Copyright

by

Jessica Sari Fleming Hay

2005

The Dissertation Committee for Jessica Sari Fleming Hay certifies that this is the approved version of the following dissertation:

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

Committee:

Randy L. Diehl, Supervisor

David P. Birdsong

Leslie B. Cohen

Catharine H. Echols

Harvey M. Sussman

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

by

Jessica Sari Fleming Hay, B.Sc.; M.A.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August, 2005

ACKNOWLEDGMENTS

Preparation of this dissertation was supported in part by a National Institutes of Health grant awarded to Randy Diehl. The author gratefully acknowledges the assistance and support of her supervisor, Randy Diehl, without whom this project would not have been possible. The author would also like to give a special thanks to Lori Holt and Andrew Lotto for their continued support and feedback as well as to her committee members, David Birdsong, Les Cohen, Cathy Echols, and Harvey Sussman. Cheryl Moran, Sidiney Gerardo Reyes, and Angel Melo were invaluable to this project, and are thanked for recruiting and running participants, among other things. Finally, the author would like to thank her parents, her friends, and her partner Kris, for their love, patience and encouragement throughout this process.

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

Publication No. _____

Jessica Sari Fleming Hay, Ph.D.
The University of Texas at Austin, 2005

Supervisor: Randy L. Diehl

Speech perception results from a complex interplay between the operating characteristics of the auditory system (i.e., auditory discontinuities) and linguistic experience. Research in human infants and animals, and research using tone-onset-time (TOT) stimuli, a type of non-speech analogue of voice-onset-time (VOT) stimuli, has suggested that there is an underlying auditory basis for the perception of stop consonants based on a threshold for detecting temporal onset asynchronies in the vicinity of ± 20 ms. Languages, however, differ in their reliance on temporal onset asynchrony-based auditory discontinuities in their [voice] categories. This dissertation sought to examine whether long-term linguistic experience with different [voice] categories (i.e., English or Spanish) affects the perception of non-speech stimuli that are analogous in their acoustic timing characteristics. This research was also designed to investigate the joint effects of linguistic experience and auditory mechanisms on phoneme structure and category learning. Three cross-linguistic studies were designed to look at (1) the production and perception of VOT and the perception of TOT, (2) the effects of stimulus range on the

perception of VOT, and (3) the effects of auditory discontinuities on non-speech category learnability. Results indicate that linguistic experience *does* affect the perception of non-speech stimuli, at least in certain circumstances. Thus, there is some commonality in the processes used to discriminate between non-speech sounds and those used to discriminate between speech sounds. Additionally, auditory discontinuities were found to influence both phoneme structure and category learning. It is suggested that English- and Spanish-speaking listeners use different cues to discriminate their [voice] categories. Results also suggest that there are perceptual asymmetries between the positive and the negative onset asynchrony-based auditory discontinuities. The relationships between auditory discontinuities, linguistic experience, discriminability, phoneme category structure, and learnability are discussed.

TABLE OF CONTENTS

	Page
LIST OF TABLES.....	x
LIST OF FIGURES.....	xi
 CHAPTER 1: Context of Dissertation.....	 1
1.1 Overview.....	1
1.2 What is VOT and How is it Distributed Cross-linguistically?	5
1.3 Theories of Perception of VOT.....	9
1.3.1 VOT Discrimination is Language-Dependent.....	9
1.3.2 VOT Categories are Linguistic and Innate.....	10
1.3.3 VOT Discrimination is Due to Auditory Mechanisms.....	12
1.4 Asymmetries in the Production and Perception of VOT.....	18
1.5 Plan of Work.....	22
 CHAPTER 2: Study 1. Effects of Linguistic Experience and Auditory Discontinuities on Voice- and Tone-Onset-Time.....	 26
2.1 Introduction.....	26
2.2 Methods.....	29
2.2.1 Participants.....	29
2.2.2 Stimuli.....	29
2.2.2.1 VOT Stimuli.....	29
2.2.2.2 TOT Stimuli.....	31
2.2.3 Procedure.....	32
2.2.3.1 VOT Production.....	33
2.2.3.2 VOT Categorization.....	34
2.2.3.3 VOT Discrimination.....	34
2.2.3.4 TOT Discrimination.....	35
2.3 Results.....	35
2.3.1 VOT Production.....	35
2.3.2 VOT Categorization.....	37
2.3.3 VOT Discrimination.....	38
2.3.4 Categorical Perception of VOT.....	40
2.3.5 TOT Discrimination.....	43
2.4 Summary.....	50
2.5 Discussion.....	50

CHAPTER 3: Study 2. The Effect of Stimulus Range on VOT Perceptual Category Boundary Locations.....55

3.1 Introduction.....	55
3.2 Methods.....	57
3.2.1 Participants.....	57
3.2.2 Stimuli.....	57
3.2.3 Procedure.....	57
3.3 Results.....	58
3.3.1 VOT Category Labeling.....	58
3.3.2 VOT Category Goodness Judgments.....	62
3.4 Summary.....	64
3.4 Discussion.....	64

CHAPTER 4: Study 3. The effects of Linguistic Experience and Auditory Discontinuities on Perceptual Learning.....69

4.1 Introduction.....	69
4.2 Methods.....	71
4.2.1 Participants.....	71
4.2.2 Stimuli.....	72
4.2.2.1 VOT Stimuli.....	72
4.2.2.2 TOT Stimuli.....	72
4.2.3 Procedure.....	72
4.2.3.1 VOT Categorization.....	72
4.2.3.2 TOT Category Training.....	73
4.2.3.3 TOT Generalization.....	74
4.2.3.4 TOT Category Refresher.....	75
4.2.3.5 VOT Categorization.....	75
4.3 Results.....	75
4.3.1 TOT Category Training.....	75
4.3.2 TOT Generalization.....	77
4.3.3 VOT Categorization.....	83
4.4 Summary.....	86
4.5 Discussion.....	86

CHAPTER 5: Summary and Interpretation of Findings.....92

5.1 Summary of Studies.....	92
5.1.1 Study 1.....	93
5.1.2 Study 2.....	94
5.1.3 Study 3.....	94
5.2 General Discussion.....	95

5.2.1 Linguistic Experience Affects Perception of Non-Speech Sounds...	95
5.2.2 Cues to the Spanish [Voice] Distinction.....	97
5.2.3 Asymmetrical Auditory Discontinuities.....	100
5.2.4 Auditory Discontinuities Do Not Yield Immutable Boundaries...	102
5.3 Proposed General Framework for Phoneme Acquisition.....	105
5.4 A Dissertation in Context: The Generalizability of Findings.....	108
5.4 A Dissertation in Context: General vs. Language-Specific Mechanisms...	109
5.5 Future Directions.....	111
APPENDIX A.....	112
APPENDIX B.....	113
APPENDIX C.....	114
APPENDIX D.....	115
APPENDIX E.....	116
APPENDIX F.....	117
APPENDIX G.....	118
APPENDIX H.....	119
APPENDIX I.....	120
REFERENCES.....	121
VITA.....	128

LIST OF TABLES

Table	Page
I. English and Spanish minimal pair words produced in Study 1.....	33

LIST OF FIGURES

Figure	Page
1. Three examples of labial VOT productions.....	6
2. Distribution of VOT productions in (a) French and (b) English.....	8
3. Schematic spectrograms of three representative TOT stimuli.....	13
4. Spectrograms of synthetic VOT stimuli.....	31
5. VOT distributions of English labial stops in initial position.....	36
6. VOT distributions of Spanish labial stops in initial position.....	37
7. VOT category labeling functions for English- and Spanish-speaking listeners.....	38
8. VOT discrimination functions for English- and Spanish-speaking listeners.....	40
9. Obtained and predicted VOT discrimination functions for English-speaking listeners.....	42
10. Obtained and predicted VOT discrimination functions for Spanish-speaking listeners.....	43
11. TOT discrimination functions for English- and Spanish-speaking listeners.....	45
12. VOT and TOT discrimination functions for English-speaking listeners...	46
13. VOT and TOT discrimination functions for Spanish-speaking listeners...	47
14. Obtained TOT and predicted VOT discrimination functions for English-speaking listeners.....	48
15. Obtained TOT and predicted VOT discrimination functions for Spanish-speaking listeners.....	49
16. VOT category labeling functions of English-speaking listeners in each of the three range conditions.....	59

17.	VOT category labeling functions of Spanish-speaking listeners in each of the three range conditions.....	60
18.	The effects of Order of presentation on the VOT category labeling functions of English- and Spanish- speaking listeners, for conditions SR and ER.....	61
19.	The effects of Order of presentation on the VOT category labeling functions of English- and Spanish-speaking listeners, for the MR condition.....	62
20.	Category goodness judgments of English- and Spanish-speaking listeners in each of the three range conditions.....	63
21.	Stimulus input distributions reflecting the relative frequency of stimulus presentation as a function of TOT.....	74
22.	Effect of Language and Training Condition on the number of training blocks to criterion by English- and Spanish-speaking listeners.....	77
23.	TOT generalization functions for English- and Spanish-speaking listeners in each of the three training conditions.....	79
24.	TOT generalization functions for English- and Spanish-speaking listeners in each of the three training conditions. Data from participants with U-shaped generalization functions were excluded.....	81
25.	Categorization data from English- and Spanish-speaking listeners who displayed U-shaped TOT generalization functions.....	83
26.	VOT identification functions for English- and Spanish-speaking listeners, before and after TOT category training, in each of the three training conditions.....	85

CHAPTER 1: Context of Dissertation

1.1 Overview

Speech perception, like many other types of perception, relies on a complex interplay of basic perceptual mechanisms and experience with external stimuli. Our ability to perceive speech can be viewed as an interplay between the perceptual biases present in the mammalian auditory system and our linguistic experience. The term ‘bias’ or ‘biases’ is used here to refer to the non-continuous perception of a continuous physical dimension. Biases are evident when there are perceptual nonlinearities (or discontinuities) in the mapping from acoustic signals to auditory responses, and they correspond to heightened sensitivity in a particular region of acoustic space. Biases in the auditory system (also called auditory discontinuities), that exist prior to linguistic experience, may influence the sound inventories used across the world’s languages. The perceptual difference between sounds that straddle an auditory discontinuity can be expected to be exaggerated relative to pairs of sounds that do not have an auditory discontinuity between them. Thus, it may be communicatively advantageous for languages to come to exploit this heightened perceptual distinctiveness in their sound inventories. While the auditory system may favor the use of certain acoustic dimensions, or regions of acoustic space, for differentiating phoneme categories, it does not uniquely determine which particular sound categories are represented within the human brain; it is only through exposure to a specific language that native-language sound categories are created.

This dissertation provides a review of the studies and the theories that have led to our current understanding of the role that the auditory system in general, and auditory

discontinuities in particular, play in the acquisition of native speech sound categories. It is designed to investigate the relationship between linguistic experience, the operating characteristics of the auditory system, and category learning of both speech and non-speech sounds.

There have been a number of different theories regarding the acquisition of native language phoneme categories. Some of the first theories posited that the ability to discriminate phoneme categories was the result of language-specific experience (e.g., Abramson and Lisker, 1970; Lisker and Abramson, 1970). The discovery that very young infants can discriminate certain speech sounds raised the possibility that phonemes may be innately specified in the human brain (e.g., Eimas, Siqueland, Jusczyk, and Vigorito, 1971). Specialized speech-perception mechanisms became an unlikely explanation, however, when it was shown that some animals could also discriminate between and categorize speech sounds in a human-like manner (e.g., Kuhl and Miller, 1975; 1978). It became increasingly evident that discrimination abilities in both infants and animals might be due not to any characteristic of language itself, but rather to the operating characteristics (i.e., biases or sensitivities) of the mammalian auditory system. This idea has been supported by perceptual experiments with non-speech stimuli (e.g., Pisoni, 1977; Jusczyk, Pisoni, Walley, and Murray, 1980).

The studies mentioned here, and described more fully in the body of the text, have focused on the categorization and discrimination of sounds that vary in the presence, absence, and degree of voicing. Voicing is primarily characterized as variation in voice-onset-time (VOT), although there are other cues that are associated with voicing distinctions. VOT refers to the time between the release of the stop consonant and the

start of vocal fold vibration. Acoustically, VOT is abstractly realized by the relative onset of low and high frequency energy. At its most basic level, the perception of VOT has been characterized as the ability to discriminate simultaneous onset from asynchronous onset of that low and high frequency energy. See below for a more detailed description of VOT.

Perceptual studies on dimensions of VOT and its non-speech analogue, tone-onset-time (TOT), have found that there appear to be some regions of high discriminability (Pisoni, 1977; Kuhl and Miller, 1975; 1978; Jusczyk et al., 1980, Holt, Lotto, and Diehl, 2004), so it is not surprising that across the world's languages, VOT is one of the dimensions used to distinguish between different sound categories. The actual distribution of VOT categories (i.e., the number of phonemes that fall along the VOT dimension for a particular language, the region of the VOT dimension that phonemes occupy, and the range and shape of the distribution of each phoneme), however, differ from one language to the next (Lisker and Abramson, 1964; Keating, Linker, and Huffman, 1983; Kessinger and Blumstein, 1997). It is through the study of VOT and TOT that we have started to develop a greater understanding of the role that both the auditory system and linguistic experience play in speech perception.

The use of non-speech stimuli to investigate issues in speech perception has been invaluable to our understanding of the processes that underlie the acquisition of phoneme categories. Non-speech analogues of speech sounds provide an appropriate stimulus set in studies for which strict control over listeners' previous experience and stimulus characteristics are important. It has been assumed that although speech sounds and their non-speech analogues are processed by similar mechanisms, experience in one domain

does not affect processing in the other. Specifically, researchers who use non-speech stimuli in attempts to learn how speech is processed have assumed that the auditory system is unaffected by linguistic experience. A recent study by Bent and colleagues (Bent, Bradlow, and Wright, 2003) suggests that this widely held assumption may be incorrect, at least in some situations. In their study, Mandarin listeners made more pitch glide identification errors (a non-linguistic auditory categorization task) than English listeners, while at the same time being more accurate at Mandarin tone identification. They were, therefore, able to demonstrate that linguistic experience can affect the perception of non-speech sounds that share certain features with speech sounds.

I am interested in the broad question of whether linguistic experience affects perception of non-speech sounds, and whether, in turn, auditory experience with non-speech sounds might affect the perception of speech sounds. Examining the effects of linguistic experience on the perception of auditory discontinuities provides a way to assess the effects of linguistic experience on non-speech auditory perception.

This dissertation is an investigation of the widely held assumption that speech and non-speech are processed independently (Liberman, Cooper, Shankweiler, and Studdert-Kennedy, 1967; Liberman and Mattingly, 1985). First, however, I review the literature on a well studied phonetic property that is found cross-linguistically, namely, VOT. It is through the study of VOT that some theories of phoneme category acquisition have been formulated and tested. I show how research with animals, non-speech stimuli, and neurophysiology has challenged assumptions of the innate specification of phoneme categories, and suggest that both auditory factors and linguistic experience are responsible for the types of cross-linguistic differences seen in the identification and

discrimination of VOT stimuli. I also review some of the peculiarities found in the VOT literature and address the specific goals and hypotheses of this dissertation.

1.2 What is VOT and How is it Distributed Cross-Linguistically?

Although languages differ in the specific phonemes that they use, many languages share commonalities in the articulatory/acoustic parameters they employ to distinguish one phoneme from another (Maddieson, 1984). Voicing¹ is a prime example of a type of articulatory/acoustic characteristic that is found in virtually all languages, but that is used differently from one language to the next. Voicing occurs when air passes through narrowly adducted vocal folds, causing the folds to vibrate. Voicing may start before, during, or after the release of a consonant and is characterized acoustically as low frequency energy in the vicinity of a consonant. Differences in the degree or duration of voicing have been commonly characterized in terms of VOT. VOT is the time between the release of the consonant and the start of vocal fold vibration (voicing), and it is measured in milliseconds (ms). By convention, when voicing starts before the release of the consonant, VOT has a negative value; when voicing and consonant release happen simultaneously, VOT has a value of 0 ms; and when voicing starts after the release of the consonant, VOT has a positive value (see Figure 1). Languages may have three or more different phonemes that exist along the voicing dimension, but over 50% of languages have only two phonemes at any given place of articulation (Maddieson, 1984).

¹ The terms ‘voiced’, ‘voiceless’, and ‘voicing’ refer to the phonetic property that is associated with the vibration of the vocal folds. Phonetic properties have a physical realization along a continuous physical scale. Although some languages may not have a phonetic voiced/voiceless distinction, the [+voice]/[–voice] feature distinction may be preserved, and is indicated by the phonological feature [voice]. Phonological features are realized along the same physical scale, but are themselves binary in nature. VOT quantifies the degree of phonetic voicing.

Wave Form

Spectrogram

Figure 1. Three examples of labial VOT productions. (A) Long-lead: Spanish /ba/ production with long-lead VOT. Voicing leads the release of the consonant by 115 ms and therefore has a VOT value of -115 ms. (B) Short-lag: Spanish /pa/ and English /ba/ production with simultaneous/short-lag VOT. Voicing lags the release of the consonant by 5 ms (VOT = +5 ms). (C) Long-lag: English /p/ production with long-lag VOT. Voicing lags the release of the consonant by 155 ms (VOT = +155 ms).

In many languages (e.g., Spanish, French, Polish, Dutch, Hungarian, Tamil), [+voice]² phonemes are produced with a substantial amount of pre-voicing (long-lead voicing), and [–voice] stops are produced with simultaneous onset of consonant release and voicing or a short lag between consonant release and the onset of voicing (short-lag voicing) (see Figure 2a). Fewer languages, of which English is one, produce their [+voice] consonants with simultaneous or short-lag onset and their [–voice] consonants with aspiration and a long lag (long-lag voicing) between the release of the consonant and the onset of voicing (see Figure 2b). Languages with two stops along the voicing dimension tend not to have a [voice] distinction that contrasts pre-voiced stops with aspirated long-lag voiceless stops (see Figure 2c), even though this would be optimal in terms of distinctiveness. Instead they choose VOT distributions that are immediately adjacent, and rely on the nature of the auditory system to enhance the distinctions. I will return to this point below.

There is substantial variation in the production of phonemes, even when they are spoken by a single talker. The variation in production creates a distribution of sounds that is unique to each particular phoneme. Although phonemes may differ from each other on a number of different dimensions, there is often a single dimension that accounts for much of the variance between two phonemes. The region between two phoneme distributions (i.e., the region that separates the two phoneme categories) is referred to as the ‘distributional boundary’.

² Recall, [+voice] and [–voice] are phonological distinctions and do not necessarily correlate with the (phonetic) presence or absence of vocal cord vibration (i.e., voicing).

Figure 2. Distribution of VOT productions in (a) French and (b) English. 0 ms VOT is marked by a dashed line. French contrasts long-lead VOT with short-lag VOT. English contrasts short-lag VOT with long-lag VOT. (Graphs taken from Kessinger and Blumstein, 1997)

Although it was already known that there are cross-linguistic differences in the use of the voicing dimension, systematic studies of VOT began with the work of Lisker and Abramson (1964). Lisker and Abramson wanted to see whether a single dimension, VOT, could serve to separate stop consonants in a number of different languages. They recorded and acoustically analyzed stop consonants produced by speakers of eleven different languages. In languages that have two (i.e., English, Cantonese, Dutch, Hungarian, Spanish, and Tamil) or three (i.e., Korean, Eastern Armenian, and Thai) stop consonants along the voicing dimension, phonetic distributions were well separated by differences in VOT. The authors also noted that the VOT region that each distribution occupies varies systematically with place of articulation, with the most fronted consonants having a distributional boundary at the shortest VOT and the most backed consonants having a distributional boundary at the longest VOT. In languages that have

four stop consonants along the voicing dimension, VOT was not a good predictor of phoneme distribution separation. So although the structure and placement of phoneme categories varies from language to language, VOT has generally been thought to be a good predictor of category separation.

1.3 Theories of Perception of VOT

1.3.1 VOT Discrimination is Language-Dependent

The first theories of VOT perception posited that sensitivities to differences in VOT were the result of linguistic experience. Pioneering work by Lisker and Abramson showed that speakers of different languages not only produce their [+voice]/[–voice] distinctions in different ways (1964), but also that their perceptual category boundaries vary in a way that corresponds to averaged production distributions (1970; Abramson and Lisker, 1970). In a series of perceptual studies, Lisker and Abramson (1970; Abramson and Lisker, 1970) found that adults from different linguistic environments (i.e., English and Thai) perceive the voiced-voiceless dimension categorically. Categorical perception refers to a joint phenomenon where a discrimination peak exists at the location of a category boundary, as determined by a category labeling task. The 50% cross-over point of a labeling function is often called the category boundary³. Categorical perception has long been assumed to be a phenomenon unique to both speech and humans (Liberman et al., 1967). Consistent with the typical definition of categorical perception, in Lisker and

³ In this manuscript, ‘perceptual category boundary’ or ‘category labeling boundary’ refers specifically to the location of the 50% cross-over point of a category labeling function or an area of heightened discriminability that falls in between two phonemes, while ‘distributional category boundary’ refers specifically to the line of demarcation between two distributions phoneme tokens. ‘Category boundary’ is used as a more general term that infers the existence of both a perceptual and a distributional category boundary that are observed in the same region of acoustic/auditory space.

Abramson's studies (1970; Abramson and Lisker, 1970) speakers of English and Thai showed discrimination peaks in the vicinity of their respective category boundaries. English-speaking adults showed a single discrimination peak at the English VOT category boundary (approximately +25 ms VOT for labials) and Thai adults showed two discrimination peaks at their respective category boundaries (approximately -25 ms and +40 ms VOT for labials). Both groups of participants had discrimination peaks and perceptual category boundaries that were consistent with the distribution of stop consonant productions found within each language. From their results, Lisker and Abramson concluded that adults' ability to detect differences along a physical dimension depends on their assignment of the stimuli to linguistic categories and that discriminability is determined by specific language experience.

1.3.2 VOT Categories are Linguistic and Innate

Early work by Eimas et al. (1971) and Trehub (1976) had suggested, however, that linguistic experience may not be necessary for aspects of categorical perception to be observed. They found that infants show categorical-like⁴ discrimination performance before they have any substantial experience with their native language. Eimas et al. (1971) showed that 1- and 4-month-old English-hearing infants can discriminate /ba/ and /pa/ and do so in a categorical-like manner with a peak that falls at roughly the same place as the English /ba-/pa/ category boundary. Categorical perception of speech sounds had been thought to be an exclusively human ability. Based on this assumption,

⁴ The term 'categorical-like' is used here because categorical perception refers to a joint phenomenon that involves a labeling function coupled with a discrimination peak at the boundary of the labeling function. Here, 'categorical-like' performance refers to discrimination being good between phoneme distributions and poor within a phoneme distribution, without the corresponding labeling function.

Eimas et al. (1971) concluded that infants respond in a “linguistically relevant manner” to synthetic speech stimuli that vary along the voicing dimension and that this ability is likely due to a biological specialization for language. Trehub (1976) found that 5 to 17 week-old English-hearing infants can discriminate the French oral/nasal contrast /pa/ – /pã/, and the Czech contrast /za/ – /řa/. Given that the infants were not exposed to either contrast prior to the testing situation, linguistic experience could not account for their discrimination performance. Again, this was taken as evidence that infants are born with the ability to discriminate speech sounds and that they do so in a linguistically relevant manner.

Infants from other linguistic environments have been shown to have discrimination peaks in regions that do not correspond to either the perceptual or distributional category boundaries in their language environment. For example, both English-hearing 5.5 to 11.0 month-old infants (Aslin, Pisoni, Hennessey, and Perey, 1981) and Spanish-hearing 4 to 6.5 month-old infants (Lasky, Syrdal-Lasky, and Klein, 1975) show enhanced discrimination of VOT in regions corresponding to the [+voice]/[–voice] distributional boundaries found in some natural languages. Spanish-hearing infants were able to discriminate synthetically produced stimuli with +20 and +60 ms VOT, corresponding to the English VOT category boundary. They were also able to discriminate –60 and –20 ms VOT stimuli, corresponding to one of the Thai category boundaries. They were not however, able to discriminate –20 and +20 ms VOT stimuli even though this is where the Spanish bilabial stop consonant perceptual category boundary lies for adults (Abramson and Lisker, 1973). These results were again seen to support the idea that discrimination abilities in the young infant are innately linguistic.

1.3.3 VOT Discrimination is Due to Auditory Mechanisms

Although the results from the Lasky et al. (1975) study could be interpreted as evidence for innate linguistic abilities, there is another possible interpretation that is based on research done with animals and with non-speech stimuli – namely that initial discrimination abilities are the result of general auditory mechanisms. Research using chinchillas, a type of mammal with auditory capabilities similar to that of human's (Miller, 1970), has provided substantial support for this claim. Kuhl and Miller (1975) found that, following training on an avoidance conditioning procedure, chinchillas show categorical labeling of consonants /t/ and /d/ in different vowel contexts. They also showed that chinchillas can generalize the labeling functions of learned VOT endpoints to new places of articulation, that the perceptual category boundaries have the same relative locations as those found in English-speaking humans (Kuhl and Miller, 1978), and that chinchillas have increased sensitivity to differences between stimuli in the vicinity of their perceptual category boundaries (Kuhl, 1981). These results suggested that categorical perception is not a uniquely human phenomenon, and they were the first demonstration that early discrimination abilities in infants might not be linguistic in nature. If animals have the same perceptual category boundary locations as human infants, then the more parsimonious explanation is that early discrimination of VOT stimuli is the result of characteristics of the mammalian auditory system.

Research with non-speech analogues of VOT have also supported the idea that the perception of phonetic categories is at least partially dependent on biases of the auditory system. Pisoni (1977) created a stimulus series that was analogous to VOT in its timing

characteristics, but that was not perceived as speech-like. Schematic spectrograms of three representative TOT stimuli can be seen in Figure 3. TOT stimuli were made up of two sine-wave components that varied in their relative onsets. The lower frequency component (500 Hz) was varied relative to the higher frequency component (1500 Hz) to create a series that varied from -50 ms TOT (lower frequency component leads the sequence by 50 ms) to $+50$ ms TOT (onset of the lower frequency components lags the onset of the higher frequency component by 50 ms).

Figure 3. Schematic spectrograms of three representative TOT stimuli. The onset of lower frequency tone (500Hz) 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 (1977) and the graph is taken from Holt et al. (2004).

The TOT stimuli that he used abstractly represent the acoustic frequency and timing characteristics of VOT variation. In pre-voiced consonants, low-frequency energy (also called the voice bar) is present before the release of the consonant. At the time of consonant release higher frequency energy is introduced, while the low frequency

stimulus components remain. Spectral energy of the consonant and following vowel is concentrated into frequency bands called formants. Depending on the consonant and following vowel, the pattern of formants varies, but the lowest frequency band of energy is called the first formant (F1), the second lowest band of energy is called the second formant (F2), and so on. During the period of consonant release, the voice bar and F1 may occupy a similar frequency range. Negative TOT stimuli have acoustic characteristics analogous to the voice bar/F1 and F2 of pre-voiced consonants in that a low frequency component precedes the onset of a higher frequency component. The inverse is true for voiceless stop consonants. In voiceless stops, when the consonant is released, high frequency energy is present (i.e., F2, F3, F4 etc.), but low frequency energy is not. Low frequency energy (i.e., F1) enters the signal when voicing commences. This has been called ‘F1 cutback’ and is characteristic of voiceless stops. Positive TOT stimuli mirror the acoustic characteristics of voiceless stops, in that high frequency energy is presented first, followed by the inclusion of lower frequency energy.

In Pisoni’s (1977) study, English-speaking subjects were trained to assign one of two labels, A or B, to the two TOT endpoints, –50 ms and +50 ms TOT. Participants then completed a labeling task and an ABX discrimination task on all 11 points in the TOT series. English-speaking adults responded categorically to non-speech signals in the same manner that they respond to analogous speech signals. Participants tended to have heightened discrimination in the same region as their 50% labeling cross-over point. Discrimination peaks and perceptual category boundary locations were not due to training; even without endpoint training, some subjects showed a discrimination peak at +20 ms TOT, which corresponds to the English VOT category boundary. Other subjects

showed two discrimination peaks, the same one at +20 ms and another one at –20 ms, which corresponds approximately to the Thai VOT category boundaries.

Infants can also perceive non-speech sounds in a categorical-like manner. Jusczyk and colleagues (Jusczyk et al., 1980) reported that 2.5 month-old infants can discriminate differences among TOT stimuli. Using a high-amplitude sucking procedure, they found that, like adults, infants have 2 regions of heightened discrimination along the TOT dimension, one in the negative region and one in the positive region.

In summary, adults, infants, and animals all have the ability to perceive speech sounds categorically. All three groups have been shown to have one, and perhaps two, distinct discrimination peaks along the VOT dimension. And both adults and infants show categorical perception with non-speech analogues of VOT. Together these results suggest that the perception of VOT in stop consonants reflects basic properties of the auditory system and that discrimination performance can be at least partially accounted for by the presence of natural auditory boundaries.

This is not an entirely new idea. Hirsch and Sherrick (1961) demonstrated that there is a threshold for judgments of non-simultaneity. Humans require an onset asynchrony of about 20 ms for two sounds to be perceived as non-simultaneous and for temporal order to be judged accurately. If the relative onsets are less than 20 ms apart, then the sounds may appear to have simultaneous onset and temporal order judgments are difficult, if not impossible. Miller and colleagues (Miller, Wier, Pastore, Kelly, and Dooling, 1976) also found that listeners were best able to discriminate differences between the onset of a noise and a buzz when the onset was separated by approximately 16 ms, and that labeling cross-over points occurred in the same vicinity. Since VOT is

characterized by the relative onset of low frequency (voicing bar or F1) and high frequency (F2 and F3) information, perception of VOT may be closely related to the types of temporal order judgments investigated by Hirsch and Sherrick (1961), Miller et al. (1976), and Pisoni (1977). The threshold account of non-simultaneity is only half of the explanation for the discrimination peaks found in the aforementioned studies. While discrimination becomes easier at the threshold for perceiving non-simultaneity, it becomes poorer again for onset asynchronies greater than 20 ms as a function of Weber's Law.

The idea that there is an underlying auditory mechanism that is responsible for the nonlinear perception of VOT becomes even more plausible when one examines the literature on neurophysiological correlates of VOT and TOT. A number of studies have found that animals (Sinex, McDonald, and Mott, 1991; Steinschneider, Schroeder, Arezzo, and Vaughan Jr., 1995) and English speaking adults (Sharma and Dorman, 1999; Simos and Diehl et al., 1998) have very different neurophysiological responses to VOT stimuli that are perceived as part of different English [voice] categories. For example, Simos et al., (1998) found that synthetically produced syllables that have VOT values of 0 ms and +20 ms tended to elicit a single neural Event-Related Magnetic Field (ERF) response peak at the onset of the syllable. Syllables with VOT values of +40 and +60 ms elicited two neural response peaks, one at the onset of the syllable and another at the onset of voicing. The response patterns to +20 and +40 ms VOT were qualitatively different; a +20 ms VOT stimulus elicited a single response peak. A +40 ms VOT stimulus elicited two discrete peaks. There was no qualitative difference in response patterns between 0 ms and +20 ms VOT, or between +40 and +60 ms VOT. Similar

patterns of results have been found for TOT stimuli in both adults' ERFs (Simos, Breier, Zouridakis, and Papanicolaou, 1998) and neonates' Auditory Evoked Responses (Simos and Molfese, 1997), suggesting that there are neurophysiological correlates of the categorical perception of VOT.

So far, it has been argued that the underlying basis for the perception of VOT lies in auditory discontinuities, and that different languages may take advantage of those discontinuities in establishing phoneme category boundaries. Aside from any perceptual benefits that auditory discontinuities may provide, it would be communicatively advantageous if their presence made categories easier to learn. A recent study by Holt and colleagues (Holt et al., 2004) set out to test this idea. In two separate experiments, English-speaking listeners were trained to categorize distributions of TOT sounds into two groups. The first group of participants was presented with a distribution of TOT stimuli in which the distributional boundary was consistent with the auditory discontinuity at +20 ms (Consistent condition). The second group of participants was also presented with a distribution of TOT stimuli, but in this case the distributional boundary was either inconsistent with the auditory discontinuity at +20 ms (Experiment 1) or inconsistent with both the auditory discontinuity at -20 ms and the one at +20 ms (Experiment 2). In these Inconsistent conditions, the distributional boundaries were at approximately +40 ms and -2.5 ms TOT, for Experiments 1 and 2 respectively. In each experiment, auditory discontinuities at +20 ms and -20 ms fell within one, or both, of the categories. Participants in the Consistent condition of each experiment required fewer blocks to reach 90% correct categorization than did those in the Inconsistent condition. Thus, the presence of auditory discontinuities at distributional boundaries enhanced

category learning, suggesting that it is advantageous for languages to place their phoneme category boundaries in the vicinity of auditory discontinuities.

1.4 Asymmetries in the Production and Perception of VOT

Based on studies with human adults, human infants, and animals, it has often been assumed that there are two symmetrical regions of enhanced discriminability along the VOT dimension (at approximately -20 ms and $+20$ ms), each corresponding to a threshold for detecting onset asynchrony of low and high frequency stimulus components. However, some researchers have suggested that different perceptual processes might apply in the positive and negative regions of VOT (see Keating, Mikos, and Ganong, 1981). There are two primary acoustic cues associated with VOT: (1) variation in the temporal onset asynchrony of higher and lower frequency components, and (2) the presence or absence of low frequency energy (created through voicing) during consonant closure. Several lines of evidence support the idea that short-lag/long-lag distinctions are based on the perception of temporal onset asynchrony, while long-lead/short-lag distinctions may not be. First, the distributional characteristics of these two types of contrasts are quite different. In languages that have a short-lag/long-lag contrast, the $[+voice]/[-voice]$ distributions are immediately adjacent to one another, with a distributional boundary in the region of $+20$ ms VOT. In languages that have a long-lead/short-lag distinction, distributions of $[+voice]/[-voice]$ productions tend to be separated by at least 20 ms VOT, but more often by about 30-40 ms VOT. Across numerous cross-linguistic production experiments, virtually no productions have been observed in the region of -20 to 0 ms VOT (Lisker and Abramson, 1970; Keating et al.,

1981; Keating, 1979; Kessinger and Blumstein, 1997; Williams, 1977a). Given this gap in the distribution of values in the negative VOT range, perception of the long-lead/short-lag distinction may require less temporal precision than perception of the short-lag/long-lag distinction.

A second, related, line of evidence suggesting that positive and negative VOT may be processed using different perceptual mechanisms comes from cross-linguistic studies of stimulus range effects. Keating et al. (1981) examined the extent of range effects in the categorization of VOT by English- and Polish-speaking listeners. English contrasts short-lag and long-lag VOT values, and the distribution of the English stop consonants are adjacent to one another. Polish contrasts long-lead and short-lag VOT values, and, consistent with the discussion in the previous paragraph, the distributions are separated by approximately 35 ms VOT. Participants were presented with synthetic /da/-/ta/ stimuli and were asked to make category judgments. Three different series of stimuli were used: the first mimicked the range of actual Polish production distributions (range = -100 to +50 ms), the second mimicked an idealized Polish distribution spanning long-lead and short-lag distributions (range = -100 to +20 ms), and the third mimicked the range of English-like production distributions (range = -20 to +80 ms). Keating et al. (1981) found that for English-speaking listeners the location of the category labeling boundary was basically unaffected by stimulus range, occurring at approximately +35 ms VOT, and that the slopes of the labeling functions were steep. Polish-speaking listeners were affected by stimulus range; the Polish category labeling boundary was in the vicinity of 0 ms VOT for the two Polish-like series and in the vicinity of +20 ms VOT for the English-like series. The authors concluded that the stability of the English category

boundary suggests that English-speaking listeners use an analysis of the temporal onset asynchrony component of VOT to make category judgments. Polish-speaking listeners were very susceptible to range effects, especially in the region of their perceptual category boundary, where sensitivities would not be predicted based on the necessary threshold of 20 ms for judging onset asynchronies. Although it is possible that Polish-speakers' threshold for detecting onset asynchronies has been lowered due to their language experience, the authors suggest the Polish-speaking listeners are not using temporal onset asynchronies to make category judgments. As there is a gap between the Polish [+voice]/[–voice] distributions, Polish speakers may never have had to establish precise VOT category boundaries that rely on temporal onset asynchrony. They suggest that Polish-speaking listeners may instead be judging the presence or absence of voicing during consonant closure. They further suggest that Polish-speakers' sensitivity to VOT differences in the vicinity of short-lag region (recall that they had perceptual category boundary of +20 ms VOT for the English-like range) may be due to the psychoacoustic salience of spectral cues that are correlated with VOT (see Diehl, 1981 for a list of spectral cues associated with VOT).

A third line of evidence that suggests that positive and negative VOT may be processed using different perceptual mechanisms comes from the literature on cross-linguistic category boundary locations. In languages that have a short-lag/long-lag distinction, like English, VOT category labeling boundaries, discrimination peaks, and distributional boundaries all line up neatly around +20 ms VOT, for labial consonants (Lisker and Abramson, 1964; 1970; Abramson and Lisker, 1970). Therefore, in English, the distribution of stop consonants may reinforce the pre-existing auditory discontinuity

and together they may enhance the reliance on the detection of temporal onset asynchronies for making category judgments. In Spanish, and presumably other languages that contrast long-lead with short-lag VOT values, VOT category labeling boundaries and discrimination peaks line up, but they are to the right of the distributional boundaries. Category labeling boundaries and discrimination peaks tend to fall in the vicinity of 0 ms VOT for Spanish-speaking listeners (Williams, 1974; 1977a; 1977b). Although the Spanish stop consonant distributions are in line with the negative auditory discontinuity at around -20 ms (Williams, 1977a), Spanish-speaking listeners appear not to benefit from the auditory discontinuity at -20 ms when learning their perceptual [voice] categories. The production gap seen in Spanish stop distributions, where no tokens occur in the short-lead VOT region, might suggest that Spanish-speakers do not establish precise VOT category boundaries based on temporal onset asynchronies. Spanish-speakers, like Polish-speakers, may instead make [voice] judgments based on the presence or absence of voicing in the vicinity of the consonant closure. Given that the negative auditory discontinuity and production distributions tend to be aligned in languages that have long-lead/short-lag voicing distinction, it is surprising that a negative VOT boundary is rarely used perceptually. This might suggest that the negative auditory discontinuity is less salient than the positive one, or that the use of presence or absence of voicing during closure as a cue to perceptual categories provides some processing benefit over the use of the negative auditory boundary, or both.

1.5 Plan of Work

This dissertation is concerned with the complex interplay between biases in the auditory system and linguistic experience. First I pose the question of whether long term experience with the VOT distributions of speech sounds in a particular language (English or Spanish) affects the perception of TOT sounds, which mimic the acoustic timing characteristics of VOT in speech. I then ask the question of how input distributions and auditory discontinuities affect phonemic category structure. Lastly, I address whether the presence of auditory discontinuities affects non-speech category learning and whether short term experience in categorizing non-speech sounds affects the perception of speech sounds with similar acoustic characteristics.

Specifically, the first study (Chapter 2) was designed to examine whether long-term linguistic experience with different distributions of VOT stimuli (i.e., through exposure to different languages) affects the perception of non-speech stimuli that are analogous in their acoustic timing characteristics. If one language (L_A) places the boundary between its stop consonants at a different VOT value than another language (L_B), will speakers of L_A and L_B differ in their ability to detect temporal onset asynchronies in non-speech stimuli? Given the enormous amount of experience that people have with speech, it is plausible that listeners might develop auditory representations that allow them to detect non-speech acoustic changes that are relevant in their language environment. Although speakers of all languages presumably engage in the same types of lower level acoustic pattern learning, the structure of any resulting perceptual sensitivity augmentation should vary from one language to the next if linguistic experience affects the perception of non-speech sounds. For the most part,

however, it has been implicitly assumed that the perception of non-speech sounds that are analogous to speech sounds should remain unaffected by language experience. The first study addressed the question of whether acoustic pattern learning involved in the acquisition of phoneme categories transfers to non-speech processing. To address this question I collected VOT production, categorization, and discrimination data, as well as TOT discrimination data from monolingual English- and monolingual Spanish-speaking participants. The within-subjects nature of the design permits an examination of the relationship between individuals' distribution of VOT productions, identification boundaries, and discrimination peaks, as well as a comparison of VOT and TOT discrimination peaks. Acoustical analysis of the production data also help to determine how VOT is realized in English and Spanish and what additional cues are used for the differentiation of stop consonants in each language.

The second study (Chapter 3) in this dissertation assessed how the internal category structure of phonemes is affected by both auditory biases and the distributional characteristics of phoneme categories in the linguistic environment. If two phoneme categories are directly adjacent to each other along an acoustic dimension, then there should be a sharp perceptual boundary between the categories. If, on the other hand, two phoneme categories are produced with non-overlapping and separated distributions along an acoustic dimension, perceptual categories may have boundaries that are less well defined. Similarly, regions of heightened discriminability should serve as natural boundaries between phoneme categories, and the distribution of speech tokens within the categories and the sharpness of the perceptual category boundaries should reflect those regions of heightened discriminability. In this study, the effect of stimulus range on VOT

perceptual categorization and category goodness judgments was examined. The acoustic characteristics and distributional structure of English and Spanish labial stop consonant categories, as measured in Study 1, were compared to listeners' internal category representations. Internal category representations should be predicted by both auditory biases and distributional properties of VOT productions.

The third study (Chapter 4) examined the extent to which auditory biases affect acoustic pattern learning and whether perceptual learning in the non-speech domain (i.e., the learning of TOT categories) can transfer to the speech domain. Several questions were addressed in this last study. (1) When listeners are trained to categorize non-speech sounds that vary in their acoustic timing characteristics, is it easier to learn the categories if the boundary between the distributions falls in a region where the auditory system is biased towards high discriminability? (2) Are the positive and negative auditory discontinuities symmetrical in their effects on category learning? (3) How does language experience affect auditory processing (i.e., does language experience shift the discrimination patterns for non-speech stimuli?) and hence the learning of non-speech categories? (4) Does training on the categorization of TOT stimuli affect the placement of VOT category boundaries? Participants (both Spanish and English) were trained to categorize various distributions of TOT stimuli. TOT distributions were either consistent with the positive auditory discontinuity, consistent with the negative auditory discontinuity, or inconsistent with both. This TOT category training study was a cross-linguistic partial replication of a previous monolingual study that found that category learning was more rapid when the boundary between category distributions falls in a region of high discriminability as opposed to when an area of high discriminability falls

within one, or both, of the category distributions (Holt et al., 2004). Participants completed a VOT identification task before and after they were trained to categorize TOT stimuli into two groups. If category training with non-speech sounds affects the perception of acoustically analogous speech sounds, the location of the distributional category boundary in the TOT training task should predict a shift in the subjects' VOT identification boundary.

CHAPTER 2: Study 1. Effects of Linguistic Experience and Auditory Discontinuities on Voice- and Tone-Onset-Time

2.1 Introduction

The first study was designed to examine how linguistic experience and auditory mechanisms affect the perception of speech and non-speech. As described in Chapter 1, English and Spanish differ in the placement of their labial stop consonants along the VOT dimension, both in perception and in production. English has distributional and perceptual category boundaries that are consistent with an area of heightened auditory sensitivity (i.e., the positive auditory discontinuity) at approximately +20 ms VOT. Of the two main cues associated with VOT, native English learners presumably use the onset asynchrony cues available in VOT stimuli when making category judgments. The presence of the positive auditory discontinuity may act as a natural boundary, to assist them in categorization. Long-term experience with English stop consonants may reinforce or augment sensitivities in the vicinity of the positive auditory discontinuity. The Spanish distributional [voice] category boundary has been described as being consistent with the negative auditory discontinuity, with a gap in production straddling the region from 0 ms to –20/30ms VOT. Spanish perceptual VOT category boundaries, however, fall to the right of the distributional boundaries at approximately 0 ms VOT (Williams, 1974; 1977a; 1977b). It is unclear if this production-perception misalignment is the result of characteristics of stimulus and experimental designs or instead a result of differential cue weighting in Spanish [voice] judgments. If Spanish-speaking listeners depend on the temporal onset asynchrony cues available to VOT when making their [voice] category judgments, it was expected that long-term experience with Spanish stop

consonants should reinforce or augment sensitivity in the region of the negative auditory discontinuity, the region around 0 ms TOT, or both. If, however, Spanish-speaking listeners are not using temporal onset asynchrony when making their [voice] judgments, long-term experience with Spanish stops consonants would not be expected to reinforce or augment either of the auditory discontinuities, or the region between them.

In order to make predictions about how linguistic experience and auditory mechanisms affect the perception of speech and non-speech, it is important to have an adequate description of the speech input distributions. Production data were collected from each participant in this study. The labial consonants of both English and Spanish speakers were acoustically analyzed in terms of VOT (i.e., the time between the release of the consonant and the onset of voicing).

Following the production portion of this study, participants performed several tasks including VOT category labeling, VOT discrimination, and TOT discrimination. The VOT category labeling and discrimination tasks served a number of purposes. First, they provided a measure of categorical perception for cross-linguistic comparison. English-speaking listeners were expected to have a category labeling boundary that is consistent with their VOT discrimination peak, in the vicinity of +20 ms VOT. Spanish-speaking listeners were expected also to show a consistent labeling boundary and discrimination peak, but instead in the vicinity of 0 ms VOT. Second, these tasks allow us to compare perception and production in the same group of participants. Third, and most pertinent to the general purpose of this dissertation, VOT perceptual category boundaries and discrimination peaks provide an appropriate comparison for TOT discrimination patterns in assessing the effects of linguistic experience on the perception

of non-speech stimuli. A number of studies have already shown that English-speaking listeners have a VOT discrimination peak in the vicinity of the positive auditory discontinuity at +20 ms (Lisker and Abramson, 1964; 1970; Abramson and Lisker, 1970). Separate studies have shown that English-speaking listeners tend to have a TOT discrimination peak in the vicinity of +20 ms TOT, and some also have a smaller peak at -20 ms TOT (Pisoni, 1977). If English-speakers process VOT by means of a temporal onset asynchrony threshold, there should be within subject consistency across the various perceptual measures. It is possible that the English speakers use the same abstract auditory representation when processing both VOT and TOT stimuli. Since auditory sensitivities would be enhanced only in the positive region for English speakers, one might expect that the positive TOT peak should be much higher than the negative TOT peak. This is in fact a pattern that has been seen in a number of different studies of TOT perception by English-speaking listeners (e.g., Pisoni, 1977).

As mentioned before, it was predicted that Spanish labial stops would be distributed along the VOT dimension in a way that *could* naturally enhance the negative auditory discontinuity at approximately -20 ms. Studies with Spanish speakers have not found discrimination peaks or category labeling boundaries in the vicinity of the negative auditory discontinuity. If Spanish-speaking listeners are using temporal onset asynchrony when making their [voice] judgments, TOT discrimination should be augmented in the region of the Spanish perceptual category boundary. Research to date has suggested that speakers of languages like Spanish may not rely on onset asynchrony cues associated with VOT in the same manner as English listeners do when categorizing their stop consonants (Keating et al., 1981). The non-reliance on onset asynchrony

judgments for stop consonant categorization in Spanish might be evident in the perception of TOT: both positive and negative TOT discrimination peaks should be evident (owing to the auditory discontinuities in those regions), but they should be small relative to the positive TOT discrimination peak for English-speaking listeners (owing to the lack of linguistic experience that might augment sensitivities).

2.2 Methods

2.2.1 Participants

Twenty-four native English-speaking listeners and twenty-four native Spanish-speaking listeners served as participants in this study. There were equal numbers of males and females in each language group. All participants were pre-screened using a language profile to determine their eligibility for the study as a monolingual speaker of either English or Spanish. Demographic and language background information can be found in Appendix A. English-speaking listeners were recruited from the Psychology Department at the University of Texas at Austin. Spanish-speaking listeners were recruited from the Biology and Psychology Departments at the University of Puebla, in Puebla, Mexico. Participants were paid approximately \$12 US or its equivalent for their participation.

2.2.2 Stimuli

2.2.2.1 VOT Stimuli

A slightly modified version of the Klatt synthesizer (Klatt and Klatt, 1990) was used to create a series of 29 /ba/-/pa/ stimuli (sampling rate = 11025 Hz) that ranged from

-70 ms to + 70 ms VOT in steps of 5 ms (see Figure 4). 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) in the synthesis parameters. In order to mimic the release of the consonant, there was 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 to 300 Hz during the period of voicing lag. F1 onset was 438 Hz and transitioned linearly over the first 40 ms of the stimulus to its steady-state value of 700 Hz, during which time the F1 bandwidth was maintained at 90 Hz until the last 30 ms of the stimulus when it increased linearly to end at 180 Hz. F2 and F3 had onset values of 1025 and 2425 Hz, respectively, and transitioned linearly over the first 40 ms of the stimulus to reach their steady-state values of 1200 and 2600 Hz, respectively. F4 and F5 remained constant at 3300 and 3700 Hz, respectively. The bandwidths of F2 through F5 remained constant throughout the stimulus (B2 = 90 Hz, B3 = 130 Hz, B4 = 400 Hz, B5 = 500 Hz). Each of the 15 stimuli that had VOT values of 0 ms or greater were 295 ms long. Fundamental frequency (F0) increased linearly from 120 Hz to 125 Hz over the first 40 ms of the stimulus and then decreased linearly to 100 Hz over the remaining 255 ms of the stimulus. Variations in negative VOT stimuli (i.e., from -5 ms to -70 ms) were 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. The frequencies of F2 through F5 were set at their onset values, but their bandwidths were increased to 500 Hz so that there was little energy at those frequencies. In natural speech the amplitude of the voicing bar is lower than the amplitude of voicing

following consonant release. In these synthetic stimuli, pre-voicing also had a lower amplitude than did the remaining voiced portion of the stimulus. Pre-voiced stimuli ranged from 300 ms to 365 ms in duration. Only a subset of these stimuli was used in the first set of experiments.

Figure 4. 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.

2.2.2.2 TOT Stimuli

TOT stimuli were modeled after those used by Pisoni (1977) and consisted of two sine-wave components in which the onset of the lower frequency component was varied relative to the onset of the higher frequency component. The two components had a constant frequency of 500 Hz and 1500 Hz, respectively. The duration of the 1500 Hz component was held constant at 230 ms. The duration of the lower frequency component was varied relative to the higher frequency component in order to create a series of TOT stimuli that ranged from -70 ms to $+70$ ms TOT in steps of 5 ms. In order to mimic the energy distribution found in natural syllables, the intensity of the higher component was

12dB lower than that of the lower component. During the first 5 ms of each of the individual components, amplitude was ramped linearly upwards. Individual sine-wave components were created using an in-house waveform editing program at a sample rate of 22050Hz and then combined in Adobe Audition and resampled at 11025 Hz. Only a subset of these stimuli was used in the first set of experiments.

2.2.3 Procedure

All verbal interactions with the participants and all written instructions were in their native language. Participants were run in four consecutive tasks in the following order: 1) VOT Production, 2) VOT Categorization, 3) VOT Discrimination, and 4) TOT Discrimination. Tasks 3 and 4 were counterbalanced. Each task took between five and fifteen minutes to complete, and participants received a short break after each task if they desired one. The entire experimental session took less than 60 minutes.

Participants were seated in front of a computer (English, Dell Workstation PWS350; Spanish, Dell Latitude D800 laptop), on which the tasks were administered. All stimuli were randomized and presented by the experimental design program, ALVIN (Hillenbrand and Gayvert, 2004). Stimuli were played through Sennheiser HD280 Professional headphones at a comfortable listening level. Productions were also elicited and recorded using ALVIN, and were spoken into a Shure WH20 headworn dynamic microphone.

2.2.3.1 VOT Production

Speakers were asked to produce a series of two-syllable, native language, minimal pair words that varied in initial labial consonant. The initial consonants /p/ and /b/ were produced in the context of the five Spanish vowels /a, e, o, i, u/ in Spanish words, and in the context of the five most similar English vowels /a, e, o, i, u/ in English words. Each word was produced three times, and the entire set of words was presented in random order. The following table includes the list of English and Spanish words produced by English- and Spanish-speakers, respectively.

Table I. English and Spanish minimal pair words produced in Study 1.

Spanish Words		English Words	
bala – pala	/ˈba.la/ - /ˈpa.la/	batty – Patty	/ˈba.ti/ - /ˈpa.ti/
beso – peso	/ˈbe.so/ - /ˈpe.so/	Betty – petty	/ˈbe.ti/ - /ˈpe.ti/
boca – poca	/ˈbo.ka/ - /ˈpo.ka/	bony – pony	/ˈbo.ni/ - /ˈpo.ni/
vida – pida	/ˈbi.ða/ - /ˈpi.ða/	beeper – peeper	/ˈbi.pə/ - /ˈpi.pə/
bullá – pulla	/ˈbu.ja/ - /ˈpu.ja/	boodle – poodle	/bu.dl/ - /pu.dl/

Acoustic Analyses. The speech tokens were recorded at a sampling rate of 16 kHz. They were acoustically analyzed using PRAAT. Both waveform displays and spectrograms were used to determine the VOT values for the initial labial consonant in each word. VOT for short- and long-lag tokens was measured from the release on the consonant to the onset of voicing. VOT for pre-voiced tokens was measured from the onset of voicing to the release on the consonant. All measurements were taken at first zero-crossing of the segment in question.

2.2.3.2 VOT Categorization

Participants were asked to categorize a subset of the synthetically produced /ba/-/pa/ series described in section 2.2.2.1. They were presented with stimuli ranging from –70 ms VOT to +70 ms in steps of 10 ms VOT, except in the range from –40 ms to +40 ms VOT where stimuli varied in steps of 5 ms VOT. The 23 stimuli included in this experiment were each presented 10 times, for a total of 230 trials. Each participant received a unique random order of the 230 trials. This task took approximately 5 minutes to complete.

2.2.3.3 VOT Discrimination

Participants performed an ABX discrimination task on a series of synthetically produced /ba/-/pa/ stimuli ranging from –70 ms VOT to +70 ms VOT. In the ABX paradigm, subjects heard three sounds. The first two sounds, A and B, were always different from each other. The third sound, X, was always identical to either A or B. The participants' task was to indicate whether the third sound was the same as the first or the second sound. Following each response, participants received feedback. It has been suggested that feedback in discrimination tasks focuses and maintains participants' attention on the relevant stimulus dimensions (Pisoni, 1977; Pisoni et al., 1982). Each stimulus pair (A and B) was separated by 6 steps (30 ms VOT). The ISI was 750 ms within a stimulus trial. The trial ended when participants made a response, which then triggered a 1 second delay before the onset of the following trial. There were 12 pairs of stimuli, each with four orders of counterbalancing (ABA, ABB, BAA, BAB), with each

order presented twice, for a total of 96 trials per block. Each block took approximately 8-10 minutes to complete. Participants completed two blocks of trials.

2.2.3.4 TOT Discrimination

Participants performed an ABX discrimination task (with feedback) on a series of TOT stimuli ranging from -70 ms TOT to $+70$ ms TOT. All other procedures were identical to those described in the VOT discrimination section.

2.2 Results

2.3.1 VOT Production

Figures 5 and 6 show the distribution of labial stop consonants in English and Spanish, respectively. A two (Consonant) x five (Vowel) ANOVA revealed that English [+voice] stops were produced with less positive VOT values (mean = -6.5 ms VOT) than English [–voice] stops (mean = $+72.8$ ms VOT) [$F(1,710)=963.264, p<.001$]. Figure 5 shows that English [+voice] stops have a quasi bimodal distribution, with approximately 15 % being produced with pre-voicing (mean = -106.7 ms VOT). The remaining 85 % of English [+voice] stops that do not have pre-voicing had a mean of $+10.8$ ms VOT. An optimal distributional boundary location was calculated using binary logistic regression. Using this technique, English was found to have a distributional boundary between [+voice] and [–voice] stop consonants at approximately $+33.4$ ms VOT. There was no effect of Vowel and no Consonant x Vowel interaction on VOT in English.

Figure 5. VOT distributions of English labial stops in initial position.

A two (Consonant) x five (Vowel) ANOVA revealed that Spanish [+voice] stops were produced with more negative VOT values (mean = -65.7 ms VOT) than were Spanish [–voice] stops (mean = $+9.4$ ms VOT) [$F(1,703)=2683.650, p<.001$]. Seven Spanish tokens were missing from the dataset due to errors in recording. Spanish had an optimal distribution boundary at approximately -12.0 ms. There was a significant effect of Vowel on VOT, for these Spanish speakers [$F(4,703)=5.297, p<.001$]; Bonferroni post-hoc tests revealed that in the context of /a/ and /e/ VOT values were generally shorter than in the context of /o/ and /u/ ($p's<.05$). The Consonant x Vowel interaction was not significant.

Figure 6. VOT distributions of Spanish labial stops in initial position.

2.3.2 VOT Categorization

In order to determine the slope and 50% cross-over point in the labeling function for each listener, a linear regression analysis was performed on the data points that ranged from 100% /b/ identification to 100% /p/ identification. The slopes and 50% cross-over points of the regression lines were analyzed in a MANOVA, with Language as an independent variable. As expected English speakers had a significantly more positive identification boundary (16.63 ms VOT) than Spanish speakers (0.86 ms VOT) [$F(1,46)=331.439, p<.001$] (see Figure 7). The effect of Language on slope approached significance, indicating a tendency for English speakers to have a steeper labeling function than Spanish speakers ($p=.058$).

Figure 7. VOT category labeling functions for English- and Spanish-speaking listeners.

2.3.3 VOT Discrimination

A 2 between (Language) x 12 within (Stimulus) repeated measures ANOVA was performed to examine whether English and Spanish speakers have different regions of high discriminability along the VOT continuum. As expected, the analysis revealed no main effect of Language, suggesting that participants from both language groups were equally good at doing the ABX discrimination task. There was a main effect of Stimulus [$F(11,506)=72.472, p<.001$], indicating that some stimulus pairs were easier to discriminate than others. The analysis also revealed a significant interaction between Language and Stimulus on discrimination performance [$F(11,506)=11.122, p<.001$]. Planned comparisons within each language showed that English-speaking listeners had a

discrimination peak (i.e., a significant increase in discrimination followed by a significant decrease in discrimination) at approximately +15 ms VOT. Spanish-speaking listeners had a discrimination peak in the vicinity of 0 ms VOT. Figure 8 shows VOT discrimination functions for English and Spanish speakers and significant differences between adjacent data points are indicated by asterisks, for each language. A MANOVA revealed that Spanish-speaking listeners were better than English-speaking listeners at discriminating differences that centered around -5 ms [$F(1,46)=38.729, p<.001$], and that English-speaking listeners were better than Spanish-speaking listeners at discriminating differences that centered around +15 to +35 ms VOT (all $p's<.005$). The results indicate that listeners from the two language groups are most sensitive to difference between stimuli at their respective category labeling boundaries. Individual VOT category labeling and discrimination data are presented in Appendix B (English, Male), Appendix C (English, female), Appendix D (Spanish, male), and Appendix E (Spanish, female).

Figure 8. VOT discrimination functions for English- and Spanish- speaking listeners. Significant differences between adjacent data points are indicated for each language.

* $p < .05$

** $p < .01$

*** $p < .005$

**** $p < .001$

2.3.4 Categorical Perception of VOT

One of the advantages of collecting various perceptual measures for the same participants is that it allows comparisons of category labeling boundaries and discrimination functions. Predicted VOT discrimination values were calculated from VOT category labeling functions and compared to obtained VOT discrimination values.

The following formula was used to calculate the predicted discrimination values;

Predicted = $0.5 + 0.5 \times (p(L_1) - p(L_2))^2$, where $p(L_1)$ is the proportion of times token 1 was identified as /b/ and $p(L_2)$ is the proportion of times token 2 was identified as /b/.

Predicted discrimination responses were calculated using the labeling responses of the

token pairs that were presented in the VOT discrimination task. There was a significant Pearson correlation between the predicted and the obtained discrimination responses in both English ($r = .706, p < .01$) and Spanish ($r = .639, p < .01$), suggesting that these participants perceived the synthetic VOT stimuli categorically. Figure 9 shows the averaged predicted and obtained discrimination values for the English-speaking participants. As found in previous research, a repeated measures ANOVA revealed that actual discrimination performance was somewhat better than what is predicted on the basis of labeling performance alone [$F(1,276)=93.693, p < .001$]. A follow-up MANOVA found that English-speaking listeners had heightened sensitivity to VOT on both sides of their discrimination peak, but generally sensitivities were slightly skewed towards the positive side of VOT. (See Figure 9 for significance values.)

Figure 9. Obtained and predicted VOT discrimination functions for English-speaking listeners. Significant differences between obtained and predicted values are indicated with asterisks.

* $p < .05$

** $p < .01$

*** $p < .005$

**** $p < .001$

Figure 10 shows the averaged predicted and obtained discrimination values for the Spanish-speaking participants. These listeners also showed somewhat better than predicted discrimination performance [$F(1,276)=52.250, p<.001$]. A follow-up MANOVA found that Spanish-speaking listeners have heightened sensitivity to VOT only on the positive side of their discrimination peak, possibly owing to the presence of the positive auditory discontinuity or the richer array of acoustic cues available in the positive region of VOT. (See Figure 10 for significance values.)

Figure 10. Obtained and predicted VOT discrimination functions for Spanish-speaking listeners. Significant differences between obtained and predicted values are indicated with asterisks.

* $p < .05$

** $p < .01$

*** $p < .005$

**** $p < .001$

2.3.5 TOT Discrimination

A 2 between (Language) x 12 within (Stimulus) repeated measures ANOVA was performed to examine whether English and Spanish speakers have different regions of high discriminability along the TOT continuum. As expected, the analysis revealed no main effect of language, suggesting that participants from both language groups were equally good at doing the ABX discrimination task. There was a main effect of stimulus [$F(11,506)=25.791, p < .001$], indicating that some stimulus pairs were easier to discriminate than other. The analysis also revealed a significant interaction between

Language and Stimulus on discrimination performance [$F(11,506)=1.882, p<.05$].

Planned comparisons within each language showed that English-speaking listeners had a broad discrimination peak at approximately +15 ms TOT. Close examination of the discrimination function shows increased discrimination between –25 and –15 ms TOT ($p<.05$), in the vicinity of the negative auditory discontinuity, followed by a general increase in discriminability up to a maximum level at +15 ms TOT. Spanish-speaking listeners, on the other hand, had two small discrimination peaks at –15 ms and +15 ms TOT. Figure 11 shows TOT discrimination functions for English- and Spanish-speaking listeners, and significant differences between adjacent data points are indicated by asterisks, for each language. Individual TOT discrimination data are presented in Appendix F (English, Male), Appendix G (English, female), Appendix H (Spanish, male), and Appendix I (Spanish, female). A MANOVA revealed that English-speaking listeners were better than Spanish-speaking listeners at discriminating differences that centered around –5 ms [$F(1,46)=4.120, p<.05$] and +5 ms [$F(1,46)=7.027, p<.05$] TOT. The effect of Language approached significance ($p=.059$) for differences in discrimination that centered around the positive auditory discontinuity (i.e., at +15 ms TOT).

Figure 11. TOT discrimination functions for English- and Spanish-speaking listeners. Significant differences between adjacent data points are indicated for each language.

* $p < .05$

** $p < .01$

*** $p < .005$

**** $p < .001$

There was a significant correlation between TOT and VOT discrimination in both English ($r = .473, p < .01$) and Spanish ($r = .358, p < .01$). Figures 12 and 13 show the averaged TOT and VOT discrimination performance for the English- and Spanish-speaking participants. A repeated measures ANOVA revealed that there was no significant overall difference between the discriminability of TOT and VOT stimuli in either language (p 's $> .23$). There was however a Discrimination Condition (TOT vs. VOT) x Stimulus interaction in both English [$F(11,276)=7.156, p < .001$] and Spanish [$F(11,276)=8.072, p < .001$]. A follow-up MANOVA revealed that English-speaking listeners show close alignment of TOT and VOT discrimination peaks (see Figure 12 for

significance values). Spanish-speaking listeners show a VOT discrimination peak that falls between their two TOT discrimination peaks (see Figure 13 for significance values), although it overlaps somewhat with the positive TOT discrimination peak.

Figure 12. VOT and TOT discrimination functions for English-speaking listeners. Significant differences between VOT and TOT discrimination performance are indicated with asterisks. Error bars represent the standard error of the mean.

* $p < .05$

** $p < .01$

*** $p < .005$

**** $p < .001$

Figure 13. VOT and TOT discrimination functions for Spanish-speaking listeners. Significant differences between VOT and TOT discrimination performance are indicated with asterisks. Error bars represent the standard error of the mean.

* $p < .05$

** $p < .01$

*** $p < .005$

**** $p < .001$

In classical examples of categorical perception, category labeling boundaries predict discrimination patterns. It is therefore of interest to examine the extent to which VOT category labeling boundaries also predict TOT discrimination. There was a significant Pearson correlation between predicted VOT discrimination and obtained TOT discrimination responses in both English ($r = .452, p < .01$) and Spanish ($r = .239, p < .01$). It is noteworthy however, that these correlations are much weaker than those found between the two perceptual measures of VOT, especially for Spanish-speaking listeners.

Figure 14 shows the averaged predicted and obtained discrimination values for the English-speaking participants. A repeated measures ANOVA revealed that TOT discrimination performance was somewhat better than that predicted on the basis of labeling performance [$F(1,276)=62.630, p<.001$]. A follow-up MANOVA found that the TOT discrimination peak fell in the region of the predicted VOT discrimination peak, although the later was a more salient peak. (See Figure 14 for significance values.)

Figure 14. Obtained TOT and predicted VOT discrimination functions for English-speaking listeners. Significant differences between obtained and predicted values are indicated with asterisks.

* $p<.05$

** $p<.01$

*** $p<.005$

**** $p<.001$

Figure 15 shows the averaged predicted VOT and obtained TOT discrimination values for the Spanish-speaking participants. These listeners also showed TOT discrimination to be somewhat better than predicted VOT discrimination performance [$F(1,276)=24.097, p<.001$]. A follow-up MANOVA found that Spanish-speaking listeners have predicted heightened sensitivity to VOT only in the region of their perceptual category boundary. Notice how the two TOT discrimination peaks flank the predicted VOT discrimination peak. (See Figure 15 for significance values.)

Figure 15. Obtained TOT and predicted VOT discrimination functions for Spanish-speaking listeners. Significant differences between obtained and predicted values are indicated with asterisks.

* $p<.05$

** $p<.01$

*** $p<.005$

**** $p<.001$

2.4 Summary

This study replicated the results of a number of previous studies on the distributional and perceptual characteristics of VOT in English and Spanish (Lisker and Abramson, 1964; 1970; Abramson and Lisker, 1970; Williams, 1974; 1977a; 1977b). Distributions of labial stop consonants differ greatly between English and Spanish. English speakers contrasted their [+voice] and [–voice] stops with a distributional boundary at approximately +30 to +35 ms VOT. Spanish speakers had a distributional boundary between 0 and –17 ms VOT. In English, VOT production distributions were fairly well aligned with, but somewhat more positive than, category labeling boundaries and discrimination peaks. The TOT discrimination peak, in the vicinity of +15 ms TOT, was consistent with the English perceptual category boundary for VOT. In Spanish, as expected, distributional boundaries were more negative than VOT category labeling boundaries and discrimination peaks. Spanish-speaking listeners showed two small TOT discrimination peaks, one at –15 ms and one at +15 ms TOT, with their VOT perceptual category boundary falling between these two regions of heightened sensitivity. While linguistic experience may enhance or augment sensitivities to TOT stimuli in English-speaking listeners, it has no apparent effect on Spanish-speaking listeners' perception of TOT.

2.5 Discussion

This study was primarily designed to investigate how long-term linguistic experience and auditory mechanisms affect the perception of speech sounds and non-speech sounds that have analogous acoustic timing characteristics. The results indicate

that the distribution of phoneme categories affects the general placement of perceptual category boundaries, and that auditory discontinuities play a larger role in English perceptual [voice] categories than they do in Spanish perceptual [voice] categories. Consequently, linguistic experience with English stop consonants has a greater effect on the perception of TOT stimuli than experience with Spanish stop consonants.

In English, the line of demarcation between the distributions of labial stop consonant tokens was at approximately 30-35 ms VOT, with over 96% of [+voice] stops being produced with VOT values less than +25 ms ([+voice] mean = +10.8 ms VOT, pre-voiced mean = -106.7 ms VOT) and over 98% of [-voice] stops being produced with VOT values over +40 ms ([-voice] mean = 72.8 ms VOT). English-speaking listeners displayed categorical perception of the synthetic VOT stimuli, with a discrimination peak and category labeling boundary in the vicinity of +15 to +20 ms VOT. Although the distributional category boundary and the perceptual category boundary were slightly nonaligned, this is likely a function of the use of synthetic stimuli that have somewhat artificially high amplitude burst and aspiration values. While these levels of burst and aspiration may have slightly shifted the perceptual category boundary for these English-speaking listeners, they were needed to make the tokens sound more natural. The TOT results were also consistent with previous findings (Pisoni, 1977). Averaged TOT discrimination data revealed that English-speaking listeners have somewhat increased sensitivity near the negative auditory discontinuity and have substantially increased sensitivity near the positive auditory discontinuity. These are similar to the findings of Pisoni (1977); approximately half of his participants had a single discrimination peak in the vicinity of the positive auditory discontinuity, while the other half had two

discrimination peaks, one in each region of auditory discontinuity. Although Pisoni did not average his data, our averaged data suggest that English-speaking listeners also have increased sensitivity in the region between 0 ms TOT and the positive auditory discontinuity. These non-speech perception results suggest that not only are English-speaking listeners using temporal onset asynchrony as the primary cue to their [voice] categories, but that long-term experience with English augments sensitivity in the region of the positive auditory discontinuity relative to the negative auditory discontinuity.

The region of demarcation between Spanish [voice] categories is more band-like than line-like; consistent with previous researchers, we found that there is a gap in production from approximately -17 ms VOT to 0 ms VOT. Ninety-four percent of Spanish [+voice] stops were produced with VOT values more negative than -30 ms VOT. As mentioned previously, Spanish stop consonants are distributed in a way that *could* take advantage of the negative auditory discontinuity. VOT category labeling and discrimination data show, however, that Spanish-speaking listeners perceive their [voice] distinction categorically, with a category boundary at approximately 0 ms VOT. If Spanish-speaking listeners use differences in temporal onset asynchrony as their primary cue for categorizing VOT stimuli, then one would expect that they should have increased sensitivity to TOT stimuli in the region of 0 ms TOT. This, however, was not the pattern observed for our Spanish-speaking listeners; they display two small TOT discrimination peaks, one at the negative auditory discontinuity and one at the positive auditory discontinuity, with a trough in performance between the two. In fact, the Spanish perceptual category boundary falls almost completely between the two TOT discrimination peaks. This suggests that Spanish-speaking listeners likely do *not* use

temporal onset asynchronies as their primary cue to their [voice] distinction. Instead, as others have suggested, they may use the presence or absence of voicing during consonant closure to make [voice] judgments.

So far, I have suggested that both a positive and a negative auditory discontinuity exist and may be exploited during the course of language acquisition for the categorization of stop consonants. There is some evidence, however, that the positive auditory discontinuity may be generally more salient than the negative auditory discontinuity. The present results suggest that, for English-speaking listeners, linguistic experience may selectively augment sensitivity to the positive auditory discontinuity. VOT and TOT discrimination performance by Spanish-speaking listeners, however, suggests that the relative salience of the positive auditory discontinuity may be psychoacoustic in nature; the VOT discrimination peak, though having a maximum at 0 ms, is somewhat skewed towards more positive VOT values, and the positive TOT discrimination peak is broader than the negative one. Both of these results suggest that Spanish-speaking listeners may also have some increased sensitivity near the positive auditory discontinuity. In fact, Williams (1974) found that a group of Puerto Rican Spanish-speaking participants had two VOT discrimination peaks, a large one at approximately -5 ms VOT and a much smaller one at the English [voice] category boundary ($+20$ ms VOT). This second peak and the sensitivities seen in our data in the vicinity of the positive auditory discontinuity are not likely due to linguistic experience, but instead probably reflect the psychoacoustic salience of the positive auditory discontinuity. The third study in this dissertation will examine this possibility more closely.

Cross-linguistic studies on the perception of non-speech are important in helping to determine the effect that linguistic experience has on general auditory processing. While the effects found here are small, they nonetheless suggest that linguistic experience can affect the perception of non-speech sounds, at least in certain situations. Because the character of Spanish stop [voice] distinctions may not provide Spanish-speaking listeners with experience judging specific temporal onset asynchronies, their TOT discrimination functions reflect the basic discrimination abilities of the auditory system. They have two small discrimination peaks, reflecting their ability to judge differences between simultaneous onsets vs. non-simultaneous onsets with a threshold of approximately ± 15 ms. I expect that the sensitivities seen for our Spanish-speaking listeners are very similar to initial-state sensitivities (i.e., sensitivities at birth), unaffected by linguistic experience. English-speaking listeners have a long history of judging temporal onset asynchronies in the vicinity of the positive auditory discontinuity, and as such have somewhat augmented sensitivities in this region; this enhanced sensitivity does not transfer to the region near the negative auditory discontinuity, with which English speaking listeners have relatively little experience. So while English-speaking listeners show sensitivities in the regions of both the negative and the positive auditory discontinuities, due to general auditory biases, linguistic experience selectively augments discrimination performance only in the vicinity of the positive auditory discontinuity.

CHAPTER 3: Study 2. The Effect of Stimulus Range on VOT Perceptual Category Boundary Locations

3.1 Introduction

The second study examined the effect that linguistic experience and auditory discontinuities have on the phonemic category structure of labial stop consonants in both English and Spanish. The internal representation of phonemic categories should depend on the structure of the distribution of sounds in the environment; after all, it is the distributions themselves that provide the necessary information for forming speech sound categories.

Example 1: if the distributions of two phonemes are close to each other, or even slightly overlapping in acoustic space, then accurate perception requires precision (or a threshold) along the relevant acoustic dimension. The presence of auditory discontinuities should result in sharp and stable perceptual category boundaries. If, on the other hand, two distributions are well separated along an acoustic dimension, then that dimension is somewhat less informative and category labeling boundaries should be less sharp and less stable. This pattern of results was reported by Keating et al. (1981) in their cross-linguistic comparison of English and Polish stop consonants. They found that English-speaking listeners had sharp category labeling boundaries that were relatively unaffected by stimulus range. Polish-speaking listeners had less sharp category labeling boundaries, and the placement of the boundary was affected by the stimulus range. Keating et al. (1981) suggested that these cross-linguistic differences are due to the structure of the input distributions in English, which has immediately adjacent short- and

long-lag VOT distributions, and Polish, in which there is a gap between their long-lead and short-lag VOT distributions.

Example 2: the range of acoustic space that an input distribution occupies should be related to people's category goodness judgments. Some distributions of phoneme tokens are stable, while others are subject to speech rate effects, among other things. Kessinger and Blumstein (1997) showed that, cross-linguistically, distributions of short-lag stop tokens were relatively unaffected by variations in speaking rate, but that the distributions of long-lead and long-lag stop tokens varied directly with the rate of speech. The distribution of short-lag stop tokens seems to be bounded by the gap in the production data in the short-lead region, owing possibly to articulatory constraints, and by the auditory discontinuity at approximately +20 ms. Both long-lead and long-lag token distributions are similarly bounded on one end, but are free to be produced with longer lead and longer lag as speech is slowed.

This second study was a cross-linguistic examination of range effects on categorization of labial stops by English and Spanish speakers. Spanish, like Polish, contrasts long-lead with short-lag VOT categories, with a production gap between the two distributions. The Spanish [+voice] distribution, which is produced with long-lead VOT, is bounded only on one side, while the English [+voice] distribution, which is produced with short-lag VOT, is bounded on both sides. If the primary cue to the [voice] distinction in Spanish involves the detection of onset asynchronies, it is predicted that Spanish-speaking listeners will have a less sharp and less stable VOT perceptual category boundary than will English-speaking listeners. If, however, Spanish-speaking listeners use the presence or absence of voicing during the consonant occlusion as their primary

cue to the [voice] distinction (as will be argued in this dissertation), then they might also exhibit a threshold-based boundary that is stable across variation in stimulus range. The relative stability of [voice] categories should be evident in category goodness judgments.

3.2 Methods

3.2.1 Participants

A new group of twenty-four native English-speaking listeners and twenty-four native Spanish-speaking listeners served as participants in this study. There were equal numbers of males and females in each language group. All other recruitment criteria and procedures were identical to those described in Study 1.

3.2.2 Stimuli

The synthetic speech stimuli were identical to those described in Study 1.

3.2.3 Procedure

Participants were asked to categorize synthetically produced /ba/-/pa/ stimuli and then to judge the goodness of each token as a member of the chosen category on a scale from 1 to 5. Participants were presented with three different ranges of VOT stimuli. The first range, from -70 ms to $+20$ ms VOT, mimicked an idealized Spanish distribution of stop consonants (SR). The second range, from -20 ms to $+70$ ms VOT, mimicked an idealized English distribution of stop consonants (ER). The third range, from -45 ms to $+45$ ms, incorporated both auditory discontinuities, but was neither Spanish-like nor English-like in its distributional characteristics (MR). There were a total of 19 different

stimuli in each range condition varying in steps of 5 ms VOT. Within each range condition, each stimulus was presented 10 times and the order of stimulus presentation was randomized across all 190 trials. Each participant was tested in each of the range conditions, with short breaks between conditions. In total, the participants responded to 570 trials over the three different range conditions. The order of the SR and the ER conditions was counterbalanced across participants. The MR condition was always presented second. Each condition took approximately 8 minutes to complete.

Experimental hardware and software was identical to that used in Study 1.

3.3 Results

3.3.1 VOT Category Labeling

A two (Language) x nine (Stimulus) x three (Range) ANOVA was performed on the proportion /b/ identification responses for the stimuli that were common across range conditions. The main effects of Language [$F(1,1242)=424.657, p<.001$], Stimulus [$F(8,1242)=394.380, p<.001$], and Range [$F(2,1242)=119.095, p<.001$] were significant, as were all two- and three-way interactions (all $p's<.005$). As found in the first study, English-speaking listeners had a greater proportion of /b/ responses than did Spanish-speaking listeners. Bonferroni corrected post-hoc tests revealed that the proportion of /b/ responses was greatest in the ER condition and least in the SR condition, with the MR condition having an intermediate proportion of /b/ responses. Both across and within languages, all range conditions were significantly different from each other (all $p's<.005$), although the effect of Range was greater for English- ($F=81.512$) than for Spanish-speaking listeners ($F=39.634$). Unfortunately, it was impossible to calculate

50% cross-over boundaries and slopes for all of the participants; some of the range conditions elicited virtually single category responses from a small number of participants. Figures 16 and 17 show the category identification data for the English- and Spanish-speaking listeners. A follow-up MANOVA revealed that although there was an effect of range on all of the tokens for the English-speaking listeners (all p 's < .05), only a subset of the stimuli were susceptible to range effects for the Spanish-speaking listeners. For Spanish-speaking listeners the perception of stimuli that ranged from -20 ms to +5 ms VOT was affected by the experimental manipulation (all p 's < .005). There was no effect of Range on the proportion of /b/ identification for the +10 ms, +15 ms, and +20 ms VOT stimuli, which were virtually all identified as /p/.

Figure 16. VOT category labeling functions of English-speaking listeners in each of the three range conditions. Stimuli within the vertical dashed lines were included in the analysis.

Figure 17. VOT category labeling functions of Spanish-speaking listeners in each of the three range conditions. Stimuli within the vertical dashed lines were included in the analysis.

Sub-analysis on effects of Order of Presentation. Half of the participants received the Spanish Range (SR) of VOT stimuli first, while the other half received the English Range (ER) first. A sub-analysis on the effects of Order of presentation revealed that when English-speaking listeners heard the ER condition first, they had a higher proportion /b/ identification than if the ER condition was the third condition [$F(1,214)=8.312, p<.005$]. The reverse pattern was observed for Spanish listeners in the SR condition [$F(1, 214)= 7.504, p<.01$]; when Spanish-speaking listeners heard the SR condition first, they had a lower proportion /b/ identification than if the SR condition was the third condition. There was no effect of Order of presentation in the SR condition for

the English-speaking listeners or in the ER condition for the Spanish-speaking listeners (all p 's > .4). (See Figure 18). English-speaking listeners showed a greater proportion of /b/ identification when the MR condition followed the ER condition than when it followed the SR condition [F(1,214)=7.616, p < .01]. There was a similar trend for Spanish-speaking listeners, although it did not reach significance (p = .083). (See Figure 19).

Figure 18. The effects of Order of presentation on the VOT category labeling functions of English- and Spanish- speaking listeners, for conditions SR and ER. Stimuli within the vertical dashed lines were included in the analysis.

Figure 19. The effects of Order of presentation on the VOT category labeling functions of English- and Spanish-speaking listeners, for the MR condition. Stimuli within the vertical dashed lines were included in the analysis.

3.3.2 VOT Category Goodness Judgments

Category goodness judgments were calculated by multiplying each goodness rating by -1 if the token was identified as a /b/ and by $+1$ if the token was identified as a /p/. The resulting values were averaged across tokens. A two (Language) x nine (Stimulus) x three (Range) ANOVA was then performed on the goodness judgments for the stimuli that were consistent across range conditions. The main effects of Language [$F(1,1242)=459.916, p<.001$], Stimulus [$F(8,1242)=335.717, p<.001$], and Range [$F(2,1242)=106.757, p<.001$], were significant, as were all two- and three-way interactions (all p 's $<.005$). Figure 20 shows the structure of the goodness judgments for each language in each range condition.

Figure 20. Category goodness judgments of English- and Spanish-speaking listeners in each of the three range conditions. Category goodness judgments were calculated by multiplying each goodness rating by -1 if the token was identified as a /b/ and by $+1$ if the token was identified as a /p/. Circles represent averaged goodness judgments for each token. A goodness judgment close to 0 indicates that that token lies on, or near, the perceptual category boundary. VOT tokens that straddle the perceptual category boundary, in terms of goodness judgments, are indicated for each language and range condition.

Not surprisingly, and consistent with the VOT identification results, English-speaking listeners' goodness judgments were more negative than were Spanish-speaking listeners'. Planned comparisons revealed that effect of Range was greater for English-listeners [$F(2,621)=90.518, p<.001$] than for Spanish-speaking [$F(2,621)=27.397, p<.001$] listeners. Figure 20 shows that across all Range conditions Spanish-speaking listeners had greater separation in their goodness judgments for tokens that straddled their perceptual category boundary than did English-speaking listeners. This difference

suggests that the Spanish-speaking listeners might have more stable category boundaries than the English-speaking listeners.

A follow-up MANOVA revealed that, in English, there was a significant effect of Range for every token (all p 's < .05) except for the -20 ms VOT token (ns, $p = .074$), which was judged to be a good member of the category /b/. For Spanish-speaking listeners, only the tokens that ranged from -15 ms to +5 ms VOT showed a significant effect of range on category goodness judgments. The -20 ms VOT token was judged to be a good member of category /b/, while tokens +10 ms through +20 ms VOT were judged to be good members of category /p/.

3.4 Summary

This study found effects of stimulus range on VOT categorization and goodness judgments, for both English- and Spanish-speaking listeners. Overall, English-speaking listeners showed a larger effect of Range than did Spanish-speaking listeners. Both the identification data and the category goodness data suggest that Range had an effect on the perception of virtually all of the analyzed stimulus tokens for English-speaking listeners, but this effect was only seen on a subset of the stimuli for Spanish-speaking listeners.

3.5 Discussion

Consistent with the findings of a number of other studies (e.g., Brady and Darwin, 1978; Rosen, 1979; Keating et al., 1981), this study found that perceptual category boundaries can be shifted as a function of stimulus range. The stimuli that were consistent across Range conditions were judged to be more /b/-like in the English Range

and more /p/-like in the Spanish Range, a pattern seen in both groups of participants. Also consistent with previous findings (Brady and Darwin, 1979; Keating et al., 1981), the order of presentation of Range conditions had a significant effect on stimulus categorization. English-speaking listeners perceived the same stimuli to be more /b/-like when they heard the English Range first than when they heard the English Range last. Spanish-speaking listeners showed the opposite effect when labeling the Spanish Range; they perceived the same stimuli to be more /b/-like when they heard the Spanish Range last, as opposed to hearing it first. English-speaking listeners perceived the Middle Range as more /b/-like when it followed the English Range and more /p/-like when it followed the Spanish Range. All of these order effects seem to be the result of the previous context; when the majority of previous tokens were perceived as /b/ it increased the probability that a following ambiguous token would be perceived as a /p/. Conversely, when the majority of previous tokens were identified as /p/, the probability that following ambiguous token would be perceived as a /b/ increased. These results are consistent with Brady and Darwin's (1979) suggestion that the ranges of both the preceding and present blocks affect the position of the category boundaries with weighting in favor of the current block. More generally, these results are consistent with the idea that category boundaries should be determined by the history of the input distributions. English-speaking listeners had a higher proportion of category /b/ responses overall than Spanish-speaking listeners, due to the differences in their long-term history with their phonemic category distributions. However, the proportion of /b/ responses was also affected by the general context of a given range (i.e., the preceding ranges presented) and by the range of the current block of stimuli.

There are a number of notable differences between the result of the current study and those of Keating et al. (1981). First, Keating found that English-speaking listeners did not show any effects of stimulus range on their perception of synthetic VOT stimuli. Our English-speaking participants showed a fairly large effect of Range, shifting their VOT boundaries by more than 10 ms VOT between the Spanish Range and the English Range. It is possible that our, but not Keating's, English-speaking listeners were susceptible to *frequency* effects (see Helson, 1964), and attempted to assign equal numbers of stimuli to each category. However, in neither the current study nor the experiment by Keating et al. were there explicit instructions regarding the proportion of stimuli in each category. In a follow-up experiment, Keating et al. (1981) manipulated expectations about the frequency of the test stimuli and found only small effects of stimulus range (approximately 4 ms VOT). It is unclear why their participants did not show an effect of stimulus range. Keating et al. suggest that due to the structure of the distribution of English stops, with the [+voice] and [–voice] categories being immediately adjacent, English-speaking listeners must form a strict category boundary. This boundary is enhanced by the psychoacoustic salience of the positive auditory discontinuity. Although the English [voice] category boundary is consistent with the positive auditory discontinuity, there is ample evidence that the category boundary can be shifted. For example, perceptual category boundaries (Summerfield, 1981; Miller, Green, and Reeves, 1986; Miller and Volaitis, 1989; Volaitis and Miller, 1992) and production distributions (Summerfield, 1981; Miller and Volaitis, 1989; Volaitis and Miller, 1992; Kessinger and Blumstein, 1997) are affected by speech rate. While the presence of an auditory discontinuity may reinforce a category boundary, auditory

discontinuities do not yield immutable boundaries (see Chapter 4; also see Holt et al., 2004). Range effects in our English-speaking listeners are possibly the result of augmented sensitivity to differences in temporal onset asynchrony, English-speaking listeners' primary cue to their [voice] distinction.

Second, Keating et al. (1981) found the Polish-speaking listeners showed a large effect of stimulus range and suggested that this was the result of the structure of their input distributions. In particular, they noted that the gap in produced VOT values of Polish speakers between -30 and $+5$ ms, means that Polish-speaking listeners need not establish precise VOT category boundaries. The Spanish stop consonant distributions are very similar to the Polish distributions, and yet our Spanish-speaking listeners showed less of an effect of range than our English-speaking listeners. Overall, for Spanish-speaking listeners, there was less than a 5-ms VOT difference between the category labeling boundaries of the Spanish and English ranges. Following Keating et al., I suggest that the distributions of Spanish stop consonants affect the use of VOT as a [voice] cue for Spanish-speaking listeners. Given that these listeners do not exploit the negative auditory discontinuity in making [voice] judgments, they may disregard onset asynchrony information. Instead, they likely use a more qualitative cue, such as the presence or absence of voicing during consonant closure (see Chapter 5 for more discussion). This may explain the relatively small range effects observed for our Spanish-speaking listeners. It would also explain why range and category goodness effects were not seen for the short-lag VOT stimuli, which were always identified as good exemplars of the /p/ category. For Spanish-speaking listeners, categorization may be quite simple: if there is no voicing during the consonant, then that token is a /p/, or,

alternatively, if there is voicing during the consonant, then that token is a /b/. The range effects seen for the short-lead stimuli may result from Spanish-speaking listeners having little experience judging the presence or absence of voicing in this region; the stimuli that are never heard in the Spanish-language environment show the greatest effects of range in Spanish-speaking listeners.

The results of the current study suggest that English- and Spanish-speaking listeners may indeed be using different cues to their respective [voice] distinctions. While differences between the two language groups may derive from the relative stability of phoneme categories in each language, it is also possible that the ranges used in the study were somewhat unnatural for English-speaking listeners. All three ranges spanned the Spanish [voice] category boundary at approximately 0 ms VOT, allowing listeners to divide the series into two categories. Only two of the ranges, however, spanned the English category boundary; the Spanish Range (from -70 ms to +20 ms VOT) does not include the English boundary value. If our English-speaking listeners were biased to divide the VOT series into two categories, then that would increase the effects of Range for this group.

CHAPTER 4: Study 3. The Effects of Linguistic Experience and Auditory Discontinuities on Perceptual Learning

4.1 Introduction

The third study examined the relationship between linguistic experience, auditory discontinuities, and the perceptual learning of non-speech stimulus distributions. One of the primary motivations for this study was to examine the relationship between linguistic experience and auditory discontinuities in a manner that did not rely solely on measures of discriminability. While the first study found slight differences between the TOT discrimination performances of English- and Spanish-speaking listeners, it is still unclear whether these differences are maintained in tasks that impose a higher cognitive load, such as learning. It has been suggested that auditory discontinuities may constrain the initial parsing of the perceptual space (Holt et al., 2004; Kuhl, 1991; 1993; 2000a; Pisoni, 1977), which can in turn be utilized by language systems to create phoneme categories. Recent research has shown that auditory discontinuities also affect non-speech category learning, and that category boundaries can shift as a function of the input distributions (Holt et al., 2004). The present study was a partial replication of Holt et al. (2004), but it extends their findings in a number of ways. First, the present study addressed the question of whether linguistic experience (English vs. Spanish) affects the learnability of non-speech categories, when the location of category distributions are manipulated vis-à-vis the presence of auditory discontinuities. Recall that the first study in this dissertation showed that English-speaking listeners have a single, rather broad, TOT discrimination peak. This discrimination peak began in the region of the negative auditory discontinuity and increased somewhat linearly to reach a maximum in the region of the positive

auditory discontinuity. Spanish-speaking listeners, on the other hand, displayed two small discrimination peaks. They had one discrimination peak at each of the auditory discontinuities, with a slightly broader positive peak.

English- and Spanish-speaking listeners were trained on one of three different distributions of TOT stimuli; there were two Consistent conditions in which the TOT distributions straddled either the positive (+Consistent) or the negative (–Consistent) auditory discontinuity. A third Inconsistent condition had a distributional boundary at –2.5 ms TOT, and both auditory discontinuities fell within the TOT category distributions. If the learnability of non-speech categories is influenced by discrimination patterns, then English- and Spanish-speaking listeners should show different patterns of TOT category learnability. Based on their discrimination performance, English-speaking listeners should learn to categorize the +Consistent TOT distributions more quickly than the –Consistent distributions, and the Inconsistent condition should fall at an intermediate level of learnability. Spanish-speaking listeners, on the other hand, should show a somewhat reduced asymmetry between the –Consistent and +Consistent conditions, and both of these conditions should be more easily learned than the Inconsistent condition. It is possible, however, that linguistic experience does not affect the learnability of these non-speech categories. In this case, learnability patterns would be similar for both groups of participants.

The second question that the present study addresses is whether the positive and negative regions of heightened sensitivity along the TOT dimension have symmetrical characteristics. It was predicted that if onset asynchronies are easier to detect in the positive region than in the negative region, participants should learn to categorize the

stimuli in the +Consistent condition faster than those in the –Consistent condition. If, on the other hand, the auditory discontinuities in the positive and negative regions have symmetrical characteristics, then the categories in both the –Consistent and the +Consistent conditions should be easier to learn than those in the Inconsistent condition.

The final purpose of this third study was to examine whether perceptual learning in the non-speech domain causes perceptual category boundary shifts in the speech domain. Both English- and Spanish-speaking listeners were presented with a VOT categorization task before and after they were trained to categorize the TOT stimuli. If VOT and TOT are being processed using abstractly similar underlying auditory representations, then one might expect that VOT category labeling boundaries may be shifted following TOT category training. This type of boundary shift was not necessarily expected, however. English category labeling boundaries are relatively stable (Keating et al., 1981, but see Chapter 2 in this dissertation), and it is still questionable whether auditory discontinuities play any role in Spanish stop categories.

4.2 Methods

4.2.1 Participants

A new group of forty-eight native English-speaking listeners and forty-eight native Spanish-speaking listeners served as participants in this study. Sixteen participants from each language background served in each of the training conditions. There were equal numbers of males and females in each language group and condition. All other recruitment criteria and procedures were identical to those described in Study 1.

4.2.2 Stimuli

4.2.2.1 VOT Stimuli

The synthetic speech stimuli were identical those described in Study 1.

4.2.2.2 TOT Stimuli

The TOT stimuli were identical those described in Study 1.

4.2.3 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. The entire experimental session lasted between 40 and 60 minutes depending on how quickly participants learned their TOT categories.

Experimental hardware and software was identical to that used in Study 1.

4.2.3.1 VOT Categorization

Participants were asked to categorize a subset of the synthetically produced /ba/-/pa/ series described in Study 1. They were presented with stimuli ranging from –70 ms VOT to +70 ms in steps of 10 ms VOT, except in the range from –40 ms to +40 ms VOT where stimuli varied in steps of 5 ms VOT. The 23 stimuli included in this experiment were each presented 10 times, for a total of 230 trials. Each participant

received a unique random order of the 230 trials. This task took approximately 5 minutes to complete.

4.2.3.2 TOT Category Training

During the training, participants heard a distribution of single tokens that varied in TOT. After listening to each token, participants made a two-alternative forced choice response by clicking on one of two buttons labeled "A" and "C" to identify the token as being a member of category A or C. After they responded, the correct answer was displayed on the computer screen beside the labeled button (e.g., "Answer: A"). There were three possible training conditions: 1) +Consistent – the two category distributions of TOT sounds straddled +22.5 ms TOT, which corresponds to the positive auditory discontinuity. 2) Inconsistent – the two category distributions of TOT sounds straddled –2.5 ms TOT, which corresponds to neither the positive nor the negative auditory discontinuity. 3) –Consistent – the two category distributions of TOT sounds straddled –22.5 ms TOT, corresponding to the negative auditory discontinuity. Each participant was assigned to only one training condition.

Within each block of trials, participants were presented with a total of 46 stimuli, according to the distributions shown in Figure 21. After each block the percent correct categorization was recorded. Participants continued to respond to stimuli until they had obtained a criterion categorization performance of 90% correct. After participants reached criterion performance, they were presented with two additional blocks of trials. These were included to ensure that the participants had actually learned the categories,

and to reduce (in proportional terms) differences among participants in amount of exposure to the stimuli. Each block of trials took approximately 3 minutes to complete.

Figure 21. Stimulus input distributions reflecting the relative frequency of stimulus presentation as a function of TOT. Arrows represent the locations of natural auditory boundaries. Both the +Consistent training condition and the -Consistent training condition are consistent with natural auditory boundaries. The Inconsistent training condition is inconsistent with both natural auditory boundaries.

4.2.3.3 TOT Generalization

Upon completion of TOT category training, participants completed a TOT generalization test. The generalization test was included in order determine how TOT category training affects perceptual category labeling boundaries in TOT stimuli. It is unlikely that there would be any effect of TOT category training on the VOT perceptual category boundaries if participants did not show category boundary shifts in their labeling of TOT stimuli. Because some of the participants in the Holt et al. (2004) study show effects of training condition on the location of TOT perceptual category boundaries, we expected to find similar results here.

Participants were presented each TOT token from the entire TOT continuum (–70 ms to +70 ms) in steps of 5 ms TOT, for a total of 29 TOT stimuli. Each stimulus was presented 10 times, for a total of 290 trials. A unique random order of presentation was assigned to every participant. Participants were asked to categorize the TOT stimuli into category A or C in the same manner as they had during the training task, but no feedback was given.

4.2.3.4 TOT Category Refresher

To reinforce the TOT category learning before testing whether TOT training condition affects VOT category boundary locations, each participant completed a TOT category refresher. The refresher consisted of two additional blocks of TOT category training as described in the TOT Category Training section.

4.2.3.5 VOT Categorization

Following completion of the TOT category refresher, participants were again asked to categorize a series of synthetically produced /ba-/pa/ stimuli ranging from –70 ms VOT to +70 ms VOT. All procedures were identical to those described in 4.2.3.1.

4.3 Results

4.3.1 TOT Category Training

The number of blocks to reach the criterion of 90% correct performance was used to assess the learnability of each of the training categories. 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 than 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 grouped across languages revealed that although the +Consistent training condition was learned in significantly fewer blocks than the Inconsistent condition ($p<.05$), there was no significant difference between the –Consistent and Inconsistent conditions ($p=.528$). Planned comparisons done on each language group individually revealed that English-speaking listeners learned to correctly categorize the +Consistent distributions in fewer blocks (1.44 blocks) than the Inconsistent distributions (3 blocks) ($p<.05$), which did not differ from the –Consistent condition (2.69 blocks) ($p=.662$). For the Spanish-speaking listeners, however, only the –Consistent (5.94 blocks) and the +Consistent (2.19 blocks) conditions significantly differed from each other ($p<.05$). The Inconsistent (4.5 blocks) training condition did not differ significantly from either the –Consistent ($p=.381$) or the +Consistent ($p=.162$) conditions. (See Figure 22).

Figure 22. *Effect of Language and Training Condition on the number of training blocks to criterion by English- and Spanish-speaking listeners. Error bars represent the standard error of the mean. Asterisks indicate significant differences at the .05 level.*

4.3.2 TOT Generalization

The data that ranged from the last 100% category A (or C) response to the first 100% category C (or A) response were extracted from each generalization function for the purposes of linear regression analysis. Certain participants never displayed 100% category identification for one or the other, or both, categories. In these situations data that ranged from the highest percent category A (or C) response to the highest percent category C (or A) response were extracted for the linear regression analysis. In the –Consistent 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 perceptual category boundaries. Additionally, one English participant in the Inconsistent condition showed a U-shaped function. 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 (–Consistent) or the boundary that was most prominent (Inconsistent) was selected for the linear regression analysis. One Spanish-speaking listener from the +Consistent condition was excluded from the analysis because she did not show any perceptual category boundary at all, and thus it was not possible to obtain a significant linear curve fit to her data.

The slopes and 50% cross-over points of the regression lines were analyzed in a MANOVA, with Language and Training Conditions as independent variables. Overall, there was a significant main effect of Training Condition on both TOT perceptual category boundary location [$F(2,89)=32.823, p<.001$] and on the slope of the category labeling function near the boundary [$F(2,89)=3.885, p<.03$]. Participants in the +Consistent condition had a more positive TOT boundary value than those in the Inconsistent condition, which in turn had a more positive boundary value than those in –Consistent (all p 's<.05). Generally the slope of the boundary was shallower for participants in the –Consistent condition than for participants in the other two conditions ($p<.05$), possibly owing to the presence of the few aforementioned u-shaped identification functions. (See Figure 23).

Planned comparisons within each language were done to insure that the training had been successful in affecting the locations of the TOT category labeling boundaries. For both English- and Spanish-speaking listeners, the value of the TOT category labeling boundary was more negative following training on the –Consistent distributions (–7.34 ms TOT and –6.21 ms TOT, respectively) than following training on the Inconsistent distributions (+7.49 ms TOT and +13.36 ms TOT, respectively; English, $p<.002$;

Spanish, $p < .001$). There were, however, no significant differences between the TOT category labeling boundaries following training on the Inconsistent distributions and +Consistent distributions (+15.02 ms TOT and +18.96 ms TOT, for English and Spanish, respectively) in either language (all p 's $> .10$).

Figure 23. TOT generalization functions for English- and Spanish-speaking listeners in each of the three training conditions.

TOT Generalization Separated by Learning Strategy. Given that participants who displayed U-shaped generalization functions obviously learned something different than those who displayed a single TOT perceptual category boundary, it seems useful to conduct the same analyses excluding these seven participants. The slopes and 50% cross-over points of the resulting regression lines were analyzed in a MANOVA, with Language and Training Conditions as independent variables. Overall, there was still a significant main effect of Training Condition on TOT perceptual category boundary location [$F(2,82)=25.376, p<.001$], but no longer a significant effect on the slope of the labeling function ($p=.081$). Participants in the +Consistent condition had a more positive TOT boundary value than those in the Inconsistent condition, which in turn had a more positive boundary value than those in the –Consistent condition (all p 's<.01). The Language X Training Condition interaction approached significance ($p=.057$) suggesting a possible tendency for TOT category boundary locations to be differentially affected by Training Condition for both language groups.

Planned comparisons within each language revealed that, for English-speaking listeners, the location of the TOT category labeling boundary did not differ significantly between the –Consistent (+1.58 ms) and the Inconsistent (+7.35 ms) conditions ($p=.09$). However, there was a significant difference between the Inconsistent and +Consistent (+15.02 ms) perceptual category boundary locations. Spanish listeners displayed a significant difference between TOT boundary locations for the –Consistent (–4.42 ms) and the Inconsistent (+13.36 ms) conditions ($p<.001$), but not between the Inconsistent and the +Consistent (+18.96 ms) conditions ($p=.156$). (See Figure 24).

Figure 24. TOT generalization functions for English- and Spanish-speaking listeners in each of the three training conditions. Data from participants with U-shaped generalization functions were excluded.

It might be expected that the degree to which Training Condition affects TOT perceptual category boundary locations is mediated by the number of training blocks to criterion, but correlational analyses revealed that this was generally not the case. There were no significant correlations between the number of blocks of training and the location of the TOT category labeling boundaries for English-speaking listeners in any of the Training Conditions (all p 's > .4). There were, however, significant Pearson

correlations between the number of blocks of training and the location of the TOT category labeling boundaries for Spanish-speaking listeners in the –Consistent ($r=.557$, $p<.05$) and the Inconsistent conditions ($r=.591$, $p<.05$), indicating that the greater the number of training blocks the more positive the TOT perceptual category boundary. This suggests that more training does not necessarily reinforce the distributional boundaries of the training categories, but instead may direct attention to the positive, and possibly more naturally salient, side of the TOT dimension.

TOT generalization data from the seven participants who displayed U-shaped functions are displayed in Figure 25. Although these participants did not learn the categories defined by the experimenter, the boundaries they did exhibit between the simultaneous onset/non-simultaneous onset categories occur in the general vicinity predicted by the positive and negative auditory discontinuities reported by Pisoni (1977).

Figure 25. Categorization data from English- and Spanish-speaking listeners who displayed U-shaped TOT generalization functions.

4.3.3 VOT Categorization

Linear regression analyses were performed on the data that ranged from the last 100% /b/ identification response to the first 0% /b/ identification response. The slopes and the 50% labeling cross-over boundaries of the regression lines were analyzed in a MANOVA, with Language, Trial (pre- vs. post-training) and Training Condition as independent variables. As expected, there was a significant main effect of Language on VOT perceptual category 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 functions 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 did not affect the perception of VOT stimuli. Interestingly, there was a significant Language x Trial interaction [$F(1,180)=8.966, p<.005$]; English-speaking listeners showed a less positive VOT labeling boundary following TOT training, whereas Spanish-speaking listeners showed just the reverse effect. While there appears to be a similar pattern in each of the training conditions, analyses on individual training conditions revealed that this pattern was only significant for the –Consistent condition ($p=.05$). (See Figure 26). Additionally, there were no significant correlations between the number of blocks to criteria performance and the VOT boundary location or the amount of VOT boundary shift post TOT category training.

Figure 26. VOT identification functions for English- and Spanish-speaking listeners, before and after TOT category training, in each of the three training conditions.

4.4 Summary

This study was successful in both replicating and extending the results of Holt et al. (2004). Both English- and Spanish-speaking listeners showed enhanced learnability of non-speech categories that were consistent with the positive auditory discontinuity at approximately +20 ms TOT. Surprisingly, neither language group showed enhanced learnability of TOT categories that were consistent with the negative auditory discontinuity at approximately –20 ms TOT. Despite differences in ease of learning, all categories were ultimately learnable, although the strategies used varied somewhat between participants. In this study, short-term experience with the categorization of TOT stimuli did not affect the perception of VOT in a predictive manner.

4.5 Discussion

The present study investigated the relationship between linguistic experience, auditory discontinuities, and perceptual learning. It was designed as a partial replication of Holt et al. (2004), with the additional intention of addressing three specific questions: (1) Does linguistic experience affect the perception and learnability of non-speech categories. (2) Are auditory discontinuities symmetrical in their effect on category learning? (3) Does learning non-speech categories affect speech perception?

The results from the present study suggest that not all auditory categories are equally learnable. Learnability is related to, but not completely dependent on, discriminability. Both English- and Spanish-speaking listeners have enhanced discriminability in the region straddling +20 ms TOT (findings from Study 1), and members of both language groups have enhanced learnability when TOT categories

straddle that same region. This result is a replication of Holt et al. (2004). Where learnability and discriminability fail to be aligned is in the region near the negative auditory discontinuity. Surprisingly, TOT category learning was either unaffected (English) or hindered (Spanish) by the presence of the negative auditory discontinuity at the distributional boundary, even though discrimination patterns show increased sensitivity in that region. Therefore, any long-term effect of linguistic experience on the perception of TOT, as seen in cross-linguistic differences in TOT discriminability, do not seem to translate into differences in learnability.

It is possible that English speakers' generally more rapid learning of the TOT categories is in fact the result of linguistic experience. If English-speaking, but not Spanish-speaking, listeners focus on onset asynchronies in perceiving VOT differences, then it might be predicted that English listeners would have greater overall sensitivity, and hence greater learnability, of TOT categories. In the first study, however, listeners from both language groups showed similar overall discrimination performance, although their patterns of discrimination were stimulus-dependent. It is likely that faster learning in this task is the result of different levels of familiarity with these types of tasks, and not the result of linguistic experience per se. Participants from the US and Mexico were selected to form a homogenous population within each language group. All participants were undergraduates at a major university, were monolingual, and were from the same region in which they were attending school. Despite the similarities between the participants, it is unlikely that the mean socio-economic level and educational experience of the language groups were perfectly matched. The English-speaking participants may

have had greater previous exposure to the kind of abstract learning tasks used in this study.

Differences in psychoacoustic salience of auditory discontinuities may be somewhat affected by, but are probably not determined by, linguistic experience. English speakers, both in the present study and in Pisoni (1977), show increased sensitivity in the region of +20 ms TOT relative to the region of –20 ms TOT. This pattern of results might be predicted if TOT perception is affected by linguistic experience with native VOT categories. Experience with VOT reinforces the positive, but not the negative, auditory discontinuity in English-speaking listeners. Because Spanish stop consonant categories do not straddle either the positive or the negative auditory discontinuity, Spanish-speaking listeners might be expected to have comparable sensitivity to both auditory discontinuities. However, this is not what was observed in this study. Both language groups showed enhanced learnability of TOT categories that straddled the positive auditory discontinuity, but received no benefit from the alignment of the negative auditory discontinuity with the distributional boundary. This suggests that asymmetrical learnability with respect to the positive and negative auditory discontinuities is at least partially due to differences in their inherent psychoacoustic salience.

The results of the TOT generalization test also support this contention. In a learning environment free from auditory discontinuities, participants should display perceptual category boundaries that are in line with the category boundaries of the input distributions. In this study, TOT perceptual category boundaries were shifted markedly from the line of demarcation of the input distribution boundaries, especially in the

Inconsistent and –Consistent conditions. In these two conditions the positive auditory discontinuity seems to have been a salient factor in the placement of TOT perceptual category boundaries. Since the line of demarcation between the input distributions in the Inconsistent condition was centered around –2.5 ms, it would be expected that perceptual category boundaries might be drawn towards the negative auditory discontinuity, if the two auditory discontinuities were of equal salience. The fact that in the Inconsistent condition, both English and Spanish listeners had positive TOT perceptual category boundaries supports the idea that the positive auditory discontinuity is more psychoacoustically salient.

The third purpose of the study was to examine whether short-term experience with these non-speech (i.e., TOT) sounds would affect the perception of VOT. Given the participants' relatively brief exposure to the TOT stimuli, it seemed unlikely that there would be any effect of training on the perception of VOT. The TOT generalization task revealed that participants were partially successful at learning the TOT categories. The results for the Inconsistent and +Consistent conditions patterned together, with similar positive TOT perceptual category boundaries. Training in the –Consistent condition tended to produce TOT generalization functions with slightly negative perceptual category boundaries. Despite the facility with which most participants learned the TOT categories, there was relatively little effect of training on the perception of VOT. The only marked effect of training was that English listeners tended to have a less positive VOT category labeling boundary after TOT category training than before training, while Spanish listeners tended to show the reverse effect. As the current study did not include a group of participants who received no TOT category training, it is not possible to know

for certain whether this change in VOT category boundary location is the result of training itself, or simply the result of doing the VOT labeling task a second time. A sub-analysis of the results of Study 2 suggests the latter. The boundary shifts were in the same direction as seen in Study 2. When English-speaking listeners hear the VOT series a second time, it is in the context of having perceived many tokens as /b/ during the first VOT labeling task. They therefore perceive fewer /b/s the second time they do the labeling task, even though the stimuli were identical. Spanish-speaking listeners perceived more /b/s the second time through the labeling task as a result of having perceived many /p/s during the first VOT labeling task.

Since this study was intended to be a partial replication of Holt et al. (2004), it is appropriate to note some similarities and differences in the results of the two studies. First, both studies found that there was a large amount of variability in the number of training blocks to criterion. For example, in the present study, one quarter of the Spanish-speaking listeners learned the –Consistent condition in a single training block while another quarter took over 15 blocks to learn the same TOT categories. This level of variability was also reported Holt et al. (2004); they found that participants took an average of 3.36 blocks to achieve criterion in the Consistent condition (identical to the +Consistent in this study) of their first experiment, while in the corresponding condition of the second experiment, subjects require an average of 6.57 blocks to achieve criterion. Additionally, Holt et al. found no significant difference between the results for the Inconsistent condition in their first experiment and the results for the Consistent condition in their second experiment. They attribute this performance difference between the two

experiments to a difference in subjects sampled at the beginning vs. the end of an academic semester (personal communication).

Second, and related to issues of variability, the participants in the current study were generally faster at learning the task than those in Holt et al. (2004). While participants in this study required an average of only 1.81 blocks to learn the +Consistent distributions, the participants in an identical condition of the Holt et al. study required an average of 3.36 blocks (Experiment 1) or 6.57 blocks (Experiment 2) to achieve the same learning criterion. The corresponding numbers for the Inconsistent condition were 3.75 (present study) and 16.86 (Holt et al. 2004, Experiment 1). It seems unreasonable to conclude that students at the University of Texas at Austin and the University of Puebla, are smarter or faster learners generally than students from Carnegie Mellon (although it is possible that this is the case!). Two of the three training distributions were identical in both studies, and yet the speed at which participants learned the categories was quite different. What might explain this discrepancy? In both studies, participants were presented with a single token on each trial, made a two-alternative forced choice categorization response, and were given feedback. The only difference between the two procedures was that in the Holt et al. study participants responded via a button box and were reinforced with a light above the correct answer. There were no labels on the button box. In the present study, participants responded by clicking a labeled button ('A' or 'C') on a computer screen, and were reinforced with a written response (e.g., Answer: A). It is possible that the process of assigning category labels enhanced categorization in the current participants; similar effects have been seen for infants as young as nine months old (see Waxman and Markow, 1995; Balaban and Waxman, 1997) .

CHAPTER 5: Summary and Interpretation of Findings

Speech and non-speech sounds are often assumed to be processed independently (Liberman et al., 1967; Liberman and Mattingly, 1985). This dissertation sought to test this assumption by examining whether linguistic experience has an effect on the perception of non-speech sounds that have analogous acoustic characteristics to speech sounds. This research was also designed to investigate the joint effects of linguistic experience and auditory mechanisms on phoneme structure and category learning. Three cross-linguistic studies were designed to look at (1) the production and perception of VOT and the perception of TOT, (2) the effects of stimulus range on the perception of VOT, and (3) the effects of auditory discontinuities on non-speech category learnability. Results indicate that linguistic experience does affect the perception of non-speech stimuli, at least in certain circumstances. Thus, there is some commonality in the processes used to discriminate between non-speech sounds and those used to discriminate between speech sounds. Additionally, auditory discontinuities were found to influence both phoneme structure and category learning.

5.1 Summary of Studies

In order to investigate the effects of linguistic experience and auditory mechanisms on (1) perception of speech and non-speech stimuli, (2) phoneme category structure, and (3) category learning, I conducted three studies that each addressed one or more of these issues. Noteworthy results are summarized here.

5.1.1 Study 1

Production data confirmed that the English labial [voice] distinction consists primarily of a short-lag/long-lag VOT contrast, with a distributional boundary in the vicinity of +30 to +35 ms VOT. Spanish speakers produce their [+voice] labial stops with a substantial amount of pre-voicing and their [–voice] labial stops with short-lag VOT values. Spanish speakers produce very few tokens in the region between 0 ms and –30 ms VOT.

Listeners from both language backgrounds showed categorical perception of a series of VOT stimuli. VOT discrimination functions predicted by VOT labeling functions were highly correlated with obtained VOT discrimination functions for the same listeners. English-speaking listeners had a perceptual category boundary at approximately +15 to +20 ms VOT, while Spanish-speaking listeners' perceptual category boundary was at approximately 0 ms VOT.

English-speaking listeners showed a single broad TOT discrimination peak, with heightened sensitivity to TOT commencing in the region of the negative auditory discontinuity and increasing to a maximum in the region of the positive auditory discontinuity, which corresponds to their perceptual category boundary (i.e., at +15 ms TOT). Spanish-speaking listeners showed two small TOT discrimination peaks, one in the region of the negative auditory discontinuity (–15 ms TOT) and one in the region of the positive auditory discontinuity (+15 ms TOT). Both their predicted and obtained VOT discrimination peaks fell between their two TOT discrimination peaks. The results of the first study challenged assumptions that speech and non-speech are processed

independently, as it was shown that linguistic experience affects the perception of non-speech sounds.

5.1.2 Study 2

The second study was designed to investigate how linguistic experience and auditory discontinuities affect phoneme category structure. In a stimulus range experiment, linguistic experience, stimulus range, and order of stimulus range condition all were found to affect category labeling. Across all range conditions, English-speaking listeners had higher percentage of /b/ identification responses than Spanish-speaking listeners. English-speaking listeners were also subject to a larger effect of stimulus range (a boundary shift of approximately 10 ms VOT) than Spanish-speaking listeners (a boundary shift of less than 5 ms VOT). English-speaking listeners perceived fewer /b/s when the preceding range conditions had many tokens that were perceived as /b/, while Spanish-speaking listeners heard fewer /p/s when the preceding range conditions had many tokens that were perceived as /p/. Results indicate that category input distributions affect phoneme category structure.

5.1.3 Study 3

The results of the third study suggest that the presence of an auditory discontinuity at a distributional boundary can facilitate perceptual learning of non-speech categories. Participants in both language groups learned the categories in the +Consistent training condition (i.e., the TOT distributional boundary was consistent with the positive auditory discontinuity) in fewer blocks than they did the categories in either of the other

two training conditions. Therefore, only the positive auditory discontinuity facilitated category learning. Results from the TOT generalization task suggest that participants may use somewhat variable learning strategies. There were no notable effects of language experience on TOT category training, suggesting that discrimination functions and learnability do not have a one to one mapping. This third study also tested whether short-term experience categorizing TOT stimuli affects subsequent perception of VOT stimuli. Neither group of listeners showed any effects of TOT training condition on the perception of VOT; however, training in general resulted in a somewhat less positive VOT labeling boundary for English-speaking listeners and a somewhat more positive VOT labeling boundary for Spanish-speaking listeners.

5.2 General Discussion

There are a number of general themes that have emerged from the results of this dissertation, some expected and some unexpected. Recall that the primary purpose of this dissertation was to investigate the role of linguistic experience and auditory mechanisms on the perception of speech and non-speech sounds, phoneme category structure, and perceptual category learning. Instead of treating each of these issues independently, I will use the findings to show how they are all inter-related in a general framework.

5.2.1 Linguistic Experience Affects Perception of Non-Speech Sounds

Germane to the original goals of this dissertation, I found that linguistic experience affects the perception of non-speech sounds, at least in some circumstances.

TOT discrimination tasks revealed that English-speaking listeners have a single, broad TOT discrimination peak, with somewhat increased sensitivity near the negative auditory discontinuity and substantially increased sensitivity near the positive auditory discontinuity. Spanish-speaking listeners have two small discrimination peaks, one in the region of each auditory discontinuity. I propose that linguistic experience with English stop consonants has a greater effect on the perception of TOT stimuli than experience with Spanish stop consonants; Spanish TOT discrimination functions reflect basic auditory sensitivities that are unaffected by linguistic experience, while English TOT discrimination functions reflect augmented sensitivities as a result of linguistic experience.

In order for linguistic experience to affect the perception of non-speech sounds, they must be processed using some of the same underlying mechanisms as speech sounds. A recent study by Bent et al. (2003) found that English- and Mandarin-speaking listeners differed in their ability to discriminate both Mandarin tones and non-speech pitch contours that were analogous to the pitch contours found in natural Mandarin speech. The two groups of listeners did not differ, however, in their discrimination of non-speech pure-tone frequency and pulse-train repetition rate. Bent et al., (2003) found that linguistic experience affects non-speech perception only when mechanisms used to perceive the non-speech sounds are also used for speech sound categorization.

Together, my results and those of Bent et al. (2003), suggest that speech and non-speech are not processed as independently as some have assumed (see Liberman et al., 1967; Liberman and Mattingly, 1985). Speech perception not only relies on the operating characteristics of the auditory system, but language experience also affects the perception

of some non-speech sounds that are processed using the same underlying auditory mechanisms as speech. While these results are exciting for those of us who want to make claims about general learning mechanisms, they should also be taken as a cautionary note by those who use non-speech sounds to make claims about language processing. Since non-speech stimuli are often used in experiments to control for prior experience, care must be taken when designing experiments to ensure that there is an understanding of how linguistic experience affects prior experience with stimuli even when they are non-speech sounds. Additionally, it should be noted that the effects found here were limited to a very specific task, namely discrimination. No interesting cross-linguistic differences were found for TOT category learning.

5.2.2 Cues to the Spanish [Voice] Distinction

When I undertook this project, I assumed that both English- and Spanish-speaking listeners used the same general strategies for making [voice] judgments, but simply developed different perceptual category boundaries as a result of experience with their respective input distributions. I, like others, believed that the misalignment of Spanish labial stop distributions and their perceptual category boundaries was nothing more than an effect of poor stimulus construction for perceptual experiments; I believed that since Spanish stop consonants were distributed in a manner that *could* take advantage of the negative auditory discontinuity, their perceptual category boundary was likely to occur in the vicinity of –20 ms VOT.

The results of the production portion of this dissertation were consistent with previous findings on the distribution of Spanish labial stops (Lisker and Abramson, 1964;

Williams, 1974; 1977a). Ninety-four percent of Spanish [+voice] stops were produced with more than 30 ms of pre-voicing and 100% of Spanish [–voice] were produced with simultaneous or short-lag VOT. The absence of Spanish stop consonants having VOT values between 0 ms and –20 ms or –30 ms may have an aerodynamic cause (see Kingston and Diehl, 1994), as similar gaps in produced VOT values are observed in other languages (Lisker and Abramson, 1964). Despite the presence of the negative auditory discontinuity near the Spanish distributional category boundary, Spanish-speaking listeners consistently had perceptual category boundaries in the vicinity of 0 ms VOT. Varying the stimulus range, as was done in the second study, provided contexts in which Spanish-speaking listeners could have been more or less affected by the presence of auditory discontinuities. However, Spanish-speaking listeners showed only very small range effects. In fact, in every range condition, Spanish-speaking listeners identified the 0 ms VOT token as /b/ more often than they identified it as /p/. The reverse was true for the +5 ms VOT token, which was identified as a /p/ more often than /b/. The consistency with which Spanish-speaking listeners use 0 ms VOT as their [voice] category boundary coupled with their inability to detect temporal onset asynchrony differences in non-speech sounds, supports the idea that Spanish-speaking listeners do not use temporal onset asynchrony when making [voice] judgments.

The first study also suggests that English-, but not Spanish-speaking listeners, use temporal onset asynchrony as an important cue to their [voice] distinction. If temporal onset asynchrony were an important cue to the Spanish [voice] distinction, our Spanish-speaking listeners should have been very good at making temporal onset asynchrony judgments of TOT stimuli in the region of their VOT perceptual category boundary (i.e.,

0 ms TOT). Surprisingly, it is English-, and not Spanish-speaking listeners who have augmented sensitivity to TOT in this region. The fact that Spanish-speaking listeners do not have enhanced TOT discriminability near their VOT perceptual category boundary, provides further evidence that they may not use temporal onset asynchrony as a primary cue to their [voice] distinction.

I suggest that Spanish-speaking listeners may be using the presence or absence of voicing during the consonant closure and/or release as the primary cue to their [voice] categories. (This idea has also been suggested by Green et al., 1998, for Spanish and by Keating et al, 1981, for Polish). Unlike temporal onset asynchrony judgments, detection of low frequency energy during the otherwise silent closure interval is a qualitatively simple task. Studies of consonant confusion matrixes have show that voicing and nasalization are the most perceptually robust features in the presence of noise or with low-pass filtering (Miller and Nicely, 1955). In both voiced and nasal consonants, there is substantial low frequency energy. When that low frequency energy is removed via high-pass filtering, the perceptual advantages of voicing and nasalization are eliminated and sounds are rendered virtually inaudible. These findings from Miller and Nicely (1955) suggest that a decision rule based on the presence or absence of voicing during consonant closure and/or release would yield fairly robust categorization. As such, Spanish-speaking listeners may not require more than a few milliseconds of pre-voicing in order to judge tokens as being members of their [+voice] category. Also, Spanish-speaking listeners may be reliably able to detect voicing at the moment of release, even though it is accompanied by higher frequency components. Voicing during the period of the burst may be easily detected (i.e., it will not get masked) since the voicing component

is generally of higher intensity and is spectrally separated from the higher frequency aperiodic burst components. These proposals, which were not formally tested in this dissertation, clearly warrant further investigation.

In the introduction, I suggested that it would be communicatively advantageous for languages to use auditory discontinuities when partitioning their phonemic space, especially since infants start out being particularly sensitive to VOT stimulus differences that straddle the auditory discontinuities (Aslin et al., 1981; Lasky et al., 1975, Eimas et al., 1971; see Jusczyk et al., 1980 for non-speech results). However, Spanish-speaking listeners may not use the negative auditory discontinuity when making [voice] judgments for two reasons. This hypothesis is supported in a number of ways. First, the presence or absence of voicing during the consonant is perceptually very salient, may be easily learned, and, as a basis for [voice] judgments, is likely to be computationally simpler than processing of relative temporal onset asynchrony. Second, as this research shows, the negative auditory discontinuity might not actually be very psychoacoustically salient.

5.2.3 Asymmetrical Auditory Discontinuities

Across both studies 1 and 3, asymmetries between the positive and negative auditory discontinuities were apparent. Although the TOT discrimination asymmetries observed for our English-speaking listeners appear to have been partly mediated by linguistic experience, linguistic experience itself cannot account for all of the data. First, Spanish-speaking listeners had a broader TOT discrimination peak in the vicinity of the positive auditory discontinuity than in the vicinity of the negative auditory discontinuity. Second, despite having a perceptual category boundary in the vicinity of 0 ms VOT,

Spanish-speaking listeners showed significantly better VOT discrimination near the positive auditory discontinuity than was predicted by category labeling performance. Third, in the TOT training study, only the positive auditory discontinuity was found to have a facilitating effect on category learning; the presence of the negative auditory discontinuity at a distributional boundary either had no effect on (English) or actually hindered (Spanish) category learnability. Together, these findings suggest that there is a psychoacoustic explanation for asymmetries between the positive and the negative auditory discontinuities.

It is well established that low frequency energy is more effective at masking high frequency energy than vice versa, owing to the response properties of the basilar membrane. Thus, for negative TOT stimuli, the onset of the high frequency component may be partially masked by the low frequency component. The situation is different for positive TOT stimuli, where the leading high frequency component cannot mask the onset of lagging low frequency component. The salience of the negative auditory discontinuity may be even more compromised by the spectral properties of natural speech where lower frequency components (i.e., the voicing bar and F1) tend to have greater intensity than higher frequency components (i.e., F2 and F3). It is therefore not surprising that the negative auditory discontinuity may not be exploited by languages in the strategic placement of phoneme categories. The TOT stimuli used in the current studies were created to mimic the energy distribution in speech, and thus it is not surprising that there were auditory discontinuity asymmetries found in both TOT discrimination and TOT category learning tasks.

A recent study by Burnham (2002) also suggests that temporal onset asynchrony differences are perceived more categorically in the positive region than in the negative region, even in a language that has three phoneme categories along the VOT dimension. Burnham found that Thai-speaking listeners had steeper labeling functions between their short-lag/long-lag [voice] categories than between their long-lead\short-lag [voice] categories, a pattern that was consistent across three age groups (6 year-olds, 8 year-olds, and adults). Also supporting the idea that the positive auditory discontinuity is more psychoacoustically salient is his finding that the degree of categorality was consistent across age group for positive VOT stimuli, but improved with age for negative VOT stimuli. While English-speaking listeners maintained high levels of performance on the positive VOT stimuli, there was no improvement with age on the categorality of labeling of negative VOT stimuli, which were always poorly categorized. Burnham also trained his participants (both Thai- and English-speaking) to identify the endpoints of either a negative TOT series (i.e., from -70 ms to $+20$ ms TOT) or a positive TOT series (i.e., from -20 ms to $+70$ ms TOT) and found that there was a tendency for the positive series to be labeled more categorically than the negative series, although these effects were much smaller than those found with VOT stimuli. He did not find any cross-language differences in the perception of TOT stimuli.

5.2.4 Auditory Discontinuities Do Not Yield Immutable Boundaries

There is direct evidence from this dissertation to suggest that auditory discontinuities do not create immutable boundaries. First, English-speaking listeners showed a substantial perceptual category boundary shift as a function of stimulus range,

despite the fact that the English perceptual category boundary is consistent with the positive auditory discontinuity. Second, in the TOT category training study, although the +Consistent distributions were easiest to learn, all of the distributions were ultimately (and surprisingly easily) learnable.

For the newborn infant, auditory discontinuities act to provide an initial partitioning of the acoustic space. As we have seen in this dissertation, not all languages take advantage of the presence of these auditory discontinuities when assigning their phoneme categories. It follows, therefore, that auditory discontinuities do not dictate the location of category boundaries.

A similar suggestion was made by Kuhl (1991, 1993, 2000a) in describing the role of “natural auditory boundaries” in language acquisition. She proposed that these innate boundaries account for infants’ initial discrimination abilities. Through experience with native-language distributions, and phoneme categorization, these boundaries are rearranged and sometimes deleted. While the view taken here is quite similar to that of Kuhl’s, the present results do not indicate that the auditory discontinuities based on temporal onset asynchrony judgments are themselves shifted or deleted through linguistic experience. There is much evidence that the ability to discriminate non-native speech sounds can be attenuated as a result of the acquisition of native-language phoneme categories (Werker and Tees, 1984; Best, 1993; 1994). However, I suggest that the underlying perceptual mechanisms that facilitated discrimination in the first place, remain intact, although sensitivities may be enhanced. This is clearly the case with the auditory discontinuities studied in this dissertation. In the current research, both auditory

discontinuities were apparent, but the size and shape of the TOT discrimination peaks were language dependent.

Additional support for the maintenance of auditory discontinuities, even when they are not used in the service of native category boundaries, comes from studies by both Williams (1974) and Streeter (1976a). Williams found that Spanish-speaking listeners had a primary discrimination peak at approximately 0 ms VOT, in accordance with their category labeling boundary, but they also had a much smaller discrimination peak at the location of the English category boundary. Since the English [voice] category boundary is consistent with the positive auditory discontinuity, this secondary peak in discrimination performance is likely the result of the presence of the positive auditory discontinuity. Streeter (1976a) found that in Kikuyu, a Bantu language spoken in Kenya, which has an apical [voice] distinction, there is a single discrimination peak at approximately +10 ms VOT. Kikuyu, however, only has a single labial [voice] category, which is produced with pre-voicing. In a labial stop discrimination experiment, adult Kikuyu-speaking listeners showed two discrimination peaks, one in the region of each of the auditory discontinuities.

In the current study, Spanish-speaking listeners did not show two distinct VOT discrimination peaks. However, the Spanish discrimination functions do suggest an effect of the positive auditory discontinuity. Spanish VOT discrimination peaks at approximately 0 ms VOT, but higher than predicted levels of discrimination are maintained throughout the positive region of VOT, especially in the vicinity of the positive auditory discontinuity.

While I do not believe that linguistic experience changes auditory discontinuities themselves, it is important to note that the actual location of these auditory discontinuities is affected by stimulus characteristics. For example, cross-linguistically, the location of the [voice] boundary is stimulus dependent. [Voice] boundaries shift from least to most positive as the place of articulation changes from labial, to alveolar, to velar. Parker (1988) also found that judgments of TOT onset simultaneity were affected by the frequency of the lower component in a manner that parallels the relationship between F1 onset and voicing judgments; as the frequency of the lower component is decreased, the TOT perceptual category boundary increases.

5.3 Proposed General Framework for Phoneme Acquisition

The results of this dissertation are suggestive of a general framework for language acquisition (see Holt et al., 1998; Lotto, 2000 for a similar framework). I propose that the initial perceptual abilities of the newborn infant are the result of general auditory mechanisms, including the type studied in this dissertation. Actual native phoneme categories are then learned through general learning mechanisms. Infants have been shown to be sensitive to statistical properties of the auditory environment. Eight-month-old infants can learn to use syllable transition probabilities to extract nonsense words or tone sequences from a continuous speech/tone stream (Saffran et al., 1996; 1999). Infants of this same age show differential discrimination patterns as a result of the type of distributional information they receive (Maye et al., 2002). Adults have also been shown to be attentive to distributional information when forming non-speech categories,

attending to both stimulus likelihoods and prior probabilities (Sullivan, Lotto, Newlin, and Diehl, 2005).

Phoneme category structure, including perceptual category boundaries and between and within category sensitivities, should depend on both input distributions and general auditory processes. The effects of linguistic experience on perceptual category boundaries suggest that the input distribution history affects phoneme categories. Stimulus range and order effects suggest that more recent distributional information can also affect perceptual category boundaries, although presumably in a more short lived way. A recent study by Guenther et al. (1999) also found that category structure can vary as a result of category and discrimination training. They found that category training on non-speech stimuli that vary in frequency resulted in some decreases in sensitivity to within category tokens. Discrimination training showed the reverse effect, increasing sensitivity to the same stimuli. It is clear from the current TOT category training study that input distributions function *jointly* with auditory discontinuities in determining perceptual TOT category boundaries. In the TOT generalization functions, category labeling boundaries were steepest in the conditions where the 50% cross-over point approached the positive auditory discontinuity. Burnham's (2002) finding that the Thai short-lag/long-lag distinction was more categorically perceived than the long-lead/short-lag distinction also supports this claim that input distributions *and* general auditory sensitivities affect category structure.

Although this dissertation has dealt with auditory discontinuities that are based on judgments of temporal onset asynchronies, there are presumably numerous auditorially natural ways to partition the acoustic space. Some cues may, however, be more naturally

attended to than others. If a particularly salient, or initially attended, auditory discontinuity is not capitalized on during the process of native-language phoneme acquisition then infants must learn to attend to other easily discriminable cues. These newly attended cues should be computationally easy to use and easy to learn. There is evidence to suggest that auditory discontinuities based on temporal onset asynchronies are particularly salient to the young infant (Eimas et al., 1971; Lasky et al., 1975; Streeter, 1976b). Although probably discriminable by young infants, there is less evidence that presence or absence of voicing during the consonant closure is attended to by the young infant. Recall that infants in numerous language environments easily discriminate tokens that fall on either side of the positive (temporal onset asynchrony-based) auditory discontinuity, but none have been found to discriminate tokens that fall on either side of 0 ms VOT (which could be discriminated on the basis of the presence or absence of voicing during the consonant closure) (Eimas et al., 1971; Lasky et al., 1975; Streeter, 1976b). There is, however, a two to one preference for the Spanish-type boundary (i.e., at approximately 0 ms VOT) across the world's languages (Maddieson, 1984). It is possible that more native-language experience is required to shift attention to the relative cue weightings, even if both types of cues are initially discriminable when tested under the right circumstances. However, like in the case of detecting the presence or absence of voicing during the closure, any cue that is not initially attended to, but becomes a salient cue to category membership, should provide some alternate form of computational advantage.

A final prediction of this general framework involves the time course of phoneme category acquisition. If phoneme categories preserve initially salient and attended to

auditory discontinuities, then adult-like phoneme structure should be obtained more quickly than if infants need to learn to attend to alternate cues. Conversely, it should be more difficult to learn a native-language phoneme category if an area of high (and attended to) discriminability falls within the phoneme category. Both the results of the third study and those of Holt et al. (2004) speak directly to this last point.

5.4 A Dissertation in Context: The Generalizability of Findings

Throughout this dissertation I have used the term ‘auditory discontinuities’ to refer specifically to the perceptual nonlinearities that result from a temporal onset asynchrony threshold for judging the simultaneity of two frequency bands or components, in the vicinity of ± 20 ms. We must be careful, however, in generalizing these findings to all phonemic categories. In [voice] categories, especially in languages that contrast short- and long-lag stop consonants, there is a close alignment of phoneme boundaries and auditory discontinuities. While there is some evidence that some form of auditory discontinuity might also underlie many other types of phoneme distinctions, the mechanisms for these are less well studied. Research with human infants (Eimas, 1974) and macaque monkeys (Kuhl and Padden, 1983) have shown that both have better discrimination between place-of-articulation categories than within place-of-articulation categories, as defined by labeling performance of adult English speakers. The English /r/-/l/ distinction is also discriminated in a categorical-like manner by human infants (Eimas, 1975), human adults (Miyawaki et al., 1975), and two species of birds (Dooling, Best, and Brown, 1995), as is a stop consonant/semi-vowel distinction (Eimas and Miller, 1980; Miller and Liberman, 1979), a transition duration distinction (Jusczyk, 1983;

Pisoni, Carrell, and Gans, 1983), and a stop/glide distinction (Cutting and Rosner, 1974; Jusczyk et al., 1977; but see Rosen and Howell, 1987; Macmillan, 1987 for counter arguments).

While the general framework for phoneme acquisition that I presented is rooted in empirical findings, it will ultimately be necessary to have a better understanding of other auditory mechanisms that affect speech and non-speech perception. It will also be necessary to have more complete descriptions of speech sound input distributions.

5.5 A Dissertation in Context: General vs. Language-Specific Mechanisms

The debate over whether language is learned through general or language-specific mechanisms has a history far older than the study of VOT. In the late 1950s, B. F. Skinner published his book *Verbal Behavior* (1957), in which he claimed that language is learned through operant conditioning, much in the same manner that a rat learns to bar press. He further claimed that infants are born as a *tabula rasa*, or blank slate, and everything that infants learn about language is the result of reinforcement.

In his review of *Verbal Behavior*, Chomsky (1959) posited a nativist view of language acquisition, making claims of Universal Grammar and Innate Phonemic Categories. Chomsky claimed that infants are born with innate, language- and species-specific mechanisms that act as a set of parameters that can be activated as a result of language experience. Learning was deemphasized as a method of acquiring language, since the claim was made that the input was too impoverished to provide enough meaningful information to the infant.

The zeitgeist has changed a number of times since the late 1950s, with new theories often denouncing the old. The current dissertation has argued for a general mechanisms approach to phoneme acquisition. General mechanisms approaches have been set out by a number of researchers, including Kuhl (2000b) (see also Holt et al., 1998; Lotto, 2000; Diehl, Lotto, and Holt, 2004). These approaches tend to agree that the process of phoneme acquisition does not require the positing of language-specific mechanisms. This emerging view of language acquisition claims that early perceptual abilities seen in infancy are not unique to speech or language, but are instead the result of general auditory mechanisms. Although learning is also general in nature, it does not come in the form of a Skinnerian conditioning. Instead perceptual learning is the result of infant abilities to attend to and utilize a rich array of information in the auditory environment, in an unsupervised manner.

Where the views presented here (see also Holt et al., 1998; Lotto, 2000) and those of Kuhl's differ is in the domain of phoneme representation. Kuhl believes that there is a warping of the perceptual space as a result of prototypical phoneme exemplars acting as perceptual magnets, drawing towards them other category members (1991). The general framework for phoneme acquisition presented in 5.3 makes no claims of perceptual space warping, instead I suggest that input distributions may be mapped and stored in the form of prior probabilities and stimulus likelihoods. Irrespective of the ultimate form that phoneme representations take, a general mechanisms approach to phoneme and possibly language acquisition has a promising future in our search to understand language and the brain.

5.6 Future Directions

My interest in general auditory and learning mechanisms stems from a desire to understand processes of phoneme acquisition. I, therefore, plan to use the general framework outlined above to study the processes of first and second language acquisition. In particular, I wish to pursue answers to questions such as the following: What factors affect functional equivalence class formation in the language-learning infant? How do auditory nonlinearities affect the rate of phoneme acquisition? How do the characteristics of the input distributions affect phoneme acquisition? Can the presence of auditory discontinuities help second language learners acquire non-native speech sound categories? What causes some cues to be weighted more heavily than others? I would like to address these questions cross-linguistically, in both infant and adult populations, using both speech and non-speech stimuli.

APPENDIX A

Total demographic information from all 3 studies for English-speaking participants

	Current Age	Age of learning L2*	Self-ratings for L2:			
			Conversation	Comprehension	Reading	Writing
Average	20.75	14.17	2.24	2.4	2.85	2.21
Standard Dev.	2.91	2.92	0.79	0.91	1.12	0.95

Total demographic information from all 3 studies for Spanish-speaking participants

	Current Age	Age of learning L2*	Self-ratings for L2:			
			Conversation	Comprehension	Reading	Writing
Average	21.03	12.77	2.51	3.11	3.29	2.93
Standard Dev.	2.62	2.36	1.36	1.43	1.54	1.47

*Includes approximated age. When age was not given, the following ages were used: 10 for Primary school, 13 for Junior high, 16 for High school, 18 for University. When a grade was listed (e.g. 3rd grade), the appropriate age approximation was made.

APPENDIX B

Individual VOT Labeling and Discrimination Functions English-Speaking Listeners, Male

APPENDIX C

Individual VOT Labeling and Discrimination Functions English-Speaking Listeners, Female

APPENDIX D

Individual VOT Labeling and Discrimination Functions Spanish-Speaking Listeners, Male

APPENDIX E

Individual VOT Labeling and Discrimination Functions Spanish-Speaking Listeners, Female

APPENDIX F

Individual TOT Discrimination Functions English-Speaking Listeners, Male (Arrows represent the approximate locations of auditory discontinuities)

APPENDIX G

Individual TOT Discrimination Functions English-Speaking Listeners, Female

(Arrows represent the approximate locations of auditory discontinuities)

APPENDIX H

Individual TOT Discrimination Functions Spanish-Speaking Listeners, Male

(Arrows represent the approximate locations of auditory discontinuities)

APPENDIX I

Individual TOT Discrimination Functions Spanish-Speaking Listeners, Female

(Arrows represent the approximate locations of auditory discontinuities)

REFERENCES

- Abramson, A. S., and Lisker, L. (1970). "Discriminability along the voicing continuum: Cross-language tests," in *Proceedings of the Sixth International Congress of Phonetic Sciences. Prague, 1967*. Prague, Czechoslovakia: Academia.
- Abramson, A. S., and Lisker, L. (1973). "Voice-timing perception in Spanish word-initial Stops," *Journal of Phonetics*, 1, 1-8.
- Aslin, R. N., Pisoni, D. B., Hennessey, B. L., and Perey, A. J. (1981). "Discrimination of voice onset time by human infants: New findings and implications for the effects of early experience," *Child Development*, 52, 1135-1145.
- Balaban, M. T., and Waxman, S. R. (1997). Do words facilitate object categorization in 9-month-old infants?" *Journal of Experimental Child Psychology*, 64, 3-26.
- Bent, T., Bradlow, A. R., and Wright, B. A. (2003). "The influence of linguistic experience on pitch perception in speech and non-speech sounds," Meeting of the Acoustical Society of America, Nashville, TN, April 28- May 2.
- Best, C. T. (1993). "Emergence of language-specific constraints in perception of non-native speech: A window on early phonological development," in *Developmental Neurocognition: Speech and Face Processing in the First Year of Life*, edited by B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. McNeilage, and J. Morton (Kluwer Academic, Norwell, MA), pp. 289-304.
- Best, C. T. (1994). "The emergence of native-language phonological influences in infants: A perceptual assimilation model," in *Development of speech perception: The transition from speech sounds to spoken words*, edited by J. C. Goodman, and H. C. Nusbaum (MIT, Cambridge), pp. 167-224.
- Brady, S. A., and Darwin, C. J. (1978). "A range effect in the perception of voicing," *Journal of the Acoustical Society of America*, 63, 1556-1558.
- Burnham, D. (2002). "Perception of temporal relations as a function of phonology: Speech and non-speech, Thai and English children and adults," in Sarah Hawkins and Noël Nguyen (eds) *TIPS: Temporal Integration in the Perception of Speech*. ISCA, 2002.
- Chomsky, N. (1959). "A review of B. F. Skinner's Verbal Behavior," *Language*, 35(1), 26-58.
- Cutting, J. E., and Rosner, B. S. (1974). "Categories and boundaries in speech and music," *Perception & Psychophysics*, 16, 564-570.

- Diehl, R.L. (1981). "Feature detectors for speech: A critical reappraisal," *Psychological Bulletin*, 89, 1-18.
- Diehl, R. L., Lotto, A. J., and Holt, L. L. (2004). "Speech perception," *Annual Review of Psychology*, 55, 149-179.
- Dooling, R. J., Best, C. T., and Brown, S. D. (1995). "Discrimination of synthetic full-formant and sinewave /ra--la/ continua by budgerigars (*Melopsittacus undulatus*) and zebra finches (*Taeniopygia guttata*)," *Journal of the Acoustical Society of America*, 97, 1835-1846.
- Eimas, P. D. (1974). "Auditory and linguistic processing of cues for place of articulation in infants," *Perception & Psychophysics*, 16, 513-521.
- Eimas, P. D., and Miller, J. L. (1980). "Discrimination of information for manner of articulation by young infants," *Infant Behavior & Development*, 3, 367-375.
- Eimas, P.D., Siqueland, E. R., Jusczyk, P., and Vigorito, J. (1971). "Speech perception in Infants," *Science*, 171, 303-306.
- Green, K.P., Zampini, M.L., and Clarke, C. (1998). "The role of preceding closure interval and voice onset time in the perception of voicing: A comparison of English and Spanish-English bilinguals," *Journal of the Acoustical Society of America*, 104, 1835.
- Guenther, F. H., Husain, F. H., Cohen, M. A., and Shinn-Cunningham, B. G. (1999). "Effects of categorization and discrimination training on auditory perceptual space," *Journal of the Acoustical Society of America*, 106(5), 2900-2912.
- Helson, H. (1964). *Adaptation Level Theory*. (Harper and Row, New York).
- Hillenbrand, J. M., and Gayvert, R. T. (2004). *Open-Source Software for Experimental Design and Control*. <http://homepages.wmich.edu/~hillenbr/>
- Hirsh, I. J., and Sherrick, C. E. (1961). "Perceived order in different sense modalities," *Journal of Experimental Psychology*, 62, 423-432.
- Holt, L. L., Lotto, A. J., and Diehl, R. L. (2004). "Auditory discontinuities interact with categorization: Implications for speech perception," *Journal of the Acoustical Society of America*, 116(3), 1763-1773.
- Holt, L. L., Lotto, A. J., and Kluender, K. R. (1998). "Incorporating principles of general learning in theories of language acquisition," in C.M. Gruber, D. Higgins, K.S. Olson, and T. Wysocki (Eds.), *CLS 34: The Panels*. Chicago: *Chicago Linguistic Society*, 34(2), 253-268.

- Jusczyk, P. W. (1983). "Infants' discrimination of the duration of a rapid spectrum change in nonspeech signals," *Science*, 222, 175-177.
- Jusczyk, P. W., Pisoni, D. B., Walley, A., and Murray, J. (1980). "Discrimination of relative onset time of two-component tones by infants," *Journal of the Acoustical Society of America*, 61(1), 262-270.
- Jusczyk, P. W., Rosner, B. S., Cutting, J. E., Foard, C. F., and Smith, L. B. (1977). "Categorical perception of non-speech sounds by 2-month old infants," *Journal of the Acoustical Society of America*, 67, 262-270.
- Keating, P. (1979). *A phonetic study of a voicing contrast in Polish*. Unpublished Ph.D. dissertation, Brown University.
- Keating, P., Linker, W., and Huffman, M. (1983). "Patterns in allophone distribution for voiced and voiceless stop," *Journal of Phonetics*, 11, 277-290.
- Keating, P. A., Mikoś, M. J., and Ganong, III, W. F. (1981). "A cross-linguistic study of range of voice onset time in the perception of initial stop voicing," *Journal of the Acoustical Society of America*, 70(5), 1261-1271.
- Kessinger, R. H. and Blumstein, S. E. (1997). "Effects of speaking rate on voice-onset-time in Thai, French, and English," *Journal of Phonetics*, 25, 143-168.
- Kingston, J., and Diehl, R. L. (1994). "Phonetic knowledge," *Language*, 30(3), 419-454.
- Klatt, D. H. and Klatt, L. C. (1990). "Analysis, synthesis and perception of voice quality variations among female and male talkers," *Journal of the Acoustical Society of America*, 87(2), 820-857.
- Kuhl, P. K. (1981). "Discrimination of speech by nonhuman animals: Basic auditory sensitivities conducive to the perception of speech-sound categories," *Journal of the Acoustical Society of America*, 70(2), 340-349.
- Kuhl, P. K. (1991). "Human adults and human infants show a 'perceptual magnet effect' for the prototypes of speech categories, monkeys do not," *Perception & Psychophysics*, 50, 93-107.
- Kuhl, P. K. (1993). "Innate predispositions and the effects of experience in speech perception : The Native Language Magnet Theory," in *Developmental Neurocognition: Speech and Face Processing in the First Year of Life*, edited by B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. McNeilage, and J. Morton (Kluwer Academic, Norwell, MA), pp. 259-274.

- Kuhl, P. K. (2000a). "Language, mind, and the brain: Experience alerts perception," in *The New Cognitive Neurosciences*, edited by M. S. Gazzaniga (MIT, Cambridge), pp.99-115.
- Kuhl, P. K. (2000b). "A new view of language," *Proceedings of the National Academy of Science*, 97, 11850-11857.
- Kuhl, P. K., and Miller, J. D. (1975). "Speech perception by chinchillas: 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.
- Kuhl, P. K., and Padden, D. M. (1983). "Enhanced discriminability at the phonetic boundaries for the place feature in macaques," *Journal of the Acoustical Society of America*, 73, 1003-1010.
- Lasky, R. E., Syrdal-Lasky, A., and Klein, R. E. (1975). "VOT discrimination by four to six and a half month old infants from Spanish environments," *Journal of Experimental Child Psychology*, 20, 213-225.
- Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (1967). "Perception of the speech code," *Psychological Review*, 74, 431-461.
- Liberman, A. M., and Matthingly, I. G. (1985). "The motor theory of speech perception revised," *Cognition*, 21, 1-36.
- Lisker, L. and Abramson, A. S. (1964). "A cross-linguistic study of voicing in initial stops: Acoustical measurements," *Word*, 20(3), 385-422.
- Lisker, L., and Abramson, A. S. (1970). "The voicing dimensions: Some experiments in comparative phonetics," in *Proceedings of the Sixth International Congress of Phonetic Sciences. Prague, 1967*. Prague, Czechoslovakia: Academia.
- Lotto, A. J. (2000). "Language acquisition as complex category formation," *Phonetica*, 57, 189-196.
- Macmillan, N. A. (1987). "Beyond the categorical/continuous distinction: A psychophysical approach to processing modes," in *Categorical Perception: The Groundwork of Cognition*, edited by S. Harnad (Cambridge U. P., New York), pp.53-85.
- Maddieson, I. (1984). *Patterns of Sound*. London: Cambridge University Press.

- Maye, J., Werker, J. F., and Gerken, L. (2002). "Infant sensitivity to distributional information can affect phonetic discrimination," *Cognition*, 83(3), B101-B111.
- Miller, G. A., and Nicely, P. E. (1955). "An analysis of perceptual confusions among some English consonants," *Journal of the Acoustical Society of America*, 27(2), 338-352.
- Miller, J. D. (1970). "Audibility curve of the chinchilla," *Journal of the Acoustical Society of America*, 48, 513-523.
- Miller, J. D., Wier, C. C., Pastore, R. E., Kelly, W. J., and Dooling, R. J. (1976). "Discrimination and labeling of a noise-buzz sequence with varying noise-lead times: An example of categorical perception," *Journal of the Acoustical Society of America*, 60(2), 410-417.
- Miller, J. L., Green, K.P., and Reeves, A. (1986). "Speaking rate and segments: A look at the relation between speech production and perception for the voicing contrast," *Phonetica*, 43, 106-115.
- Miller, J. L. and Liberman, A. M. (1979). "Some Effects of Later-Occurring Information on the Perception of Stop Consonant and Semivowel," *Perception & Psychophysics*, 25, 457-465.
- Miller, J. L., and Volaitis, L. E. (1989). "Effect of speaking rate on the perceptual structure of a phonetic category," *Perception & Psychophysics*, 46(6), 505-512.
- Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., and Fujimura, O. (1975). "An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English," *Perception & Psychophysics*, 18, 331-340.
- Pisoni, D. B. (1977). "Identification and discrimination of the relative onset time of two component tones: Implications for voicing perception in stops," *Journal of the Acoustical Society of America*, 61(5), 1352-1361.
- Pisoni, D. B., Aslin, R. N., Perey, A. J., and Hennessy, B. L. (1982). "Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants," *Journal of Experimental Psychology: Human Perception and Performance*, 8, 297-314.
- Pisoni, D. B., Carrell, T. D., and Gans, S. J. (1983). "Perception of the duration of rapid spectrum changes in speech and nonspeech signals," *Perception & Psychophysics*, 34, 314-322.

- Rosen, S., and Howell, P. (1987). "Auditory, articulatory and leaning explanations of categorical perception in speech," in *Categorical Perception: The Groundwork of Cognition*, edited by S. Harnad (Cambridge U. P., New York), pp.113-160.
- Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996). "Statistical learning by 8-month-old infants," *Science*, 274(5294), 1926-1928.
- Saffran, J. R., Newport, E. L., and Aslin, R. N. (1999). "Statistical learning of tone sequences by human infants and adults," *Cognition*, 70(1), 27-52.
- Sharma, A., and Dorman, M. F. (1999). "Cortical auditory evoked potential correlates of categorical perception of voice-onset time," *Journal of the Acoustical Society of America*, 106(2), 1078-1083.
- Simos, P. G., Breier, J. I., Zouridakis, G., and Papanicolaou, A. C. (1998). "Magnetic fields elicited by a tone onset time continuum in humans," *Cognitive Brain Research*, 6, 285-294.
- Simos, P. G., Diehl, R. L., Breier, J. I., Molis, M. R., Zouridakis, G., and Panpanicolaou, A. C. (1998). "MEG correlates of categorical perception of a voice onset time continuum in humans," *Cognitive Brain Research* (7), 215-219.
- Simos, P.G. and Molfese, D.L. (1997). "Electrophysiological response from a temporal order continuum in the newborn infant," *Neuropsychologia*, 35(1), 89-98.
- Sinex, D. G., McDonald, L. P., and Mott, J. B. (1991). "Neural correlates of nonmonotonic temporal acuity for voice onset time," *Journal of the Acoustical Society of America*, (90)5, 2441-2449.
- Skinner, B. F. (1957). *Verbal Behavior*. New York: Appleton-Century-Crofts.
- Steinschneider, M., Schroeder, C. E., Arezzo, J. C., and Vaughan Jr. H. G. (1995). "Physiologic correlates of the voice onset time boundary in primary auditory cortex (A1) of the awake monkey: Temporal response patterns," *Brain and Language*, 48, 326-340.
- Streeter, L. A. (1976a). "Kikuyu labial and apical stop discrimination," *Journal of Phonetics*, 4, 43-49.
- Streeter, L. A. (1976b). "Language perception of two-month-old infants shows effects of both innate mechanisms and experience," *Nature*, 259, 39-41.
- Sullivan, S. C., Lotto, A. J., Newlin, E. T., and Diehl, R. L. (2005) "Sensitivity to stimulus distribution characteristics in auditory categorization," *Journal of the Acoustical Society of America*, 117, 2596.

- Summerfield, Q. (1981). "On articulatory rate and perceptual constancy in phonetic Perception," *Journal of Experimental Psychology: Human Perception and Performance*, 7, 1074-1095.
- Trehub, S. E. (1976). "The discrimination of foreign contrasts by infants and adults," *Child Development*, 47, 466-472.
- Volaitis, L. E., and Miller, J. L. (1992). "Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories," *Journal of the Acoustical Society of America*, 92, 723.
- Waxman, S. R., and Markow, D. B. (1995). "Words as invitations to form categories: Evidence for 12- to 13-month-old infants," *Cognitive Psychology*, 29, 257-302.
- Werker, J. F., and Tees, R. C. (1984). "Cross-language speech perception: Evidence for perceptual reorganization during the first year of life," *Infant Behavior & Development*, 7, 49-63.
- Williams, L. *Speech perception and production as a function of exposure to a second language*. Unpublished PhD dissertation, Harvard University, 1974.
- 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 & Psychophysics*, 21(4), 289-297.

VITA

Jessica Sari Fleming Hay was born in San Francisco, CA on August 8th, 1975, the daughter of Alison S. Fleming and Thomas F. Hay. After completing her work at Northern Secondary School, Toronto, Canada, in 1993, she spent a year traveling and living on a kibbutz in Israel. In 1994 she entered McGill University, Montreal Canada. She received the degree of Bachelor of Science with First Class Honors in psychology in May, 1998. In September, 1998, she entered The Graduate School at the University of Texas at Austin. She earned the degree of Master of Arts in psychology from the University of Texas at Austin in August, 2002. During her continued enrollment in the Graduate School, she was an instructor for Introduction to Psychology in the summer of 2003 and worked as a graduate research assistant in The Speech Perception Lab at the University of Texas at Austin from 1998 to 2005.

Permanent Address: 137 Walmer Road
Toronto, ON
Canada M5R 2X8

Email: hay@psy.texas.edu
jessicahay@gmail.com

This dissertation was typed by the author.