



How Auditory Discontinuities and Linguistic Experience Affect the Perception of Speech and Non-Speech in English- and Spanish-Speaking Listeners

Jessica F. Hay¹, Lori L. Holt², Andrew J. Lotto³, & Randy L. Diehl¹

¹University of Texas at Austin, ²Carnegie Mellon University, ³Boys Town National Research Hospital

INTRODUCTION

- Speech perception can be viewed as an interplay between linguistic experience and perceptual biases present in the mammalian auditory system.
- As a result of linguistic experience, languages have different mappings between voice-onset-time (VOT) and voicing categories [1].
- In English, [-voice] and [+voice] labial stops have a category boundary at approximately +20 ms VOT.
- Young infants [2] and animals [3] also display boundary-type behavior at +20 ms VOT.
- Research using a non-speech analogue of VOT, called tone-onset-time (TOT), suggests that there is an underlying auditory basis for the perception of stop consonants based on a threshold for detecting onset asynchronies in the vicinity of ±20 ms [4].
- Both young infants [5] and English-speaking adults [4] have two regions of heightened sensitivity near ±20 ms TOT.
- It would be communicatively advantageous for languages to exploit these regions of heightened sensitivity (also called auditory discontinuities) in their sound inventories by placing stop consonant categories on either side of a discontinuity.
- Not all languages exploit these auditory discontinuities (ADs) in their stop consonant categories.
- In Spanish, for example, [-voice] and [+voice] labial stops have a category boundary at approximately 0 ms VOT, and the two ADs fall within each of the voicing categories [6].
- The present study addresses the question of whether long-term linguistic experience with VOT categories affects the perception non-speech stimuli that are analogous in their acoustic timing characteristics.

OVERVIEW OF THE STUDY

- Speakers of two languages, English and Spanish, that differ in their mappings of VOT onto their voicing categories were tested in three perceptual tasks:
 - VOT Category Labeling
 - VOT ABX Discrimination
 - TOT ABX Discrimination
- Category labeling boundaries and discrimination peaks were assessed for English and Spanish listeners.
- Differences between language groups are interpreted as the result of linguistic experience.

METHOD

Participants

- 24 monolingual English speakers (12 male and 12 female) from University of Texas at Austin, USA
- 23 monolingual Spanish speakers (11 male and 12 female) from the University of Puebla, Mexico

Stimuli

VOT Stimuli

- A slightly modified version of the Klatt (1988) synthesizer was used to create a series of 23 labial VOT stimuli.
- VOT stimuli varied in 10 ms steps between -70 ms and +70 ms except in the region of -40 ms to +40 ms, where stimuli were varied in 5 ms steps.
- Variation in positive VOT stimuli was created by manipulating the amplitude of voicing (AV), aspiration (AH), and frication (AF).
- Variation in negative VOT stimuli was created by appending the 0 ms VOT stimulus to a low frequency voicing bar.

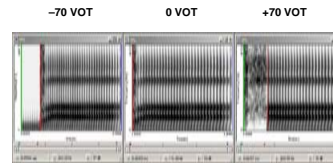


Figure 1: Spectrograms of synthetic VOT stimuli. The onset of voicing relative to the release of the consonant is varied to create a VOT continuum from -70 ms to +70 ms VOT.

TOT Stimuli

- TOT stimuli were modeled after those used by Pisoni (1977), see Figure 2 for a schematic representation.
- TOT stimuli consisted of 2 sine-wave tones where the onset of the lower frequency tone (500 Hz) was varied relative to the onset of the 230 ms higher frequency tone (1500 Hz).
- The duration of the lower frequency tone was varied relative to the higher tone to create a TOT series that ranged from -70 ms to +70 ms TOT, in 5 ms steps.
- The intensity of the higher tone was 12dB lower than the lower tone.

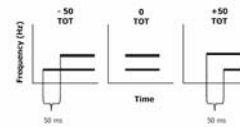


Figure 2: Schematic spectrograms of three representative TOT stimuli. The onset of the lower frequency tone (500 Hz) is varied relative to the onset of the higher frequency tone (1500 Hz) to create a continuum from -70 ms to +70 ms TOT. TOT stimuli are modeled after those of Pisoni [4] and the graph is taken from Holt et al. [6].

Procedure

(1) VOT Category Labeling

- Participants were asked to categorize the series of synthetic /ba-/pa/ stimuli (using language-appropriate categories).
- Each of the 23 stimuli was presented 10 times, for a total of 230 randomly ordered trials.

(2) VOT ABX Discrimination

- Participants were asked to do an ABX discrimination task (with feedback) on the series of synthetic VOT stimuli ranging from -70 ms to +70 ms VOT.
- There was always a VOT difference of 30 ms between stimulus A and B.
- In each block there were 12 pairs of stimuli, with four orders of counterbalancing (ABA, ABB, BAA, BAB). Each stimulus was presented twice, for a total of 96 trials per block.
- Participants completed two blocks of trials.

(3) TOT ABX Discrimination

- All procedures were identical to those described in the VOT discrimination section, except that TOT instead of VOT stimuli were used.
- The order of VOT and TOT discrimination tasks was counterbalanced between participants.

DISCUSSION

- Consistent with previous results, both English and Spanish speakers have VOT discrimination peaks in the vicinity of their 50% cross-over VOT labeling boundary [1,7].
- We additionally found that English speakers have TOT discrimination functions that resemble their VOT discrimination functions.
- Spanish speakers, on the other hand, have very different VOT and TOT discrimination functions, with their VOT discrimination peak falling between their two TOT discrimination peaks.
- Results suggest that Spanish speakers' discrimination of TOT may be unaffected by their linguistic experience.
- We suggest that Spanish speakers may not rely on VOT to distinguish voicing categories, but instead may use a qualitative cue such as detecting the presence or absence of voicing at the release of the consonant (see Green et al. [8] and Keating [9] for similar suggestions).
- Our results support the idea put forth by Holt et al. [6], that learning can overcome psychoacoustic boundaries, or ADs.

RESULTS

VOT Category Labeling Boundary

- Linear regression analysis was performed on the data points that ranged from the last 100% /b/ identification response to the first 100% /p/ identification response.
- The slope and 50% cross-over point of the regression line were analyzed in a MANOVA, with Language as an independent variable.
- English speakers had a significantly longer identification boundary (16.63 ms) than Spanish speakers (0.84 ms) [$F(1,45)=318.421, p<.001$].
- There was a non-significant trend for English speakers to have a steeper boundary than Spanish speakers ($p=.052$).

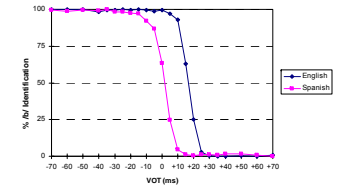
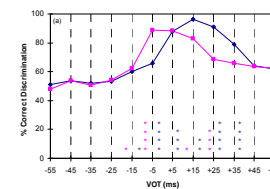


Figure 3: VOT identification functions for English and Spanish speakers.

VOT Discrimination Peaks

- A 2 (Language) x 12 (Stimulus) repeated measures ANOVA revealed a significant interaction between Language and Stimulus on discrimination performance [$F(11,495)=10.671, p<.001$].
- Figure 4 (a) shows that English speakers have a discrimination peak in the vicinity of +15 ms VOT.
- Spanish speakers have a discrimination peak in the vicinity of 0 ms VOT.
- Speakers of both languages show VOT discrimination peaks in the same region as their labeling boundaries.



TOT Discrimination Peaks

- A 2 (Language) x 12 (Stimulus) repeated measures ANOVA revealed a significant interaction between Language and Stimulus on discrimination performance [$F(11,495)=2.223, p<.02$].
- Figure 4 (b) shows that English speakers have a broad discrimination peak in the vicinity of +15 ms TOT.
- Spanish speakers have two small discrimination peaks in the vicinity of ±15 ms TOT.

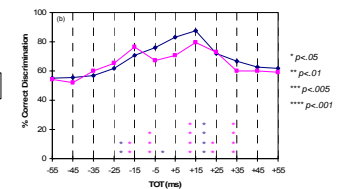


Figure 4: (a) VOT discrimination functions for English and Spanish speakers, and (b) TOT discrimination functions for English and Spanish speakers. Significant differences between adjacent data points are indicated for each language.

REFERENCES

[1] Abramson, A. S. and Lisker, L. (1970). "Discriminability along the voicing continuum: Cross-language tests." In *Proceedings of the Sixth International Congress of Phonetics Sciences*, Prague, 1967. Prague, Czechoslovakia: Academia.

[2] Eimas, P.D., Siqueland, E.R., Jusczyk, P., and Vigorito, J. (1971). "Speech perception in infants." *Science*, 171, 303-306.

[3] Kuhl, P.K. (1981). "Discrimination of speech by non-human animals: Basic auditory sensitivities conducive to the perception of speech-sound categories." *J. Acoust. Soc. Am.*, 70(2), 340-349.

[4] Pisoni, D.B. (1977). "Identification and discrimination of the relative onset of two component tones: Implications for voicing perception in stops." *J. Acoust. Soc. Am.*, 61(5), 1552-1561.

[5] Jusczyk, P.W., Pisoni, D.B., Wailey, A., and Murray, J. (1980). "Discrimination of relative onset time of two-component tones by infants." *J. Acoust. Soc. Am.*, 67(1), 282-270.

[6] Holt, L.L., Lotto, A.J., and Diehl, R.L. (2004). "Auditory discontinuities interact with categorization: Implications for speech perception." *J. Acoust. Soc. Am.*, 116(3), 1763-1773.

[7] Williams, L. (1977). "The voicing contrast in Spanish." *J. Phonetics*, 5, 169-184.

[8] Green, K.P., Zampini, M.J., and Clarke, C. (1988). "The role of preceding closure interval and voice onset time in the perception of voicing: A comparison of English and Spanish-English bilinguals." *J. Acoust. Soc. Am.*, 104, 1835.

[9] Keating, P.A., Mack, M.J., and Gannon, W.F. (1981). "A cross-language study of range of voice onset time in the perception of initial stop voicing." *J. Acoust. Soc. Am.*, 70(5), 1261-1271.

ACKNOWLEDGEMENTS

* This work was supported by research grant R01 DC00427-15-16, from the National Institutes of Health, National Institute on Deafness and Other Communication Disorders, to the 4th author, and by Grant No. 0345773, from the National Science Foundation, to the 2nd, 3rd, and 4th authors. Special thanks to Angel Melo of the University of Tlaxcala, Mexico, Sidney Gerardo Reyes, and Cheryl Moran.