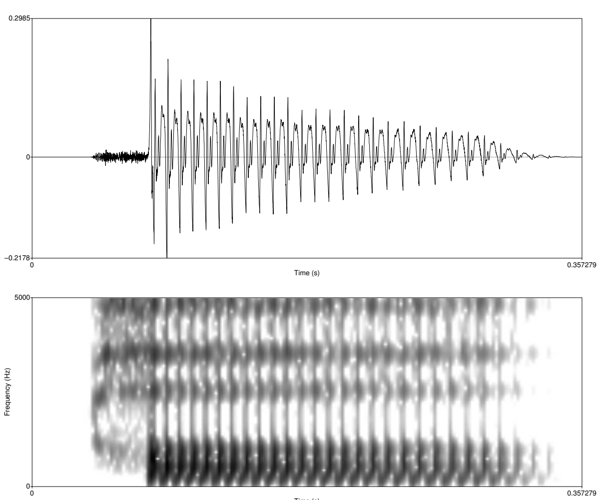


Fig. 3: Long lag [$\mathrm{p^h o}$] stimulus (VOT = $+40$ ms)

### Procedure

This experiment was conducted separately for each of the 12 subjects, who sat in a quiet location and watched the videos on a 15” laptop screen, with headphones adjusted to a comfortable volume level. After viewing a screen with directions in both Spanish and English, the subjects recorded their responses to each stimulus by checking “po” or “bo” on a provided answer sheet. A 3-second silent delay frame displaying the number of the upcoming stimulus was shown between each video clip to allow the subjects time to look down, record their response, and look back at the screen.

## Results

The responses from all individuals to the VOT-varying stimuli were combined into 4 different sets:

1. English listeners’ responses to stimuli with the apparent Anglo speaker
2. English listeners’ responses to stimuli with the apparent Hispanic speaker
3. Bilingual listeners’ responses to stimuli with the apparent Anglo speaker
4. Bilingual listeners’ responses to stimuli with the apparent Hispanic speaker

### English Listeners

The percentages of “po” and “bo” responses for all English listeners were plotted against the VOT for each stimulus with the apparent Anglo speaker. Linear interpolation placed the VOT value of the /b/-/p/ category boundary at 16.63 ms (**Figure 4**), while logistic regression placed it at 15.85 ms (**Figure 5**).

The English listener responses to the apparent Hispanic speaker stimuli were then analyzed in the same manner, with linear interpolation placing the boundary at 13.14 ms (**Figure 6**) and logistic regression placing it at 11.25 ms (**Figure 7**).

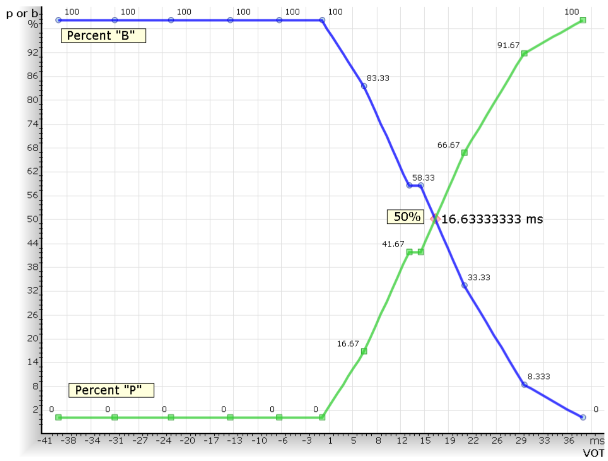


Fig. 4: English listener responses to Anglo speaker stimuli / *Linear Interpolation* / VOT boundary = 16.63 ms

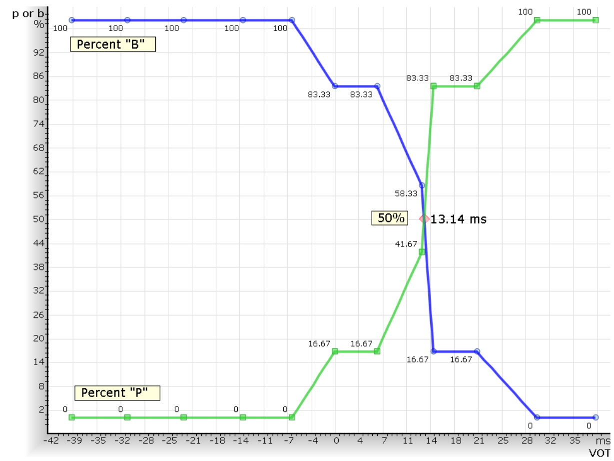


Fig. 6: English listener responses to Hispanic speaker stimuli / *Linear Interpolation* / VOT boundary = 13.14 ms

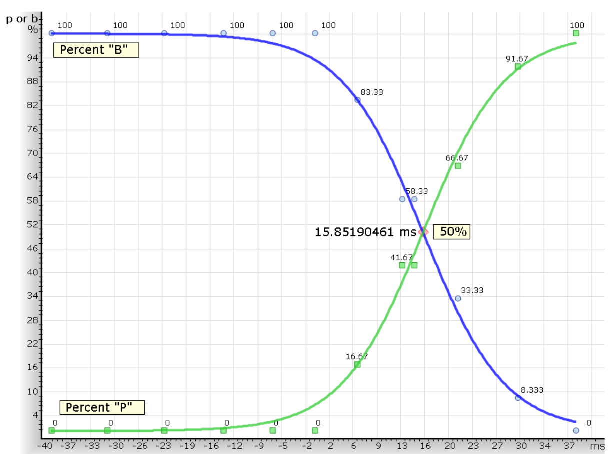


Fig. 5: English listener responses to Anglo speaker stimuli / *Logistic Regression* / VOT boundary = 15.85 ms

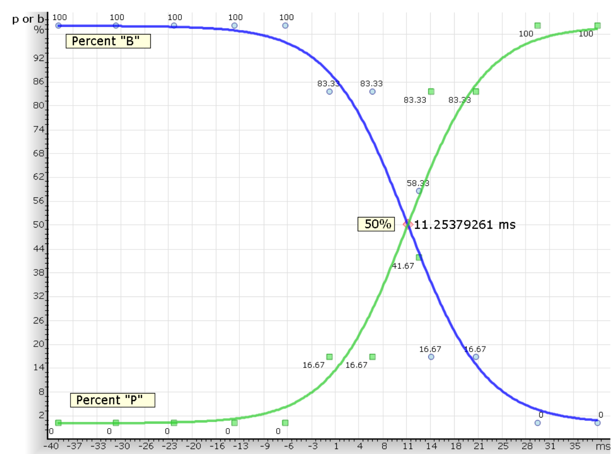


Fig. 7: English listener responses to Hispanic speaker stimuli / *Logistic Regression* / VOT boundary = 11.25 ms

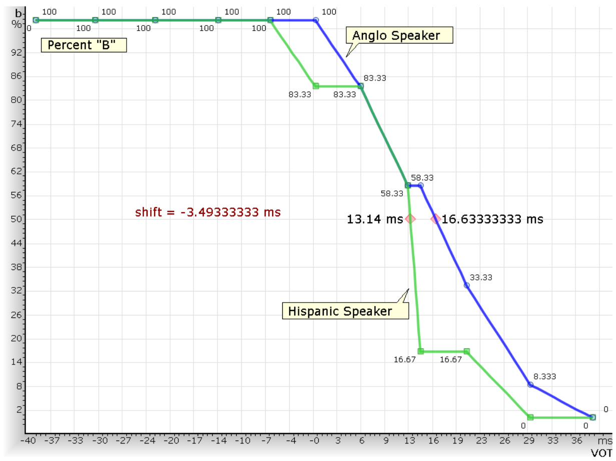


Fig. 8: English listener responses to both speaker stimuli / *Linear Interpolation* / VOT boundary shift = $-3.49$ ms

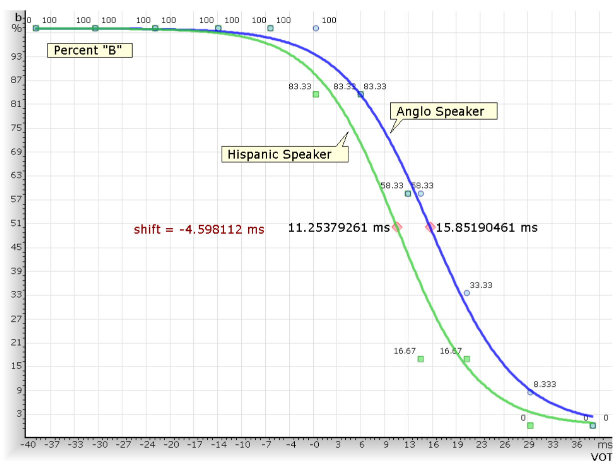


Fig. 9: English listener responses to both speaker stimuli / *Logistic Regression* / VOT boundary shift = $-4.60$ ms

Thus, the English listeners' category boundary shifted to a lower VOT value from the Anglo speaker to the Hispanic speaker. This shift was $-3.49$ ms according to linear interpolation (**Figure 8**) and $-4.60$ ms according to logistic regression (**Figure 9**).

### Bilingual Listeners

The same analysis was applied to the responses from the bilingual listeners.[6] Linear interpolation yielded a VOT value for 7.52 ms for the apparent Anglo speaker, while logistic regression placed it at 5.23 ms. For the apparent Hispanic speaker, linear interpolation placed the VOT boundary at 4.65 ms, while logistic regression yielded 4.75 ms. The shift in VOT value of the category boundary from the Anglo speaker to the Hispanic speaker was therefore $-2.87$ ms according to linear interpolation and $-0.48$ ms according to logistic regression.

### Individual Average

To provide an additional means of looking at the data, the responses of each listener were analyzed separately to pinpoint that individual's category boundaries. The individual VOT boundaries in response to both apparent speakers were then averaged for each group of listeners, yielding 17.98 ms (Anglo speaker) and 10.79 ms (Hispanic speaker) for the English listeners (a shift of $-7.19$ ms) and 5.61 ms (Anglo speaker) and 6.36 ms (Hispanic speaker) for the bilingual listeners (a shift of **+**0.75 ms). The results of all methods are presented in **Tables 1-3**.

## Discussion

**Table 4** summarizes the VOT boundary shifts from the three methods and presents a 3-way average of the results. As can be seen, the group of bilingual listeners showed no significant boundary shift in comparison with the group of English listeners. The presence of a positive average

[6] The bilingual graphs are similar to those for the English listeners and are omitted for the sake of space.

| | English listeners | Bilingual listeners |
|---|---|---|
| Anglo speaker VOT boundary (ms) | 16.63 | 7.52 |
| Hispanic speaker VOT boundary (ms) | 13.14 | 4.65 |
| VOT boundary shift (ms) | **-3.49** | **-2.87** |

Tab. 1: Linear Interpolation results

| | English listeners | Bilingual listeners |
|---|---|---|
| Anglo speaker VOT boundary (ms) | 15.85 | 5.23 |
| Hispanic speaker VOT boundary (ms) | 11.25 | 4.75 |
| VOT boundary shift (ms) | **-4.60** | **-0.48** |

Tab. 2: Logistic Regression results

| | English listeners | Bilingual listeners |
|---|---|---|
| Anglo speaker VOT boundary (ms) | 17.98 | 5.61 |
| Hispanic speaker VOT boundary (ms) | 10.79 | 6.36 |
| VOT boundary shift (ms) | **-7.19** | **+0.75** |

Tab. 3: Individual Average results

| | English listeners | Bilingual listeners |
|---|---|---|
| Linear boundary shift (ms) | -3.49 | -2.87 |
| Logistic boundary shift (ms) | -4.60 | -0.48 |
| Individual Average boundary shift (ms) | -7.19 | +0.75 |
| **3-way Method Average (ms)** | **-5.09** | **-0.87** |

Tab. 4: VOT boundary shifts from Linear, Logistic, and Individual Average methods

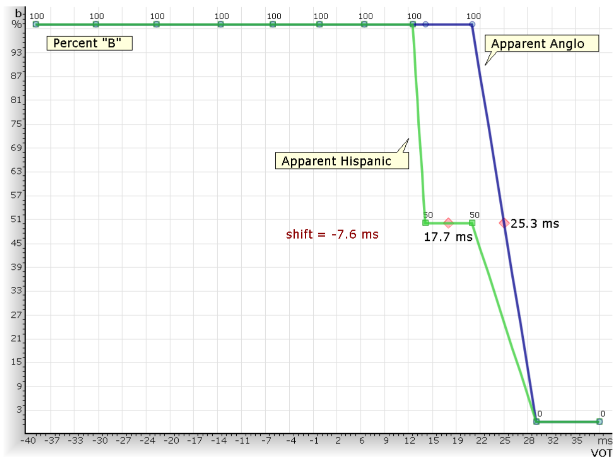


Fig. 10: English monolingual w/0 years of Spanish, shift = $-7.6$ ms

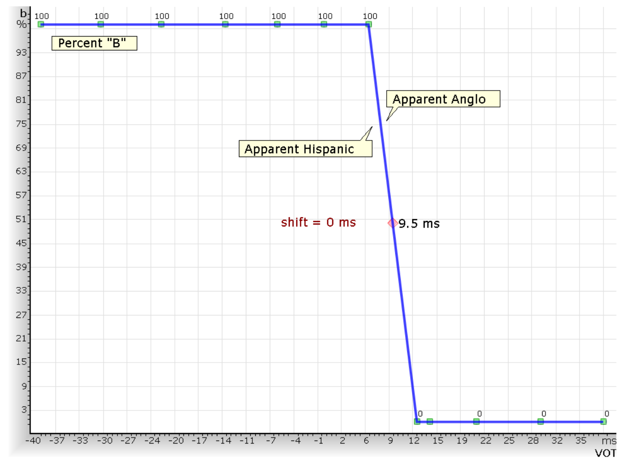


Fig. 11: Bilingual w/17 years of English, shift = 0 ms

boundary shift of +0.75 ms from one of the methods (**Table 3**) also indicates that some bilinguals exhibited a greater VOT boundary value for the apparent Hispanic speaker than for the apparent Anglo speaker. Given that the VOT value of the /b/-/p/ perceptual category boundary in Spanish is less than that of the same category boundary in English, these were unexpected results.

Another unusual finding was the consistent negative shift exhibited by the English listeners as illustrated in **Table 4**. The English listeners responded to the visual stimuli in a manner that was expected only of the bilinguals. The response of one such subject is shown in **Figure 10** as an example of this anomaly. This 20-year-old monolingual English female had no Spanish experience and yet exhibited a VOT shift in her category boundary of $-7.6$ ms from the Anglo speaker to the Hispanic speaker. In contrast, a 21-year-old bilingual male exhibited no VOT shift whatsoever (**Figure 11**) despite his 17 years of English (L2) experience.

One interpretation of these results leads to the conclusion that the factor of speaker ethnicity must reside under the level of consciousness required to trigger a category boundary shift. Elman et al. (1977) demonstrated that a change in language set can cause such a shift in bilinguals, thereby distinguishing that factor as an upper bound on triggering a perceptual switch. A change in apparent speaker ethnicity, on the other hand, caused no boundary shift in bilinguals and therefore acts as a lower bound on the level of consciousness required for such a trigger.

To explain why the English listeners did indeed shift, it may be conjectured that listeners with no Spanish experience employ stereotypes of Hispanic English (e.g. Speedy Gonzales) in speech perception. Hence, although the English listeners did not possess the necessary linguistic experience with Spanish to be aware, either consciously or subconsciously, of the true difference in category boundaries, they were still able to utilize a stereotype of the Spanish /p/ as sounding like the English [b]. Assuming this is true, it appears that the bilinguals, all of whom were Latin Americans and who failed to exhibit the same shift, operated under no similar stereotypes of American English. This is not a surprising conclusion, given that popular caricatures of Hispanic English speakers

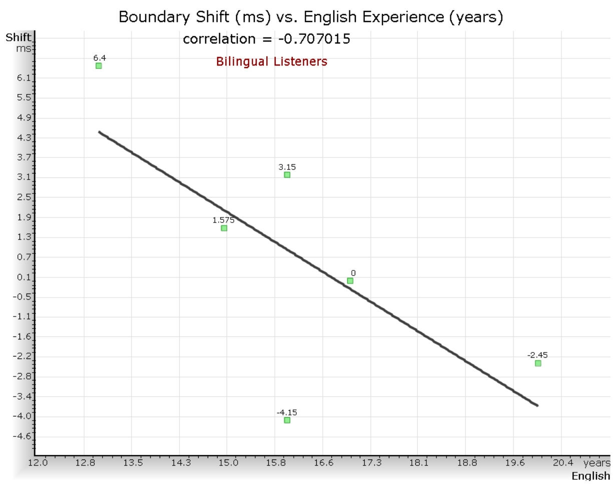


Fig. 12: Bilingual listeners: Boundary shift (ms) vs. English experience (years), correlation = $-0.707$

(Speedy Gonzales, Cheech Marin, Chico and the Man, etc.) are far more prominent than caricatures of American English speakers.

Another interpretation of the bilingual behavior arises when the boundary shift of each listener is analyzed with respect to that individual's L2 experience. For the English listeners, there was no significant correlation between the amount of negative boundary shift and their limited years of experience with Spanish. The bilingual listeners, on the other hand, exhibited a correlation of $-0.707$ between the amount of boundary shift and their years of experience with English, as demonstrated in **Figure 12**. That is, the more experience the bilinguals had with English, the more likely they were to shift their category boundary in the direction expected (toward smaller VOT from Anglo speaker to Hispanic speaker). There are at least two possible explanations for this behavior:

1. Apparent speaker ethnicity can indeed trigger a perceptual shift in bilingual individuals, but one which is only significant in listeners who possess extensive experience (>20 years) with the L2 language.
2. Spanish bilinguals develop stereotypes of American English as they acquire more experience with it.

However, more data is needed to verify this finding before a suitable explanation can be determined.

## Corroborative Results

Flege (1991) examined VOT in stops produced by Spanish-English bilinguals and reports that early L2 learners of English exhibit VOT values comparable to those of monolingual English speakers, while late L2 learners produce "compromise" VOT values between those of Spanish monolinguals and English monolinguals. The current research found this connection between age of learning and VOT to exist in perception boundaries as well. As shown by the data in **Tables 1-3**, the average English monolingual category boundary for the apparent Anglo speaker is around VOT = 17 ms. An analysis of the bilingual category boundaries for the same speaker ethnicity showed a strong correlation ($-0.888$) between the VOT boundary value and the age at which English language learning began (**Figure 13**). That is, the earlier a bilingual listener was introduced to English, the closer his/her VOT category boundary was to that of an English monolingual. The most dramatic example of this was a 20-year-old Mexican female bilingual who began learning English at birth and exhibited a category boundary of VOT = 15.25 ms, which is essentially an English monolingual value.

## Improvements

Several issues arose during this research that revealed areas upon which improvements could be made, were this experiment to be repeated.

One subject early in the process reported perceiving the Anglo speaker as Hispanic. The possible impact of this confusion was circumvented by confirming the ability of the other subjects to distinguish the ethnicity of the speakers before con-

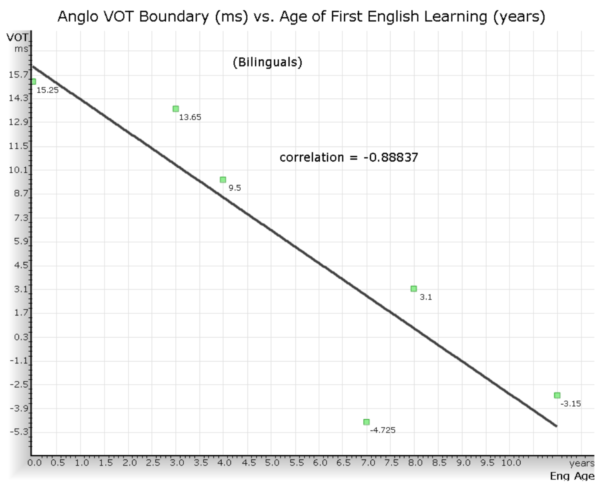


Fig. 13: Bilingual listeners: Anglo VOT boundary (ms) vs. Age of first English learning (years), correlation = $-0.888$

ducting the experiments. This may have affected the results, however, by causing the listeners to give conscious attention to speaker ethnicity instead of allowing it to act as a subconscious trigger. It is therefore recommended that a repeat of this research follow the example of Johnson et al. (1999) and conduct a separate experiment to determine which speakers are perceived as "most Anglo appearing" or "most Hispanic appearing." The choice of speakers for the stimuli can then be made on the basis of those results.

Subjects also reported attempting to identify the utterances by watching the mouths of the speakers. According to Lubker and Parris (1970), intraoral breath pressure can be used to reliably discriminate /p/ and /b/. It is therefore conceivable that listeners utilize cheek expansion of the speaker as an aid for making this discrimination, which is a cue that did not vary across the videos used in this experiment. To prevent listeners from relying on this invariant cue, the mouths of the speakers should be obscured while preserving their ethnic identities.

Previous research involving VOT-varying stimuli have utilized various implementations of a parallel-resonance synthesizer, which gives better control over formants and VOT than the tube model. Future researchers would be well-advised to follow this tradition and purchase such a synthesis package for their experiments.

Several participant-related changes would also be beneficial:

- Test a larger number of participants to obtain more reliable statistics.
- Test a third group of listeners: monolingual Spanish individuals.
- Restrict the monolingual English group to individuals with no Spanish instruction whatsoever.
- Develop a better test for an individual's level of bilingualism than "years of experience with the L2 language."

Another improvement to the experiment could be accomplished by increasing the amount of individual sample repetition in the stimuli, yielding more precise listener responses than the 100%, 50%, and 0% obtained from presenting each stimulus only twice (e.g. **Figure 10**).

A final consideration should be given to the language context in which the experiments are conducted (c.f. Elman et al., 1977). The principal investigator in this research communicated with the subjects in English, although instructions in both Spanish and English were presented in the videos in an attempt to prevent biasing the results toward a specific language. However, a better strategy would employ a bilingual investigator to conduct the experiments with the bilingual listeners. By fostering a dual-language atmosphere with the listeners, the investigator can better avoid biasing the subjects toward a specific language and thereby skewing the data.

**Acknowledgments**

Special thanks to Dr. Erik Thomas and Phillip Carter of NC State University for their expert assistance in this research. The author is also indebted to Tanner Leggett, Santiago Ortega, and Cassandra Neher for their aid in designing and conducting the experiments.

### References

Abramson, A.S. and Lisker, L. (1970) Discriminability along the voicing continuum: cross-language tests. *Proceedings of the 6th International Congress of Phonetic Sciences*, 569-573. Prague: Academia.

Benkí, J. R. (2005) Perception of VOT and first formant onset by Spanish and English speakers. *ISB4: Proceedings of the 4th International Symposium on Bilingualism*, 240-248. Somerville, MA: Cascadilla Press.

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

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.

Flege, J. E. (1991) Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language. *Journal of the Acoustical Society of America*, **89**, 395-411.

Hill, D., Manzara, L., and Taube-Schock, C. R. (1995). Real-time articulatory speech-synthesis-by-rules. *Proceedings of AVIOS '95*, 27-44. San Jose.

Johnson, K., Strand, E. A., and D'Imperio, M. (1999) Auditory-visual integration of talker gender in vowel perception. *Journal of Phonetics*, **27**, 359-384.

Ju, M. and Luce, P. A. (2004) Falling on sensitive ears: constraints on bilingual lexical activation. *Psychological Science*, **15**, 314-318.

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.

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

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.

Lubker, J. F. and Parris, P. J. (1970) Simultaneous measurements of intraoral pressure, force of labial contact, and labial electromyographic activity during production of the stop consonant cognates /p/ and /b/. *Journal of the Acoustical Society of America*, **47**, 625-633.

Williams, F., Whitehead, J. L., and Miller, L. M. (1971) Ethnic stereotyping and judgments of children's speech. *Speech Monographs*, **38**, 166-170.

Wood, C. (1976) Discriminability, response bias, and phoneme categories in discrimination of voice onset time. *Journal of the Acoustical Society of America*, **60**, 1381-1389.