Effects of linguistic experience on perception and learnability of non-speech categories

Jessica F. Hay[1] *& Adrian Garcia-Sierra*[2]

[1]Department of Psychology
[2]Department of Communication Sciences and Disorders
University of Texas at Austin
hay@psy.utexas.edu

Abstract

The present study addresses whether linguistic experience affects the learnability of non-speech categories and in turn whether learning non-speech categories affects speech perception. Speakers of English and Spanish, languages that have different mappings between voice-onset-time (VOT) and their voicing categories, were trained to categorize a series of analogous non-speech (tone-onset-time) sounds into two groups. The boundaries between the distributions of tone-onset-time (TOT) stimuli were either (1) consistent with the negative auditory discontinuity (–CC), (2) consistent with the positive auditory discontinuity (+CC), or (3) inconsistent both auditory discontinuities (IC), along the VOT/TOT dimension. Results suggest that linguistic experience does not play a role in the learnability of these non-speech categories. While language experience affects VOT category labeling boundaries, short-term experience with TOT categorization does not appear to influence VOT boundary locations. Speech and non-speech appear to be processed independently. The relationship between auditory discontinuities and learnability is discussed.

1. Introduction

Experience with a native language causes a drastic re-organization of speech sounds into language-specific categories; such that listeners from two languages may hear the same acoustic signal, yet perceive it as belonging to different categories. The representation of language-specific speech categories in the adult brain is likely the result of both biases present in the mammalian auditory system and experience with distributions of speech sounds in the environment. Biases (or auditory discontinuities) are evident when there are perceptual non-linearities in the mapping from acoustic signals to auditory responses, resulting in heightened sensitivity in a particular region of acoustic space. The presence of auditory discontinuities (ADs) is evident in both young infants [1] and non-human mammals [2] and is believed to provide the basis for early discrimination abilities in infants. The perceptual difference between sounds that straddle an AD can be expected to be exaggerated relative to pairs of sounds that do not have an AD between them. Thus, it may be communicatively advantageous for languages to exploit this heightened perceptual distinctiveness in their sound inventories. In order for the presence of ADs to provide an advantage, however, it must be the case that their presence makes categories easier to learn.

A recent study by Holt and colleagues [3] addressed this issue. They trained English-speaking listeners to assign distributions of tone-onset-time (TOT) sounds to two categories. The first group of participants was presented with distributions of TOT stimuli where the category boundary was consistent with the AD at +20 ms. (English-speakers have been shown to have two discrimination peaks, one at –20 ms TOT and one at +20 ms TOT [4].) The second group of participants was presented with distributions of TOT stimuli, with the category boundary at approximately +40 ms TOT. (The AD, at +20 ms, fell within one of the categories.) In the Consistent Condition, the AD at the distributional boundary was predicted to reinforce the category distinction. In the Inconsistent Condition, the AD, which was in the middle of one of the categories, was predicted to hinder category learning. Holt et al. found that found that participants in the Consistent Condition required fewer blocks to reach 90% correct categorization than did those in the Inconsistent Condition. This was the first direct demonstration that the presence of ADs at category boundaries enhances learning, and it suggests that it would in fact be communicatively advantageous for languages to place their phonetic category boundaries in the vicinity of ADs.

The present study is a partial replication of the Holt et al. study, but it addresses a number of additional questions. (1) Does linguistic experience affect the learnability of non-speech categories? Although some speech sounds (e.g., VOT) can be represented in terms of non-speech analogues (e.g., TOT), it has generally been assumed that experience in one domain does not affect processing in the other. TOT perception has never been measured in non-English-speakers. Listeners from two language groups, English and Spanish, who have different mapping between VOT and their voicing categories, were tested using the Holt et al. training paradigm. (2) Are auditory discontinuities symmetrical in their effect on category learning? Researchers have found both positive and negative ADs in VOT/TOT perception [4, 1], consistent with cross-linguistic distributions of stop categories, but it is unclear whether they are equally perceptually salient. Listeners were trained on one of three possible distributions: (a) a distribution that is consistent with the negative auditory discontinuity (–CC), (b) a distribution that is consistent with the positive auditory discontinuity (+CC), or (c) a distribution that is inconsistent both auditory discontinuities (IC). We expect that our listeners will learn the consistent distributions more quickly than the inconsistent distribution, and any differences between the two consistent distributions will reflect the relative perceptual salience of the ADs. (3) Does learning non-speech categories affect speech perception? Prior to and following training, participants were asked to label a series of VOT stimuli, providing a measure of the effect of non-speech category training on VOT labeling boundary location. Results have implications for the effects of linguistic experience on auditory processing and learnability, and also address whether short-term experience with non-speech stimuli can affect speech perception.

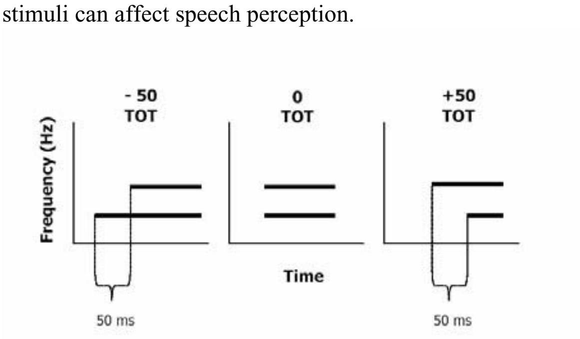

Figure 1: Schematic spectrograms of three representative TOT stimuli. The onset of lower frequency tone (500 Hz) is varied relative to the onset of the higher frequency tone (1500 Hz) to create a continuum from −*70 ms to +70 ms TOT. TOT stimuli are modeled after those of Pisoni* [4] *and the graph is taken from Holt et al.* [3].

2. Methods

2.1. Participants

Forty-eight native English speakers and forty-eight native Spanish speakers served as participants in this study. Sixteen participants from each language background participated in each of the three training conditions. English-speakers were recruited from the Psychology Department at the University of Texas at Austin. Spanish-speakers were recruited from the Biology Department at the University of Puebla, in Puebla, Mexico. Participants were paid $12 per hour, or its equivalent, for their participation.

2.2. Stimuli

2.2.1. VOT Stimuli

A slightly modified version of the Klatt [5] synthesizer was used to create a series of 23 VOT stimuli that varied in 5 ms steps between –40 ms and +40 ms VOT and in 10 ms steps between ±40 ms and ±70 ms VOT. Labial CV syllables were synthesized at a sampling rate of 11025 Hz. Variation in positive VOT stimuli (i.e., from 0 ms to +70 ms) was created by manipulating the amplitude of voicing (AV), aspiration (AH), and frication (AF). The release of the consonant was mimicked with a 10 ms burst at the onset of each stimulus, which was created by adding 10 ms of frication to the beginning of the signal. Voiceless stimuli were created by turning off voicing, turning on aspiration, and increasing the bandwidth of F1 from 90 Hz during voicing to 300 Hz during the period of voicing lag. Variation in negative VOT stimuli (i.e., from −5 ms to −70 ms) was created by appending the 0 ms VOT stimulus to a low frequency voicing bar with an F0 of 120 Hz. The voicing bar was created by setting F1 at 180 Hz with a bandwidth of 150 Hz. All voiced stimuli were 295 ms long. Pre-voiced stimuli ranged from 300 ms to 365 ms in duration.

2.2.2. TOT Stimuli

TOT stimuli were modeled after those used by Pisoni [4] and consisted of two sine-wave tones where the onset of the lower frequency tone was varied relative to the onset of the higher frequency tone. Both tones had a constant frequency of 500 Hz and 1500 Hz, respectively. The duration of the 1500 Hz tone was held constant at 230 ms. The duration of the lower frequency tone was varied relative to the higher frequency tone in order to create a series of TOT stimuli that ranged from −70 ms to +70 ms in 5 ms steps. In order to mimic the intensity profile of natural VOT syllables, the intensity of the higher tone was 12dB lower than that of the lower tone. During the first 5 ms of each of the individual tones, the amplitude was ramped upward linearly. Individual tones were created and combined in Adobe® Audition™ 1.0 at a sampling rate of 22050 Hz and were then re-sampled at 11025 Hz.

2.3. Procedure

Participants were run in five consecutive tasks in the following order: (1) VOT Categorization, (2) TOT Category Training, (3) TOT Generalization, (4) TOT Category Refresher, and (5) VOT Categorization.

2.3.1. VOT Categorization

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

2.3.2. TOT Category Training

Participants were run in one of three training conditions: (1) +Consistent Condition (+CC) – the distributions of TOT sounds straddled +22.5 ms TOT, (2) Inconsistent Condition (IC) – the distributions of TOT sounds straddled –2.5 ms TOT, (3) –Consistent Condition (–CC) – the distributions of TOT sounds straddled –22.5 ms TOT. Within each block of trials participants were presented with a total of 46 stimuli, according to the distributions shown in Figure 2. Following the presentation of each token, participants were asked to decide whether that token belonged to category A or C. Feedback was given after every trial. Participants were deemed to have learned the categories once they reached the criterion categorization performance of 90% correct. After participants reached criterion performance, they were presented with two additional blocks of trials.

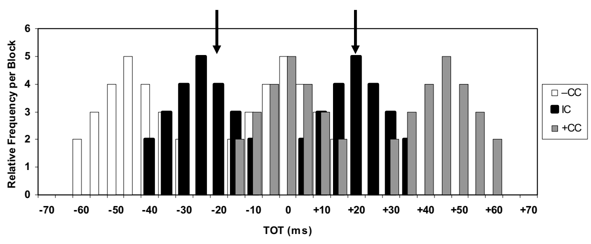

Figure 2: Stimulus input distributions reflecting the relative frequency of stimulus presentation as a function of TOT. Arrows represent the locations of ADs. −*CC and +CC distributional training conditions are consistent with respective to the ADs. The IC training condition is inconsistent with ADs.*

2.3.3. TOT Generalization

Upon completion of the TOT category training, participants completed a TOT generalization test. The generalization test was included in order see how auditory discontinuities affect category labeling boundaries in TOT stimuli. Participants were randomly presented with the entire set the 29 TOT stimuli, 10 times each and were asked to categorize the stimuli into category A or C using the same criteria they applied during the training task. No feedback was given.

2.3.4. TOT Category Refresher

Before being tested on whether TOT training condition affects VOT category boundary locations, each participant completed a TOT category refresher consisting of two additional blocks of TOT category training as described in the TOT Category Training section.

2.3.5. VOT Categorization

Following completion of the TOT category refresher, participants were again asked to categorize a series of synthetic /ba/-/pa/ stimuli ranging from –70 ms VOT to +70 ms VOT. All procedures were identical to those described in *2.3.1*.

3. Results

3.1. TOT Training

A 2 (Language) x 3 (Training Condition) ANOVA was performed to examine the effects of both linguistic experience and the type of training condition on the number of blocks it took participants to learn their respective TOT categories. Overall, English-speaking listeners learned their categories in fewer blocks that did the Spanish-speaking listeners [F(1,90)=6.409, p<.02]. There was also a main effect of Training Condition [F(2,90)=4.373, p<.02]; Planned comparisons revealed that although +CC was learned in significantly fewer blocks that IC (*p<.05*), there was no significant difference between –CC and IC (p=.528). (See Figure 3).

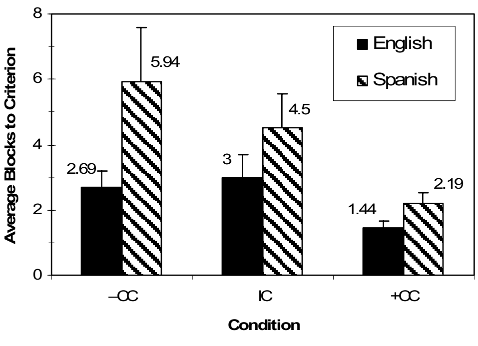

Figure 3: Effect of Language and Training Condition on the number of blocks to criterion performance by English- and Spanish-speaking listeners. Error bars represent the standard error of the mean.

3.2. TOT Generalization

The data that ranged from the last 100% category A response to the first 0% category A response were extracted from each generalization function for the purposes of linear regression analysis.[1] The slope and 50% cross-over point of the resulting regression line were analyzed in a MANOVA, with Language and Training Conditions as independent variables. There was a significant main effect of Training Condition on both TOT boundary location [F(2,89)=32.823, *p<.001*] and on the slope of the boundary [F(2,89)=3.885, *p<.03*]. Participants in +CC had a higher TOT boundary value than those in IC, which in turn had a higher boundary value that those in –CC (all *p's<.05*). Generally the slope of the boundary was more shallow for participants in the –CC training condition than for participants in either of the other conditions (*p<.001*). (See Figure 4).

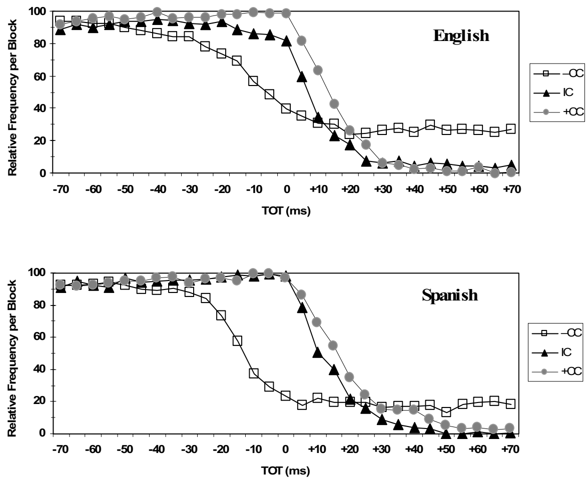

Figure 4: TOT Generalization functions for English- and Spanish-speaking listeners in each of the three training conditions.

3.3. VOT Categorization

Linear regression analysis was performed on the data that ranged from the last 100% /b/ identification response to the first 0% /b/ identification response. The slope and the 50% labeling cross-over boundary of the regression line were analyzed in a MANOVA, with Language, Trial (before vs after training) and Training Condition as independent variables. As expected, there was a significant main effect of Language on VOT boundary location [F(1,180)=392.019, *p<.001*]; English- and Spanish-speaking listeners had average VOT labeling boundaries of 13.87 ms and 1.71 ms, respectively. There was a significant main effect of Trial on the slope of the labeling function [F(1,180)=4.457, *p<.05*]; participants had steeper VOT labeling slopes before the training phase than afterwards. There were no Language x Training Condition or Language x Training Condition x Trial interactions, suggesting that training on the categorization of TOT stimuli does not affect the perception of VOT stimuli. Interestingly, there was a significant Language x Trial interaction [F(1,1800=8.966, *p<.005*]; English-speaking listeners showed a decrease in the VOT value of their labeling boundary following TOT training, whereas Spanish-speaking listeners showed just the reverse effect.

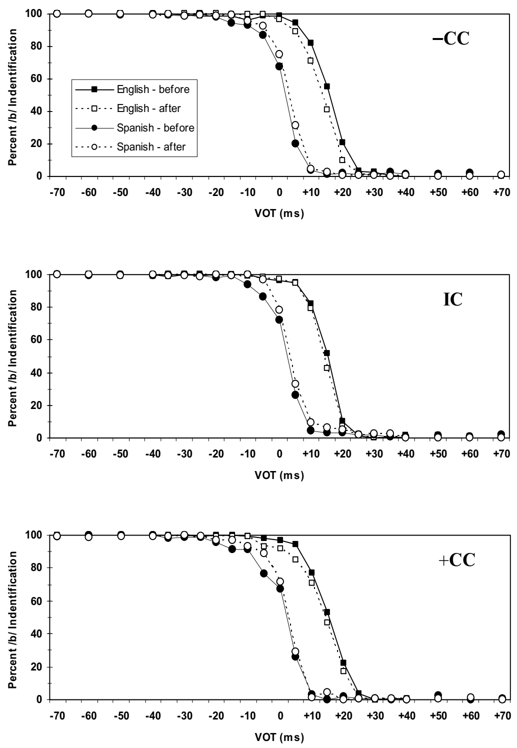

Figure 5: VOT identification functions for English- and Spanish-speaking listeners, before and after TOT category training, in each of the three training conditions.

4. Discussion

The present study examined the relationship between linguistic experience, auditory discontinuities, and the perceptual learning of non-speech stimulus categories. The study also addressed the question of whether learning in the non-speech domain can be transferred to the perception of speech. It has already been shown that auditory discontinuities affect perceptual category learning, and that category boundaries can shift as a function of the input distributions [3], a result which was partially replicated here. We, however, found that although TOT category learning was aided by the presence of the positive AD at the distributional boundary, it was either unaffected or even hindered by the presence of the negative AD. There does not seem to be a simple mapping from perceptual discriminability to perceptual learnability. An ongoing study in our lab is looking at the effect of linguistic experience on the discriminability of TOT stimuli. Given the extensive experience that listeners have with VOT stimuli, one might expect some effect of native language on the perception of non-speech stimuli that have analogous characteristics. The fact that no such effect was observed suggests that linguistic experience does not affect perceptual learning in a non-speech domain (but see Bent et al. [6] for a qualification).

5. Conclusions

Research on the effects of linguistic experience on the perception of speech has a long history. However, there have been relatively few studies of the long-term effects of linguistic experience on auditory processing of non-speech sounds. In this study we chose to examine the effect of linguistic experience with VOT stimuli on learnability of analogous TOT stimulus categories. Our results suggest that phonetic re-organization through language acquisition may affect only the perception of speech but cause little change in the basic auditory system.

6. 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 Randy Diehl. Special thanks to Lori Holt of Carnegie Mellon University, Andrew Lotto of Boys Town Medical Hospital, Angel Melo of the University of Tlaxcala, Mexico, Sidiney Gerardo Reyes, and Cheryl Moran.

7. References

[1] Jusczyk, P. W., Pisoni, D. B., Walley, A., and Murray, J. (1980). "Discrimination of relative onset time of two-component tones by infants," *J. Acoust. Soc. Am., 61(1)*, 262-270.

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

[3] 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.

[4] Pisoni, D. B. (1977). "Identification and discrimination of the relative onset time of two component tones: Implications for voicing perception in stops," *J. Acoust. Soc. Am., 61(5)*, 1352-1361.

[5] Klatt, D.H. and Klatt, L.C. (1990). "Analysis, synthesis and perception of voice quality variations among female and male talkers," *J. Acoust. Soc. Am., 87(2),* 820-857.

[6] Bent, T. Bradlow, A.R., and Wright, B.A. (2004). "The influence of linguistic experience on pitch perception in speech and non-speech sounds," *J. Acoust. Soc. Am., 113*, 2256.

---

[1] In the –CC training condition, 4 English-speaking listeners, and 2 Spanish-speaking listeners did not form a single generalization labeling function, but instead displayed U-shaped functions with two boundaries. Additionally, 2 English participants in the IC condition showed U-shaped functions. The U-shaped functions suggest that these participants did not learn strict TOT category boundaries, but instead categorized the stimuli as having either simultaneous or non-simultaneous onset. In these few cases, either the boundary in the negative TOT region (–CC) or the boundary that was most prominent (IC) was selected for the linear regression analysis. One Spanish-speaking listener was excluded from the analysis because she did not show any boundary at all and thus we could not obtain a significant linear curve fit to her data.