PHONETIC CATEGORY LEARNING

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the Graduate
School of The Ohio State University

by

Grant Leese McGuire, B.A., M.A.

****

The Ohio State University

Dissertation Committee:	Approved by:
Professor Mary Beckman, Co-Adviser	_____
Professor Keith Johnson, Co-Adviser	Co-Adviser Linguistics
Professor Susan Nittrouer	_____
Professor Shari Speer	Co-Adviser Linguistics

Copyright by
Grant Leese McGuire
2007

ABSTRACT

This dissertation discusses the role of phonetic cues and warping in perceptual learning. The experiments described herein use a two dimensional stimulus set derived from the Polish alveopalatal ~ retroflex sibilant contrast with one dimension varying in fricative noise cues and the other formant transition information. In the first set of experiments English listeners demonstrated that they largely assimilate the contrast to their native palato-alveolar fricative, but can detect some differences, primarily using vowel formant transition information. A second set of experiments confirmed that English listeners do not use the full range of cues available to make the distinction in the stimuli. However, another language group, Mandarin listeners, who have a similar contrast natively, do use both fricative noise and vocalic information to categorize these stimuli. Moreover, Mandarin listeners show evidence of cue-integration that English listeners do not.

A final set of experiments used a selective attention training paradigm to explore English listeners' perceptual space for these sounds. It was found that these listeners could indeed use the fricative noise cues for categorization if sufficiently trained and that listeners could be trained to use both cues separately to categorize the stimulus space. In these experiments listeners showed an effect of heightened sensitivity to differences in the trained di-

mension of categorization without any corresponding loss of sensitivity to differences in the other dimension. These results are discussed with respect to theories of perceptual learning and cue acquisition.

ACKNOWLEDGMENTS

I am very much indebted to my committee, Mary Beckman, Keith Johnson, Susan Nittrouer, and Shari Speer. They challenged me and made this project possible; I could not have gotten to this point without their professional guidance. I would also like to thank Alexander Francis for specific feedback on background literature and interpreting experimental results. His help went beyond the call of duty and made my research infinitely better. Thanks also must go to Alejandrina Cristià for helping me with tedious proofreading as well as conceptual review work.

Parts of this work have been presented to audiences at the at the 4th Joint Meeting of the Acoustical Society of America and the Acoustical Society of Japan, the 2007 Annual Meeting of the Linguistic Society of America, Purdue University, and Indiana University. I thank all of those in attendance for their helpful comments.

I would like to thank my teachers at Ohio State for their help in my professional development. I would especially like to thank Brian Joseph and Beth Hume for pushing me to be a creative as well as critical thinker.

Thanks go to my many fellow students who helped me get to this point. I am especially grateful to Mike Armstrong, Andrea Sims, Jeff Mielke, Steve Winters, Na'im Tyson, Angelo Costanzo, Vanessa Metcalf, and David Durian for academic help as well as getting me out of my apartment occasionally, even if it was just to the 'Dube. I am also grateful to the many members of the Berkeley Phonology Laboratory for their input into my work.

I would like to thank the Linguistics Departments at Ohio State and UC Berkeley for help in getting me to this point in my career. Thanks especially to Jane Harper, Francis Crowell, Paula Floro, and Belen Flores for navigating me through the byzantine bureaucracies of both universities.

I thank my family for their support in my graduate school endeavor. Thanks Mom and Dad, Sean, Amy, Ryan, and Mandy. Thanks for making me the human I am. And I have to thank my canine companion Hemingway who put up with the occasional late dinner time or shortened walk, a move to a studio apartment in California, and the odd trip to the kennel while I traveled to present work..

VITA

November 27, 1978.....Born – Davenport, Iowa

2001.....B.A. Linguistics, University of Illinois,
Urbana-Champaign

2001 – 2002.....University Fellow, The Ohio State University

2003 – 2005.....Graduate Teaching and Research Associate,
The Ohio State University

2005.....M.A. Linguistics, The Ohio State University

2005 – 2007.....Graduate Teaching and Research Associate,
University of California, Berkeley

Fields of Study

Major Field: Linguistics

Minor Field: Phonetics

TABLE OF CONTENTS

Abstract....ii

Acknowledgments....iv

Vita....vi

Table of Contents....vii

List of Tables....xi

List of Figures....xii

Chapters:

1. Introduction: Perceptual category learning....1

1.1 Overview....1

1.2 Warping of perceptual space: Perceptual learning mechanisms....2

1.3 Phonetic category acquisition overview....5

1.3.1 Acquisition of native language categories: From differentiation to unitization....6

1.3.2 Older children and perceptual skill: Attentional weighting....9

1.3.3 Warping and low-level information: Stimulus imprinting....13

1.3.4 Discussion....14

1.4 Adult perception and malleability....17

1.4.1 Learning new speech categories....18

1.4.2 Dimensional warping: Perceptual learning mechanisms in adults....25

1.4.3 Summary of adult research....30

1.5 General conclusion....31

2. English listeners' perception of Polish alveopalatal and retroflex voiceless sibilants: A pilot study....32
2.1 Introduction....32
2.2 Polish sibilants....33
2.3 Experiments....35
2.3.1 Stimuli....35
2.3.2 A Polish listener's perception....40
2.3.3 English listeners' perception of the Polish stimuli....43
2.3.4 English orthographic labeling....43
2.3.5 English labeling with brief training....47
2.3.6 Results....49
2.4 General discussion....54
3. Perception of the Polish alveopalatal ~ retroflex sibilant contrast by English and Mandarin listeners....56
3.1 Introduction....56
3.2 Discrimination experiment....58
3.2.1 Materials....59
3.2.2 Subjects....59
3.2.3 Procedure....60
3.2.4 Results and discussion....61
3.3 Labeling experiment....71
3.3.1 Materials....71
3.3.2 Subjects....71

3.3.3 Procedure....72
3.3.4 Results and discussion....73
3.4 General discussion....79
4. English listeners' perceptual learning of the Polish post-alveolar sibilant contrast....82
4.1 Introduction....82
4.2 Experiment 1: Fricative dimension categorizers....86
4.2.1 Materials....86
4.2.2 Subjects....87
4.2.3 Procedure....88
4.2.4 Results and discussion....89
4.3 Experiment 2: Vocalic dimension categorizers....94
4.3.1 Methods....94
4.3.2 Results and discussion....96
4.4 Experiment 3: Two-dimensional categorizers....100
4.4.1 Methods....100
4.4.2 Results and discussion....101
4.5 Experiment 4....107
4.5.1 Methods....108
4.5.2 Results and discussion....108
4.6 Post-hoc analysis....108
4.6.1 Methods....109
4.6.2 Results....109
4.7 General discussion....111

5. Conclusion.....114
5.1 Summary of results and discussion.....114
5.2 Future work.....117
List of References.....119

LIST OF TABLES

Table 2.1: Formant values for each step of the vowel continuum as measured at +25ms and +100ms from the onset of voicing.....38

Table 2.2: Labeling responses from the five subjects. 44

Table 2.3: Response tallies for each subject for the label "sha" and any other label consistently used.....46

LIST OF FIGURES

Figure 2.1: Spectra of fricative step 0 (fully alveopalatal), step 3, step 6, and step 9 (fully retroflex).....37

Figure 2.2: Spectrograms of the fully alveopalatal (top) and fully retroflex (bottom) syllables. 39

Figure 2.3: Stimuli design. Each circle represents a particular combination consonant and vowel from each continuum. Circles with darker outlines represent a subset of tokens for orthographic labeling (see text.).....40

Figure 2.4: Labeling results of the stimuli set by a native speaker. Black indicates for 3/3 trials the stimulus was labeled as alveopalatal; dark gray, 2/3; light gray 1/3; white, 0/3, i.e. labeled 100% as retroflex. 42

Figure 2.5: Pilot training sets; dark filled circles represent tokens used for training, white filled tokens were used for testing (along with the training tokens.) Heavier lines indicate tokens used for orthographic labeling. Clockwise from top: 1) natural category group, 2) fricative categorizers (retroflex transitions), 3) fricative categorizers (alveopalatal transitions).....49

Figure 2.6: Results from the natural category training subjects. Dark shading indicates the stimulus was labeled as alveopalatal >66%; gray shading indicates between 33% and 66% alveopalatal labeling; white indicates <33% alveopalatal labeling. The top left panel is the subject that categorized along the fricative dimension.....51

Figure 2.7: Results of fricative categorizers. Top row shows subjects trained on alveopalatal vocalic tokens and bottom row shows retroflex.....52

Figure 2.8: Times series results from subject 303. Left panel represents block 1/5, central panel shows block 3/5, and right panel shows block 5/5. Black indicates stimulus labeled as alveopalatal, white as retroflex.....53

Figure 3.1: Pairs used in discrimination experiment. Each circle represents one stimulus. Lines connect pairs; single lines connect pairs differing in a single dimension, double lines connect pairs differing in two dimensions.....61

Figure 3.2: Sensitivity by block. "M" and "E" indicate Mandarin and English, respectively...63

Figure 3.3: Mean reaction times for Mandarin (M) and English (E) listeners. Asterisks indicate significant differences in means between language groups.....65

Figure 3.4: Mean reaction times for English listeners (E) and Mandarin listeners (M) by dimension. B = both dimensions differ, f = fricative dimension differs only, v = vocalic dimension differs only. The "*" indicates a significant difference in means ($p < 0.5$).....67

Figure 3.5: Reaction times for discriminating conflicting cue, correlated cue, and mixed conflicting/correlated cue pairs. M= Mandarin listeners, E= English listeners.....69

Figure 3.6: Labeling results for English listeners. White squares indicate > 66% labeling as "sha", light gray squares indicate 33%-66% labeling as "sha", and dark squares indicate < 33% labeling as "sha".....73

Figure 3.7: Labeling results for Mandarin listeners. White squares indicate > 66% labeling as "sha", light gray squares indicate 33%-66% labeling as "sha", and dark squares indicate < 33% labeling as "sha".....74

Figure 3.8: Correlated and conflicting cue definitions for the analysis. Stimuli that are considered correlated are in black and gray; black stimuli were fully correlated.

White squares were considered conflicting for the analysis.....76
Figure 3.9: Comparison of reaction times for conflicting and correlated cue stimuli for Mandarin (M) and English (E) listeners.....77
Figure 4.1: Category assignments for the stimuli in experiment 1. Heavy circles denote discrimination stimuli.....87
Figure 4.2: Cumulative labeling in pre- (left panel) and post-test (right panel) for all learners. Horizontal dimension represents vocalic continuum and vertical dimension represents fricative dimension. Black filled squares indicate > 67% labeling as "A". Gray squares indicate 33%-67% labeling as “A”. White squares represent < 33% labeling as "A".....91
Figure 4.3: Fricative dimension learner's change in sensitivity from pre-test to post-test for the fricative "F" and vocalic "V" dimensions.....92
Figure 4.4: Category assignments of the stimuli for experiment 2, vocalic dimension categorizers. Stimuli marked with heavy circles were used in the discrimination task 95
Figure 4.5: Cumulative labeling in pre- (left panel) and post-test (right panel) for all vocalic dimension learners. Horizontal dimension represents vocalic continuum and vertical dimension represents fricative dimension. Black filled squares indicate > 67% labeling as "A". Gray squares indicate 33%-67% labeling as “A”. White squares represent < 33% labeling as "A". 96
Figure 4.6: Vocalic dimension learner's change in sensitivity from pre-test to post-test for the fricative "F" and vocalic "V" dimensions.....98
Figure 4.7: Vocalic dimension learner's change in sensitivity from pre-test to post-test cross-

category "X" and within category ".".....99
Figure 4.8: Category assignments of the stimuli for experiment 3, two-dimension categorizers.....101
Figure 4.9: Cumulative labeling in pre- (left panel) and post-test (right panel) for all learners. Horizontal dimension represents vocalic continuum and vertical dimension represents fricative dimension. Each square represents one stimulus and each letter indicates that for a particular stimulus > 25% of responses were for the corresponding category.....102
Figure 4.10: Examples of post-test labeling spaces demonstrating a variety of categorizations of the stimulus space. Subject number is indicated in the upper left corner of each panel. See Figure 4.9 for an explanation of this graphic representation. 104
Figure 4.11: Two-dimension learner's change in sensitivity from pre-test to post-test for the fricative "F" and vocalic "V" dimensions.....105
Figure 4.12: Two-dimension learner's change in sensitivity from pre-test to post-test cross-category "X" and within category ".".....106
Figure 4.13: Sensitivity to each region of the stimulus space for the fricative "F" dimension and vocalic "V" dimension for all subjects' discrimination pre-tests.....110

CHAPTER 1

INTRODUCTION: PERCEPTUAL CATEGORY LEARNING

1.1 Overview

One of the central goals of linguistics is an understanding of the extent of humans' phonetic and phonological knowledge. It is generally accepted that this knowledge, especially at higher cognitive levels, is categorical in nature. This categorical knowledge can be modeled in terms of such constructs as features, segments, syllables, etc., and the grammatical (i.e., phonological) rules for combining such meaningless differentiating ingredients into words and larger meaningful units. However, the speech signal itself is not inherently categorical; it is a stream of auditory events that a listener must parse and map onto categories. Thus, a key element in understanding human language is an understanding of categories—their formation, structure, and use. This dissertation attempts to add to our knowledge of human linguistic capabilities by exploring how changes in perceptual space affects category formation and its relevance to our phonetic and phonological knowledge. It will explore differences in how language groups perceive sounds and how these differences can be manipulated.

Specifically, the first chapter will provide a brief overview of perceptual learning mechanisms, research into the acquisition of phonetic categories, and the processes that lead to the phenomena seen in adults. The second chapter will explore a linguistic contrast and a stimulus set used to explore that contrast. The third chapter will compare two different language groups using the stimulus set from Chapter two to explore how different perceptual learning situations (i.e. different language experiences) affect perceptual space. The fourth chapter will then further explore the modification of the perceptual space of one of the linguistic groups. Finally Chapter five will provide a summary and jumping off-point for further research.

1.2 Warping of perceptual space: Perceptual learning mechanisms

It has been shown that perceptual space and acoustic space differ (Miller and Nicely 1955, Stevens 1957, Liberman et al. 1957, Zwicker 1961, Kuhl 1991, Guion 1998, Johnson 2003) and that these differences are often language specific (Terbeek 1977, Harnsberger 2001, Lambacher, et al. 2001, Huang 2001, Mielke 2003). This dissertation is concerned with how such warpings occur and how they can be manipulated.

A classic finding is that many speech sounds are heard categorically, rather than continuously. The first study to report this phenomenon is Liberman, Harris, Hoffman, and Griffith (1957). This study demonstrated that adult listeners of English presented with a continuously varying continuum from voiceless [p] to voiced [b] in the context of [a] hear two distinctly different categories rather than a gradual change. That is differences between categories are perceived as being large while equal acoustic differences within categories are

perceived as being small, if they are perceived at all. Kuhl (1991) further refined this idea by claiming that even within a category stimuli that are peripheral are more easily discriminated than differences closer to a category's center (or prototype).

This phenomena is not language specific, but a general phenomena within the realm of perception, regardless of the mode (Gibson 1991). It is viewed as an aid in identifying categories, such that differences within a category are minimized while those separating categories are maximized resulting in categories being perceived as separate units rather than ends of a continuum (Gibson 1991, Goldstone 1998). A key aspect, therefore, of learning a category is learning to maximize differences while minimizing similarities. However, these are just two phenomena involved in perceptual learning. Goldstone (1998) identifies four mechanisms relevant to perceptual learning: attentional weighting, stimulus imprinting, differentiation (acquired distinctiveness), and unitization (acquired equivalence).

Attentional weighting refers to the phenomenon of paying more attention to relevant dimensions of categorization and less attention to dimensions that are irrelevant or less informative. In speech, this means directing attention at relevant cues while reducing attention to or ignoring cues that are uninformative or even contradictory to a categorization at hand. For example, Japanese listeners learning to distinguish /r/ from /l/ must learn to attend to relative changes in F3 (Bradlow et al. 1999). Similarly, Francis and Nusbaum (2002) showed that English listeners learning the Korean stop contrast spontaneously changed which cues they attended to to better perform at the task at hand. Such a change in weighting is found in children's use of cues, also (e.g. Nittrouer and Miller 1997).

A second phenomenon, stimulus imprinting, means that specific receptors are specialized for stimuli, parts of stimuli, or whole categories. This can be an internalized trace memory, as found in some exemplar models (e.g. Logan 1988, Johnson 1997) or even general theories of implicit memory (Schacter 1987). For example, Goldinger (1996) found that word identification is facilitated when a previously heard word is repeated by the same voice it was originally heard in; that is the word and voice seem to be stored as a unit.

Differentiation, which includes acquired distinctiveness (discussed later), is the commonly observed phenomenon whereby stimuli or dimensions that were once difficult to distinguish become more perceptually distinct or separate. In language this phenomenon has been reported in many experiments where subjects showed improved discrimination of some difficult contrast (e.g. Strange and Dittman 1981, Jamieson and Morosan 1986). As will be discussed later, some authors feel this is a key phenomenon early in perceptual learning.

Unitization, including the phenomenon known as acquired equivalence, is the loss of discriminability or sensitivity to some contrast; that is treating distinct units as a whole. This can be viewed as learning to ignore irrelevant distinctions for a given categorization task. In speech perception, the Perceptual Magnet Effect (PME; Kuhl 1991) is an example of unitization. Under this theory, stimuli closer to a category prototype become less perceptually distinct resulting in a general shrinking of perceptual space around a category center. Similarly, different dimensions or cues to a contrast can be integrated together such that they are perceived as a unit rather than distinct, separate parts (Grau and Kemler Nelson 1988, Czerwinski et al. 1992).

The following sections of this chapter will explore how the phenomena briefly described above operate in first language acquisition as well as in adulthood. This will act as a base of knowledge for understanding the extent to which and how such phenomena can be manipulated. Subsequent chapters will explore how different populations of listeners with different linguistic experiences (i.e. different language groups) perceive the same sounds differently and how these differences can be manipulated.

1.3 Phonetic category acquisition overview

Adults demonstrate a diverse ability to perceive speech sounds, from clearly discriminating fine distinctions, even among different tokens of the same word, to broad generalizations collapsing across acoustically distinct tokens. The traditional view of the development of human consonant and vowel perception is that infants are born with a universal ability to discriminate phonetic distinctions and that as they age this ability narrows to only relevant native contrasts. This account is disproved by evidence of the adult ability to discriminate difficult non-native contrasts (e.g. Werker and Tees 1984b), as well as studies showing infant's apparent loss and subsequent regaining of sensitivity to fine phonetic detail during initial word acquisition (Werker, et al. 2002). However, the account can be refined to accommodate such seemingly contradictory data by examining the role of lexical acquisition in the processing of speech. This section will begin by sketching out the relevant phenomena and then attempting to show how the results are part of a broader picture where children acquiring language must learn to incorporate low-level sublexical information along with higher-level lexi-

cal information to process speech in a truly adult-like manner. That is, the development of speech category perception seems to proceed in steps from the development of the basic perceptual system to maturation of the strategies employed by that system.

1.3.1 Acquisition of native language categories: From differentiation to unitization

In an important early study, Werker and Tees (1984a) demonstrated that upon birth, infants show a tremendous ability to discriminate phonetic distinctions that are not part of their native language, but this ability eventually disappears near the end of the first year of life. This seminal study compared the performance of several different groups on native and non-native subjects. Their first experiment, intended to replicate earlier results, compared the performance of 6-10 month old English learning infants, Thompson Salish speaking adults, and English speaking adults on a task that required them to differentiate English /ba/ ~ /da/ and Thompson Salish /k'i/ ~ /q'i/ syllables. Infants were tested using the conditioned head turn (HT) paradigm and adults pressed a button when they heard a change in stimulus category. All of the Thompson speaking adults reached criterion, while 80% of the English learning infants and only 30% of English speaking adults reached criterion, demonstrating the infants' expertise at perceiving non-native contrasts.

Werker and Tees expand on this result in a subsequent experiment where English learning infants 8-10 months and 10-12 months were tested on the Thompson velar versus uvular contrast as well as the Hindi retroflex versus dental /ʈa/ ~ /ta/ contrasts. A third experiment was similar except that is was a longitudinal study. The results from both studies show that infants have a significantly reduced ability to perceive the contrasts tested starting

around 8-10 months and that 10-12 month old infants perform as poorly as or worse than English speaking adults. Follow-up results with 11-12 month old Hindi learning infants and Salish learning infants showed that each group performed at or near criterion with their respective native contrast.

Overall, this study gives the basic evidence and time line for the onset of native-language specific perception for consonant place contrasts. Starting around 8 months infants show a decrease in perceptual attunement to non-native contrasts that reaches roughly adult-like status by 12 months; however, native contrasts show no such loss in discriminability. Later work has refined these results and looked at the phenomenon in more detail, in some cases finding contradictory results.

In a 1992 paper, Kuhl, Williams, Lacerda, Stevens, and Lindblom explored the onset of native language specific perception using vowels. For evidence of language specific perception the authors looked for the Perceptual Magnet Effect (PME; Kuhl 1991) in infants. The PME is a phenomenon where perceptual space is warped. Specifically equally spaced sounds (in an auditory space, e.g. Bark) within a category are warped such that tokens closer to a category prototype are more difficult to discriminate and are therefore closer in perceptual space than tokens further from a category prototype which are easier to discriminate from each other and therefore farther apart in perceptual space. In their study, English and Swedish infants 6 months of age were tested on synthetic continua centered around English /i/ and Swedish /y/ using the HT paradigm. The results demonstrated that English and

Swedish infants both showed a PME for tokens in their native language, but not the non-native tokens. This result pushes the boundaries for the onset of native-language perception forward to 6 months for vowels.

Further evidence for an early PME for vowels is provided by Polka and Werker (1994). In this study, experiments were conducted with English learning 4 month olds, 6-8 month olds, and 10-12 month olds on German stimuli contrasting tense and lax front rounded vowels with back rounded vowels. It was found that 4 month old infants are significantly better at discriminating the non-native contrasts than 6 month old infants and that 6 month old infants perform better than the 10-12 month infants, as well. Further, the 10-12 month old infants did not perform as well as English and German speaking adults (who performed similarly.)

A study by Maye, Werker, and Gerken (2002) demonstrated one way to create such categorical behavior. They exposed young infants (6 and 8 months) to a synthetic continuum from [da] to [ta] in differing distributions. One group for each age group heard a unimodal distribution of stimuli having a peak in the center of the continuum and the other group heard a bimodal distribution centered near the endpoints. This resulted in only the bimodal group showing an ability to discriminate the endpoints of the continuum. This shows that infants have a strong ability to process statistical information and that different distributions of stimuli can lead implicitly to unitization and differentiation resulting in categorical perception.

Together, these results indicate that language specific vowel perception strategies emerge earlier than those for consonants, that the development is gradual, and importantly, that adults perform better than older infants in non-native discrimination. Moreover, the distribution of stimuli in an experiment can drive perceptual learning.

The studies above all indicate that between six months and one year infants begin unitization for sounds such that they become largely insensitive non-native contrasts. That older children and adults treat contrasts used in other languages as the same category in certain tasks is particularly strong evidence for language specific perception. The following section explores in more detail the fine-tuning of the perceptual system that results in a fully adult system.

1.3.2 Older children and perceptual skill: Attentional weighting

Adults are able to easily recognize and distinguish words; however, infants and young children do not demonstrate adult-like skill either in recognizing words or in using perceptual cues. For example, when the speech signal is degraded children show much more difficulty in recognition than adults.

Eisenberg, Shannon, Schaefer Martinez, Wygonski, and Boothroyd (1999) processed several types of stimuli, including sentences, words, nonce syllables, and digits (for recall), through 4-, 6-, 8-, 16-, and 32-band noise filters to remove spectral cues and used these processed stimuli in recognition tasks. Generally, adults and 10-12 year olds performed the same on all tasks while 5-7 year olds performed significantly worse. A further exploration of the sentence recognition task found that the young children did not use contextual information

to the same extent that the older children and adults did. Similarly, even with considerably less degraded speech (16- and 32-band), the young children did not achieve the level of accuracy that adults and older children did for most tasks.

Another study, by Ewards, Fox, and Rogers (2002), provides a similar picture. In this study, children 3-4yrs, 5-6yrs, 7-8yrs old, and adults1 discriminated a *tap/tack* and *cap/cat* contrast where the final portion of the coda was successively gated, that is the coda information necessary for differentiating the minimal pair was gradually removed. All groups had more difficulty with the contrasts in the gated conditions than in conditions with full information. However, the differences between age groups was generally gradient such that the youngest children performing the poorest and the adults performed the best. Interestingly, in the whole-word condition, the youngest children performed significantly worse than the other children, who were all at or near adult performance. Further analysis found a strong positive correlation between performance and vocabulary size, a result that will be returned to later.

These findings suggest that young children are still acquiring the full range of skills needed to perceive speech in an adult-like manner. Part of the reason for this can be elucidated by looking at how children use cues to detect phonetic categories and how they selectively attend to the speech signal.

Nittrouer and colleagues have done considerable work examining children's perception of place cues in fricatives. Generally, this work uses synthetic or modified natural stimuli arranged into continua where either transition or fricative noise cues are minimized or both are varied independently. Nittrouer and Miller (1997) and Nittrouer (2002) found that young

English learning children (3.5 years old) generally weight transition cues heavily when discriminating place in native sibilant fricatives (/s/ and /ʃ/), gradually reaching adult-like weighting (fricative noise more than transition) around 7-8 years old. Nittrouer (2002) demonstrated that for the /f/ ~ /θ/ distinction children predictably used cues like adults, who weight formant transition much more heavily than fricative noise. Nittrouer's general conclusion, the Developmental Weighting Shift hypothesis, is that children initially give more attention to large scale changes in the acoustic signal as driven by major changes in articulators, only later using the more finely detailed fricative noise information. Importantly, this work shows that as children age and acquire experience, they change their weighting of acoustic information available in the signal. This process apparently takes many years and amounts to a gradual refinement of the perceptual system.

Similarly, Hazan and Barrett (2000) examined children 6-12 years old and their ability to perceive several phonemic contrasts, with and without all available cues. The contrasts ranged from robust to fragile as based on frequency in the world's languages, respectively /k/~/ɡ/, /d/~/ɡ/, /s/~/z/, /s/~/ʃ/, and were represented by minimal pairs. All contrasts (except /s/ ~ /z/) were tested in such a way that static and dynamic cues were in isolation as well as in their natural combinations. Their results show a general increase in categorization accuracy, but 12 year olds were still not quite at adult-like levels. Generally, the children showed more inconsistency in isolated cue conditions compared to combined cue conditions; adults, however, seem to be easily able to switch to different perceptual strategies.

In a study from 1999, Walley and Flege examined children's and adult's category boundaries. In this study English learning children (five and nine years old) and adult native speakers of English labeled two continua, a native one from /ɪ/ to /i/ and mixed native /ɪ/ to non-native /y/. For both continua, adults had a more categorical, steeper slope identification function, indicative of stronger, more precise categories. Moreover, older children and adults label fewer stimuli as the native category /ɪ/ (as opposed to non-native /y/) than the younger children. This is taken as evidence demonstrating that children's categories are more inclusive of variation than adults' and that one aspect of learning is the range of variation allowed for a given category.

Taken together, the results from these various studies show that younger children are not as sophisticated in their use of redundant information as adults and that their category boundaries are somewhat fuzzy; i.e. they are not as consistent as adults in categorizing ambiguous stimuli and their categories lack the full effects of acquired equivalence and distinctiveness. Adults seem to be able to use a wide variety of cues and perceptual strategies, both in conjunction and independently, while children are much more limited. Specifically, Hazan and Barrett's work suggests that children discriminate phonemically in a much more holistic manner than adults. Similarly, Nittrouer and colleagues' results indicate that children's attention to information available in the speech signal gradually changes as a result of experience (Nittrouer 2006).

1.3.3 Warping and low-level information: Stimulus imprinting

The work discussed in the previous sections demonstrates that infants and children acquiring their first language tune their perceptual systems in various ways as new categories are formed and cues to identification are discovered. However, as has been shown in adults, knowledge of fine-grained phonetic detail is not lost. Recent work has shown that implicit memory tasks reveal a wide range of detailed knowledge in adults (Schacter 1987, Schacter and Church 1992). Fisher and Church (2000) present data from two experiments demonstrating that 2 - 3 year old children and adults both repeat words more accurately if they had previously heard that word in the experiment and conclude that this offers a mechanism and window into the effects of memory on lexical storage.

In another study, Fisher, Hunt, Chambers, and Church (2001) examine in more detail the role of implicit memory in children using novel words. In an initial experiment, they replicated the results of the previously mentioned study, using CVC nonwords. Here 2.5 year olds repeated studied novel words more accurately than new novel words. In the second, third, and fourth experiments CVCCVC non-words were used for familiarization where one syllable had high frequency phonotactic structure and the other low. In experiment two, the syllables were interchanged (and rerecorded) in the test phase such that the children heard study syllables in a new context (either preceding or following a unique syllable). The results of this experiment showed more accuracy for studied syllables in a new context, indicative of some degree of abstraction. In experiment three no new syllables were presented to the children in the test phase, but half of the studied bisyllables were interchanged as in experiment two. Here there was more accuracy for the syllables in the same context as the

study phase as well as more accuracy for the higher frequency syllables. This indicates that the children are encoding very specific representations of the non-words as well as making generalizations across the lexicon (because these novel words show frequency effects.) The final experiment was identical to the third, except that the test phase stimuli consisted of only the second syllable where the initial syllable had been spliced out. The result is that half of the syllables are the exact same tokens as the study phase, but with only the second syllable present; and the other half had contextual information that was different from the study phase. As with the previous experiment, the same context tokens were repeated more accurately than the changed context syllables. Together, these experiments display an ability in children, like adults, to rapidly learn new tokens with context-specific detail, as well as the ability to abstract away from that detail (i.e. the "different context" effects).

Overall, the evidence from implicit memory research provides strong evidence that children are able to rapidly learn token-specific details. This behavior is extremely similar to that seen in adults and is essentially the "stimulus imprinting" perceptual learning mechanism described by Goldstone (1998). Such knowledge forms a basis of knowledge of the distribution and variety of stimuli in perceptual space allowing more developed categories and aiding in identification.

1.3.4 Discussion

The literature discussed so far points to several phenomena in the development of human speech perception. Infants initially show a strong ability to discriminate phonetic categories. This ability seems to decline through the first year of life as language specific abili-

ties take precedence, driven by exposure solely to the infant's native language. However, it takes children many years to achieve fully adult-like perceptual abilities. Specifically they do not seem to use all of the cues available for recognition and weight cues differently than adults do. As suggested by Beckman and Edwards (2000) and Beckman, Munsen, and Edwards (to appear), the onset of word learning greatly affects speech perception and plays a role in how well children are able to use fine phonetic detail for word recognition and categorization.

Being able to discriminate phonemes means being able to discriminate categories that are based in higher levels of abstraction, i.e. meaning and contrast. In order to make such abstractions a large database of knowledge is required (the lexicon.) What seems to happen initially is that infants in the early stages of experience with language can rely on raw auditory processing, but native-language distributions eventually tune the perceptual system into categories based on those distributions (e.g. Maye et al. 2002). As the sound - meaning relationship becomes more important higher cognitive loads (i.e. processing words with complex concepts rather than raw syllables) and lexical access (neighborhood competition) begin to affect the processing of fine phonetic detail, require the steady accumulation of tokens in a lexicon and an understanding of what is necessary to phonemically contrast words.

At this point it is useful to take a brief look at non-linguistic categorization phenomena for some support of these ideas. Mareschal, Powell, and Volein (2003) examined 7 and 9 month olds' categorization of dogs and cats seeking to learn if children first formed coarse-grained categories that were later refined as fine detail was added or if categories grow out

of fine-grained detail. The authors argue, and find support for the view that infants build categories from the ground up by analyzing features (fine-grained detail.) The categorization of cats and dogs is interesting because the features that distinguish them are not all symmetrical. Specifically, cat features are generally subsumed by dog features, i.e. "for most features, cats are plausible dogs whereas dogs are not plausible cats" (103). Based on this, initial familiarization with a variety of cats should cause infants to reject dogs as members of that category, while familiarization with dogs should cause infants to accept cats as members of the "dog" category. In fact, this is the result that they found using small toy dogs and cats (and an eagle as a control). They conclude that unlike adults, the infants are processing the information bottom-up by analyzing individual features which leads to the asymmetry in categorization. This conclusion is further supported by a follow-up study, French, Mareschal, Mermillod, and Quinn (2004) who found that the results could be reversed or negated by using subsets of the stimuli to change the distribution of features.

With respect to language, we can view infants' earliest experiences with language as representative of this bottom-up processing using fine phonetic detail. At later stages in development higher-level concepts (lexical entries) begin asserting themselves and children must learn to incorporate both top-down and bottom-up information to achieve the full range of perceptual skills seen in adults. Ultimately, the development of consonant and vowel categories requires an infant not only to be sensitive to variation in minor details as well as variation across broad categories but to be able to incorporate both types of knowledge in a holistic manner.

The basic idea of perceptual development that children begin with language-universal perception which is gradually refined into language-specific perception can now be seen as process where information is gradually being layered allowing larger generalizations to be made. However, the acquisition of phonemes, as in that both *cat* and *cap* mean something and are two contrasting words that are phonetically distinct having separate phonemes and several cues to those phonemes, requires learning the concepts and their relations to each other and integrating that information with the distributional information that [kæ], [æt], and [æp] are distinct sounds. Organizing and integrating such a complex array of information requires many years of development.

1.4 Adult perception and malleability

Despite, or possibly because of, the accumulated knowledge that results in robust, stable perceptual categories, adult perceptual abilities are still quite malleable and new contrasts can be learned with differing degrees of success (Pisoni et al. 1982, Strange and Dittman 1984, Jamieson and Morosan, 1986, Logan et al. 1991, Case et al. 2003, Guion and Pederson 2007). Phonetic training studies, as well as research into perceptual learning in general, have shown that the phenomena demonstrated above in children (dimensional weighting, stimulus imprinting, differentiation, unitization) are also operant in adults learning new speech categories. The following sections detail some of this research and current issues in perceptual learning.

1.4.1 Learning new speech categories

In general, attempts at training sound categories have met with mixed results, especially when attempting to generalize from specific stimulus knowledge to broader, more phonemic knowledge. Most of this research has focused on the ability of learners to discriminate two categories, especially Japanese speakers learning /r/ and /l/.

An important early study is Strange and Dittman (1984). This study attempted to train Japanese listeners to distinguish the English /r/~/l/ contrast. This study used fixed AX discrimination with a 1s ISI using a synthetic *rock – lock* continuum. Testing was done using naturally produced minimal pairs having the target sound in various contexts in an identification task and an oddity discrimination task using the training series and a similar *rake – lake* series. Clear improvement occurred throughout the training and most subjects showed moderate improvement from pre- to post-test on the synthetic stimuli. However, transfer to the natural stimuli was poor. These results indicate that the training was too narrow in focus with little variety. Further, as shown by Guenther, Husain, Cohen, and Shinn-Cunningham (1999), discrimination training has the effect of increasing sensitivity to stimuli and does not help category consolidation.

Logan, Lively, and Pisoni (LL&P, 1991) addressed some of the shortcomings of the Strange and Dittman (1984) work by increasing the variety and duration of training . The authors reasoned that training in a variety of contexts (word initial, word final, in clusters, etc.) using a variety of stimuli (natural tokens from multiple talkers) should improve performance. Further, they used identification training using stimuli from minimal pairs to encourage classification into categories rather than discrimination training. These modifications re-

sulted in a general improvement from pre-test to post-test within most contexts. The improvement was small, but consistent across listeners and supported by percent correct and reaction time data from the training sessions. A follow up test using a novel talker and novel stimuli showed some generalization to the new stimuli, and this generalization was more apparent in novel stimuli produced by a talker from the training stimuli.

A second study by Lively, Logan, and Pisoni (1993) further explored the role of variability. This study consisted of two experiments where the first was largely identical to the earlier study except training was concentrated on the more difficult syllabic contexts and the second experiment used stimuli produced by single talker. Both experiments resulted in general improvement in testing. However, the subjects trained in talker-based variation showed more robust performance in a generalization task using new tokens and a new talker than subjects in the single talker condition.

A third study, this time by Lively, Pisoni, Yamada, Tohkura, and Yamada (1994), sought to examine the long-term effects of the training paradigm they had used previously. Using essentially the same stimuli and training as Logan et al. (1991), subjects were trained to differentiate minimal pairs of words with /r/ and /l/. However, in this case the subjects were retested three months after training and again six months later. Like their earlier studies, subjects showed consistent improvement throughout the training and that generalization to new tokens was best when the stimuli were presented in the voice of a previously heard talker (general improvement of 11% over pretest). Three months after training, subjects showed only a small decrease in accuracy (still 9% above pretest). Six months later, subjects

still showed some improvement over pretest levels for most contrasts as well as an old talker advantage for new stimuli (4.5% above pretest). These results demonstrate that training effects can persist for quite some time despite lack of further experience with the stimuli or contrast.

These results were further studied by Callan, Tajima, Callan, Kubo, Masaki, and Akahane-Yamada (2003). They tested the neural aspects of learning difficult and easy contrasts by examining Japanese speakers trained on the on the /r/-/l/ distinction (Bradlow et al. 1997) using fMRI. Subjects responded to two other contrasts that were not trained, /b/-/v/ (difficult for Japanese listeners), and /b/-/g/ (easy for Japanese listeners). The imaging data showed that activity for the trained contrast is not the same as that of the easy contrast and indicated that learning a difficult non-native contrast involved close associations with articulatory mappings. This also supports the findings of Bradlow Akahane-Yamada, Pisoni, and Tohkura (1999) who found that the Logan et al. (1991) perceptual training paradigm also improved production of the contrast and had long-lasting effects.

Taken together, the results of these studies with the /r/ - /l/ distinction demonstrate that perceiving a non-native phonemic difference between minimal pairs requires a tremendous amount of training in a variety of contexts. However, effective training of this distinction can improve subjects' perception of the contrast and such training has long-lasting effects, although the results still do not quite correspond to learned easy contrasts nor do subjects achieve native-like perception.

Such difficulties are supported by anecdotal and experimental evidence from L2 learners in a natural environment. Specifically, Pallier, Bosch, and Sebastián-Gallés (1997) examined an adult population of extremely proficient bilinguals who learned their second language before the age of six, Spanish speakers in Catalonia. Catalan has two distinct vowel categories /e/ and /ɛ/ where Spanish has only one /e/. Subjects with Spanish speaking or Catalan speaking parents labeled, discriminated, and judged a continuum of stimuli from [e] to [ɛ]. For labeling and discrimination, subjects with Catalan speaking parents showed classic categorical perception of two categories; those with Spanish speaking parents treated the stimuli as a single category. The goodness judgments show even more striking results. Subjects were asked to rate the continuum based on how good an exemplar of Catalan /e/, Catalan /ɛ/, and Spanish /e/ each stimulus was. The result show that Catalans and Spanish treat their native vowels as expected. Interestingly, the Spanish dominant group, when judging the stimuli as Catalan speakers, showed similar, but not the same results as the Catalan dominant group. This seems to indicate the Spanish dominants are aware of the two categories and have some knowledge of them, but not in the same way that Catalan dominants are. It would seem from such data that perceptual categories are very rigid and unless acquired early and used almost exclusively, native speaker perception is impossible to achieve.

In general, such results imply that experimentally modifying perceptual categories is a long and difficult process. Further, it seems likely that it may be impossible for training (or even considerable linguistic experience) to result in native-like speech perception. However, there are some results indicating that training other contrasts can be accomplished in much shorter time and with considerably greater ease.

In a study from 1982, Pisoni, Aslin, Perey, Hennessy examined English listeners perception of VOT. Generally, English speakers categorize negative and small positive VOT values into one category and longer lag times into another. In their first experiment, they asked subjects to label a VOT continuum using either two categories [ba] and [pa] or three as [b], [p], [p^h] after receiving brief listening experience to each category. Interestingly, subjects that were given three labels to assign to the continuum were able to fairly consistently label the continuum into three categories (although the negative VOT category had a shallow slope identification function.) Further, groups receiving discrimination testing showed a slight increase in discriminability in the negative VOT portion of the continuum, whether or not they had previously used two or three labels for the continuum. This latter result was confirmed in a second experiment. A third experiment used ABX discrimination training followed by identification testing over the course of four days (one hour each day). After just two sessions subjects showed steep identification functions demonstrating categorical perception of three contrasts in the VOT continuum. Contrary to previous studies, this shows an ability to overcome the ingrained native categories with brief training.

However, it is important to consider the differences between English speakers learning VOT categories and Japanese speakers learning the /r/ ~ /l/ distinction. First, as pointed out by Logan and Pruitt (1995), VOT is a temporal contrast that may be easier to learn that the spectral contrast in liquid discrimination, that English allows considerable variation in VOT, much of which is allophonic (i.e. pre-voicing in VCV contexts), and the generalization of the results to novel stimuli was not tested. The last two points are especially important. English speakers have considerable experience with different lead and lag times in VOT perception and it is highly likely that allophonic and sub-allophonic differences are used in word recognition and perception generally. The Pisoni et al. (1982) experiment may have directed subjects' attention to these distinctions. Further, it is not known whether subjects can generalize this distinction to more difficult tasks, especially word discrimination. As demonstrated by Strange and Dittman (1984) in comparison with Logan et al. (1991) the generalization of low-level phonetic distinctions to higher-level distinctions requires considerably more training using high-variability stimuli as well as a rethinking of the tasks necessary to produce the generalizations.

Similarly, Jamieson and Morosan (J&M, 1986) studied French speakers acquisition of a voicing contrast in interdentals. This study used synthetic and natural CV stimuli, the natural stimuli having wide variation in several dimensions and the synthetic stimuli having variation only in fricative duration. The synthetic stimuli were used in identification training over the course of three days (total of 90 minutes) using a fading technique where the most extreme tokens were identified initially; later training used more difficult stimuli. Background

noise from a cafeteria was added in on the last day. Identification and discrimination (AX, 850ms ISI) tests using both natural and synthetic stimuli were administered as pre- and post-tests. In general, the subjects showed significant improvement in discrimination of synthesized stimuli across categories but not within categories, while a control group showed no such behavior. Importantly, the subjects significantly improved in the identification of most natural tokens. A later study (Jamieson and Morosan 1989) found similar results with only 40 minutes of training and slightly less robust results using less variable (prototype) training.

The results of these studies seem to be in contrast with those of the /r/-/l/ literature. Here, a relatively small amount of training with highly specific variation resulted in generalization to natural token variation. Based on the /r/-/l/ literature, one would expect that this somewhat limited training would not result in robust transfer to natural stimuli. One major question is whether the switch to labeling training by LL&P is the primary cause of the difference between their results and S&D's. Like LL&P, J&M used labeling training and had significant success. So perhaps the reason why S&D could not find generalization to natural stimuli is primarily their use of discrimination training. This is somewhat contradicted by the results of Pisoni et al. (1982) where discrimination training in fact resulted in categorization (although the issue of temporal vs. spectral contrast is still present.) Another possible factor is the lexical status of the stimuli. All of the stimuli in LL&P and S&D were words with clear lexical status (chosen from dictionaries). However, the J&M stimuli were solely the syllables [θʌ] and [ðʌ], the first is a non-word and the second is a function word, *the*. This would seem to make the task more of a non-word, stimuli specific task that does

not involve any lexical competition (i.e. not word level processing.) Thus, this task is more akin to that of Pisoni et al. (1982). Further, the two contrasts may have different inherent difficulties. The /r/-/l/ distinction could be more perceptually difficult than /θ/-/ð/ distinction in terms of general acoustic saliency. However, beyond that, the two contrasts could have different status for the listeners in the tasks based on cues that differentiate the contrast. In this case that would mean that Japanese speakers do not have as much experience with F3 as a cue to a phonetic difference and have difficulty using it for discrimination while French speakers have experience with the cues to voicing in fricatives (e.g. [f] ~ [v]) and can easily extend those cues to the new contrast. Of course, it is most likely that some combination of these explanations produce the results in these studies.

1.4.2 Dimensional warping: Perceptual learning mechanisms in adults

Recent training studies have begun to explore the changes that happen to the perception of the categories themselves as they are learned. Much of this work was spurred by the Perceptual Magnet Effect (PME, Kuhl 1991). This effect is somewhat controversial; in fact Lotto, Kluender, and Holt (1998) argue that based on their results that the PME is essentially the same as categorical perception and is an epiphenomenal result of certain methodologies. Whether or not the PME actually exists as a unique phenomenon, there is clear evidence of warping of a categories' space such that tokens within a category are more difficult to discriminate than those across categories, as in classical categorical perception (Liberman et al. 1957). In other perceptual domains it is well known that forming a category requires not

only increased perceptual distance between categories, but also increased similarity within the category itself (e.g. Gibson 1991, Goldstone 1998). This section discusses some key studies looking at changes in the architecture of a given category that results from training.

Goldstone (1994) examined perceptual learning using categories that varied on two dimensions. Goldstone was especially interested in how dimensions are related together and whether they acquire distinctiveness across boundaries through training or equivalence within category. Although this research is not done with speech stimuli, it has relevance to perceptual learning in general and especially the acquired equivalence versus acquired distinctiveness of categorization. Much of speech literature traditionally assumes acquired equivalence, especially the literature on the PME and much of that concerning categorical perception. Much of this assumption is based in the idea that children initially can discriminate a wide range of sounds and exposure to their native language tunes the perceptual system to native contrasts. However, as discussed in the earlier question, the situation is considerably more complicated.

In order to begin answering such questions, Goldstone (1994) trained subjects using stimuli where the two dimensions are integrated and are difficult to separate perceptually, as well as separable dimensions where subjects can easily attend to one dimension and ignore the other. Two sets of visual stimuli were used, 16 stimulus items in each set. The first set of stimuli consisted of squares that varied in their size and brightness, two separable dimensions. The second set, the integral dimensions, varied on brightness and saturation. Four groups of subjects were assigned to each set of stimuli: a control group, a size learning

group, a brightness or saturation learning group, and a combined dimension learning group. The single dimension learning groups each were trained to categorize based on the relevant dimension and ignore the variation in the other dimension. The combined dimension group learned four categories, essentially learning to differentiate along both dimensions. This group allows an exploration of how the two dimensions compete with each other, if at all. Primarily Goldstone found strong evidence for acquired distinctiveness. All three dimensions exhibited large increases in *d'* values for stimuli on the boundaries between categories. Interestingly, d' values also increased within categories for the category relevant dimension. Evidence was generally lacking for acquired equivalence, except for a single category-irrelevant distinction (size for brightness categorizers). There was also some evidence for competition between dimensions, especially in the separable dimension stimuli.

These results indicate that the process of category learning results in general sensitization to the relevant dimension of categorization and that discriminability increases across category boundaries, but within-category discrimination does not decrease significantly and can actually increase.

Somewhat contradictory evidence is provided by another study. Livingston, Andrews, and Harnad (1998) examined perceptual learning of single-dimension and multiple dimension stimuli. The experimenters found significant within category equivalence for learning multiple dimension stimuli and no learning effect for single dimension stimuli. In one experiment the stimuli consisted of images roughly based on single-celled organisms. The two categories for these stimuli were easily discriminated and no features overlapped. In

a second experiment similar stimuli were used, but these stimuli had overlapping features such that some examples of each category were ambiguous. In this case similar results were found, but the effect was much smaller. In a final experiment subjects categorized actual, rather than abstract objects. These were examples of black and white drawings of day-old chick genitals[1] and subjects were asked to categorize them with accuracy feedback until a pre-determined criterion was reached; this was followed by a similarity judgment task. Using a multi-dimensional scaling analysis of the rating data, only within-category compression was found when compared to a control group, consistent with the other experiments. Importantly the MDS analysis suggested that subjects fairly consistently focused attention on three important dimensions to differentiate the stimuli, despite not being told (implicitly or explicitly) which dimensions were most relevant.

This study, unlike Goldstone (1994) only provided evidence for acquired equivalence. The authors suggest that these differences are due to the difficulties of the tasks involved and the nature of testing in each study. They argue that Goldstone's tightly constructed stimuli were difficult for subjects to discriminate (only 1 JND apart) and consequently subjects had to concentrate more on differentiating stimuli and focusing on differences. They further argue that this is enhanced by the discrimination testing paradigm which presents an equal number of stimuli from across the continuum and directs subjects to highlight differences between categories rather than similarities within categories. Overall, they suggest that in ini-

1 Chick-sexing is somewhat famous (and mythologized) for being a difficult task performed by highly trained professionals. In the experiment subjects were told the picture were the larynxes of two species of monkey.

tial category learning acquired distinctiveness is in operation to highlight differences in relevant dimensions and subsequent acquired equivalence compresses the categories once sufficient, robust discrimination is present.

This view is supported by Francis and Nusbaum (2002). In this study English listeners were trained to differentiate the three-way Korean phonation-type contrast found in stops. The results were analyzed using an MDS analysis in order to determine the changes in dimensional structure due to category learning. They found that subjects shifted attention as a result of training to dimensions relevant for distinguishing the stimuli. Moreover, subjects attended to a novel dimension unattended to in the pre-test condition. Evidence was also found for both within category compression and cross-category expansion. Specifically, expansion was found for categories that were difficult for subjects to differentiate in the pre-test while compression within categories was found where discriminability was higher initially.

The results of these studies suggests that the perceptual learning mechanism that drives categorization is a combination of acquired distinctiveness and acquired equivalence, but the two act in different situations. Specifically, acquired distinctiveness operates when a contrast is difficult and the learner must acquire knowledge of the distinctions in a given dimension. Acquired equivalence operates when categories are already easily differentiated but a higher degree of difference is necessary for proper categorization. In such a situation differences within a category are ignored.

A final point is that most of these studies so far described do not train subjects using different distributions of stimuli. As mentioned earlier, Maye, Werker, and Gerken (2002) induced categorical effects in infants using only bimodal distribution of stimuli, following Maye's (2000) results with adults. Subjects being trained and tested using a flat distribution of stimuli hear "bad" tokens as frequently as good ones, which may emphasize differences rather than similarities within category.

1.4.3 Summary of adult research

The experiments and results detailed above demonstrate that adults can be trained to perceive new speech categories. The success of such attempts vary greatly, depending on the difficulty of the contrast to be learned and the degree of learning that is expected. For example, in the *r*~*l* learning literature, Japanese listeners are able discriminate the sounds relatively easily. However, learning the phonemic distinction between the sounds and being able to use this in the context of minimal pairs is considerably more difficult.

In adults, training new perceptual categories results in several phenomena that result in warped perceptual spaces. This warping generally results in a smaller perceptual space within a category and expanded space between categories. Such results are generally supported by research in other modes of perception.

1.5 General conclusion

Although perceptual learning in adults and children show many differences, most obviously the differences between essentially building a language from scratch and tweaking aspects of a fully developed system, there are some important similarities. Notably, the phenomena that drive warping of the perceptual system can be witnessed both in children and adults. Similarly, both children and adults learning a new contrast must learn what cues in the auditory signal are important and weight them accordingly. In these aspects acquisition and adult perceptual learning can be seen as having valuable similarities for researchers.

The next three chapters describe a series of experiments into perception paying special attention to cues used in categorization. In Chapter 2 a stimulus set using Polish sibilant fricatives is described and preliminary tests of its suitability for further experiments with English listeners is described. Chapter 3 presents two experiments contrasting English listeners' and Mandarin Chinese listeners' perception of the stimulus set. Chapter 4 describes a training study where English listeners are trained to use different perceptual dimensions to categorize the stimuli. The final chapter provides a summary and brief discussion of the results and possible future research directions.

CHAPTER 2

ENGLISH LISTENERS' PERCEPTION OF POLISH ALVEOPALATAL AND RETROFLEX VOICELESS SIBILANTS: A PILOT STUDY

2.1 Introduction

Training studies using adult listeners allow for an exploration of category formation and relevant theoretical issues. One important factor in perceptual learning is an understanding of which perceptual cues aid in category identification and how they are weighed by listeners. This chapter examines the suitability of a stimulus design method and a phonetic contrast, Polish post-alveolar voiceless sibilants, for use in a series of perceptual learning experiments examining cue learning.

As established in Chapter 1, in order to learn a contrast and to behave as a native listener, the proper cues must be attended to and weighted appropriately by a listener. For native cues the learning process takes many years to acquire the relevant experience, with children near puberty still showing not-quite adult-like perceptual abilities. Similarly, experimental evidence with adults shows that with proper training subjects can be directed to attend to

relevant dimensions and can also learn novel cues if necessary for categorizing. Therefore it is an suitable goal of any perceptual learning study to take these issues into account and design stimuli accordingly.

2.2 Polish sibilants

A phonetic contrast must be sufficiently difficult to for subjects to show improvement over the course of training. As suggested by Best et al. (2001) listeners confronted with a new contrast will map them onto their own native categories if possible. When a contrast is subsumed under the variation of a native contrast, then that contrast will be very difficult for listeners to distinguish. For English listeners, one such non-native contrast is the Polish alveopalatal /ɕ/ and retroflex /ʂ/ sibilant categories[2], which they generally collapse under the native category /ʃ/ (Lisker 2001). This distinction is also found in many varieties of Mandarin Chinese including the Beijing dialect and the PRC Putonghua standard that is based on it.

Although both sounds are similar to English /ʃ/, they are both articulatorily distinct from it. Ladefoged and Maddieson (1996) report that while both English /ʃ/ and Polish /ʂ/ have similar constriction locations and widths along with concurrent lip-rounding, /ʂ/ has a much flatter tongue shape. For /ʃ/ and /ɕ/, they note that the tongue blade and body are higher for /ɕ/ and that both exhibit lip rounding and have very similar place of articulation

2 There is some disagreement on the exact phonetic transcription of these sounds. See Nowak (2006) and Zygis and Hamman (2003) for discussions. I follow the conventions found in Nowak (2006).

to the retroflex. In comparing the Polish sibilants with the Mandarin sibilants, they find very similar articulation strategies for both sibilants with the exception that Mandarin /ʂ/ does not exhibit lip rounding but has a larger sub-lingual cavity than its Polish equivalent.

For native speakers of Polish this distinction has several cues with the primary ones being fricative pole frequency and F2 onset (Lisker 2001, Zygis and Hamann 2003, Nowak 2006), along with slight post-consonantal vowel quality differences (Nowak, 2006). Specifically, Nowak (2006) demonstrated that proper formant transition information is necessary for native speakers identifying syllables containing these sibilants; however, the author also found that isolated fricatives could be identified reliably. He suggests that subjects used very different perceptual strategies in the different conditions, possibly perceiving the isolated fricatives as non-speech. In any event, it is clear that both fricative noise and formant transition information are used by Polish listeners for identification of these fricatives.

In contrast, English listeners show very poor discrimination of this contrast (Lisker 2001). Lisker's study, using brief training, found that English speakers could not discriminate these sounds above chance in the context of full syllables, but could discriminate if presented with either the fricatives alone or the following vowels in isolation. Lisker suggests a non-speech mode of perception to account for the discrepancies between full-syllable and isolated segment conditions. However, the fact that English listeners could, with very little training, discriminate the isolated segments suggests that more extensive training could expand on these abilities.

Given these findings, this contrast seems to be an ideal candidate for studying perceptual learning. In order to compare the two sources of identification information (fricative noise and vocalic transition) it is necessary to have a stimulus set that varies in both dimensions and has sufficient internal structure such that changes within as well as across categories can be examined. The following stimulus design and experiments address these concerns.

2.3 Experiments

2.3.1 Stimuli

The stimuli for the following experiments consist of modified, naturally produced examples of Polish alveopalatal and retroflex sibilant consonants followed by [a] in a two-dimensional space varying by fricative noise in one dimension and by vocalic transition information in the other dimension. To produce these stimuli, several productions of [ʂa] and [ɕa] were recorded in a sound-proof booth by a male native speaker of Polish using a head-mounted microphone and a Marantz PMD670 solid state recorder at 44.1 kHz sampling rate. One example of each syllable was selected based on clarity and similarity to the acoustic analyses of Polish fricatives reported in Nowak (2006). The selected retroflex syllable had a peak located at 2890 Hz and an F2 onset at 1420 Hz with a midpoint F2 of 1280 Hz. The alveopalatal has a peak located at 3890 Hz with an F2 onset of 1720 Hz and a midpoint F2 of 1320 Hz.

Each syllable was split in two at the boundary of the fricative and vocalic portions, determined by the onset of voicing. The two fricatives were brought to the same length by excising 32ms from [ɕ] in four 8ms chunks located at 20% intervals of the total length. The vowels were modified using Praat (Boersma and Weenik 2002) to have the same length, pitch, and RMS through PSOLA resynthesis. Both the fricative and vowel portions were then separately interpolated to form fricative and vowel continua consisting of 10 steps where each step was one of ten graded proportions in terms of intensity. That is, fricative step 0 consisted of 9/9 [ɕ] and 0/9 [ʂ], while step 1 consisted of 8/9 [ɕ] and 1/9 [ʂ], step 2 was 7/9 [ɕ] and 2/9 [ʂ], etc. Figure 2.1 displays spectra of selected fricative steps (20ms window from the center of the fricative, cepstral-smoothed 500Hz bandwith.) Table 2.1 displays vowel formant measures taken at 25ms and 100ms from onset of voicing.

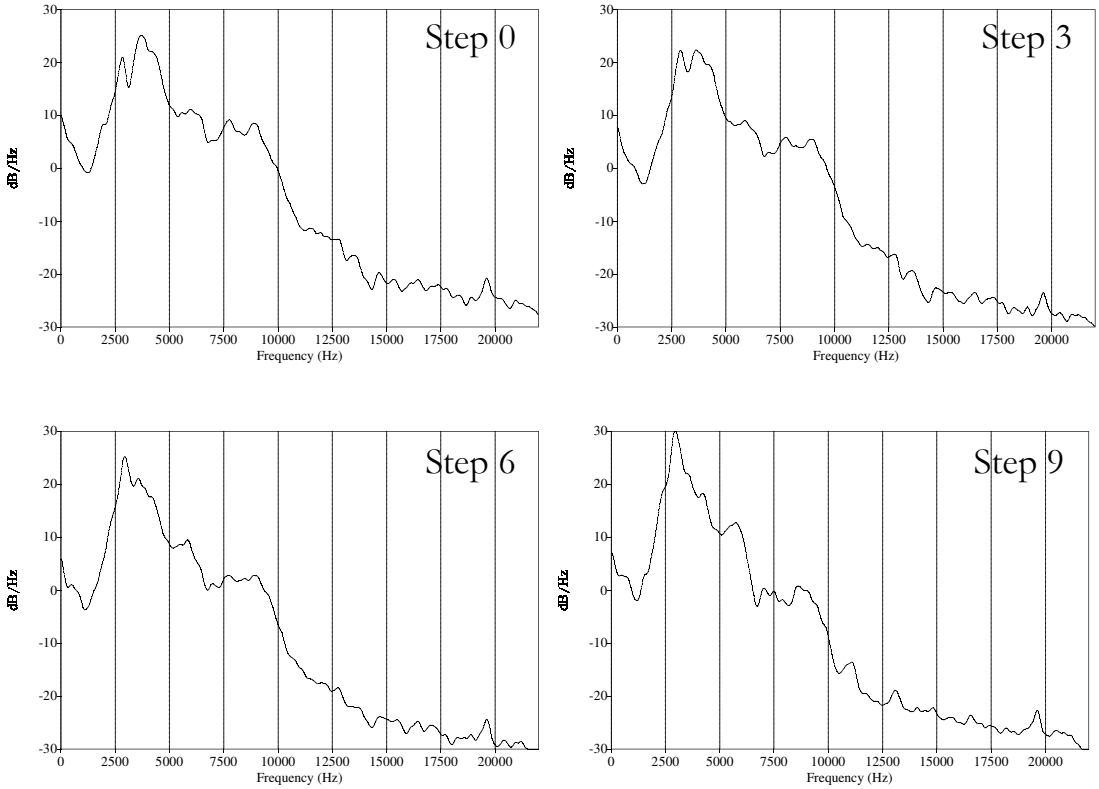

Figure 2.1: Spectra of fricative step 0 (fully alveopalatal), step 3, step 6, and step 9 (fully retroflex).

Step	25ms		100ms	
	F1	F2	F1	F2
0	680	1629	775	1315
1	684	1624	778	1317
2	690	1617	777	1315
3	701	1607	777	1311
4	716	1597	776	1308
5	734	1581	775	1304
6	752	1556	776	1300
7	763	1513	775	1295
8	765	1455	776	1289
9	762	1392	777	1283

Table 2.1: Formant values for each step of the vowel continuum as measured at +25ms and +100ms from the onset of voicing.

CV syllables were produced by concatenating each fricative with each vowel yielding 100 tokens varying in two dimensions, vowel transition and fricative noise. The concatenated "natural" syllables are shown in Figure 2.2. The 10 X 10 stimulus set is graphically represented in Figure 2.3.

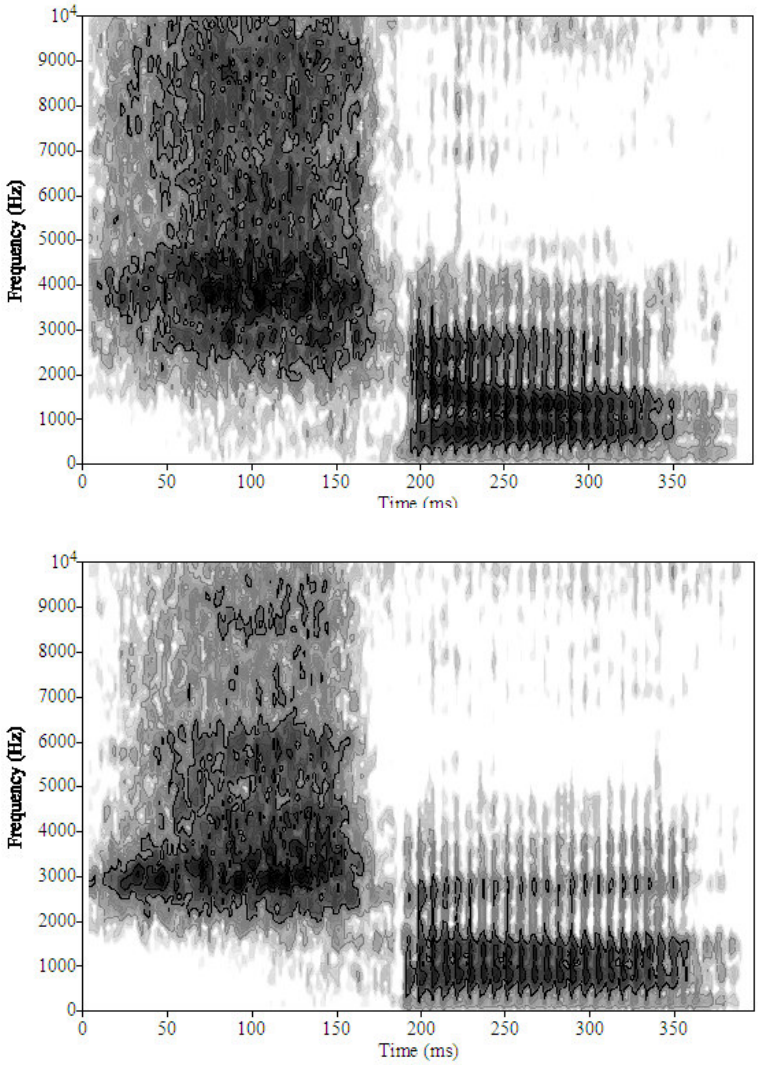

Figure 2.2: Spectrograms of the fully alveopalatal (top) and fully retroflex (bottom) syllables.

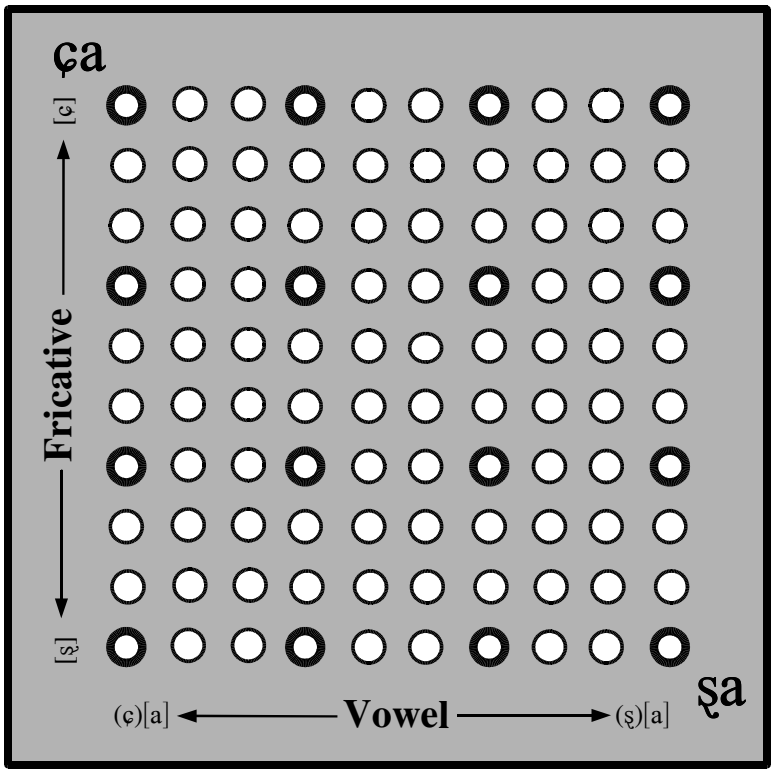

Figure 2.3: Stimuli design. Each circle represents a particular combination consonant and vowel from each continuum. Circles with darker outlines represent a subset of tokens for orthographic labeling (see text.)

2.3.2 A Polish listener's perception

In order to assure that the stimuli still represent consistent linguistic categories after modification, a native speaker of Polish was asked to label them. The participant was the same speaker who produced the stimuli and is a trained linguistic phonetician who has studied Polish fricatives, but was not aware of how the stimuli had been manipulated. Each stimulus was presented in random order, in a single block (n=100). Three blocks were presented

for a total of 300 trials. The subject was asked to label each stimulus by responding using a five-button box with the leftmost button labeled *sza* (retroflex) and the rightmost labeled *śa* (alveopalatal). The interval between trials was 3s, there was no feedback given.

The labeling results show a clear categorical distinction between the two categories (see Figure 2.4). The boundary along the fricative dimension is approximately between fricative step 4 (f4) and fricative step 5 (f5), the center of that dimension. The vocalic boundary is shifted considerably towards the retroflex end of that dimension, around vocalic step 6 (v6) and vocalic step 7 (v7).

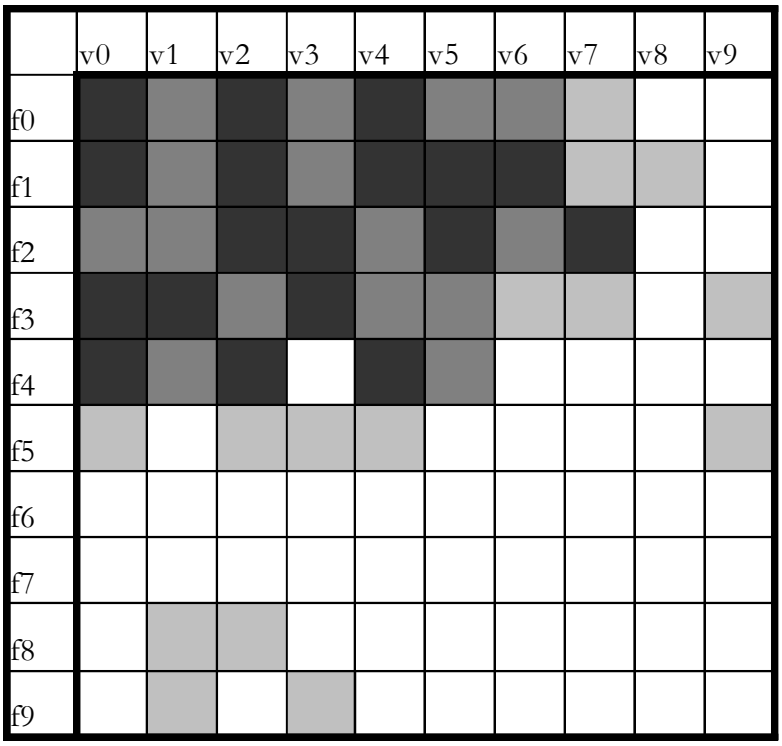

Figure 2.4: Labeling results of the stimuli set by a native speaker. Black indicates for 3/3 trials the stimulus was labeled as alveopalatal; dark gray, 2/3; light gray 1/3; white, 0/3, i.e. labeled 100% as retroflex.

These labeling results indicate a clear categorical boundary for this listener. Both dimensions were used for classification, though the boundary was not symmetrical as many more tokens were labeled as retroflex than alveopalatal. It is difficult to draw too many conclusions beyond this from one listener, especially since the listener was a trained linguist. Important to the questions at hand, however, is how English listeners perceive these sounds. The following studies attempt to shed light on that question.

2.3.3 English listeners' perception of the Polish stimuli

In order to establish the suitability of these stimuli for a training experiment, English listeners' perception of the stimuli was examined with the following questions in mind: 1) Do English listeners uniformly assimilate the Polish contrast to English /ʃ/, or are they sensitive to differences and able to categorically label differences? 2) If English listeners are not sensitive to the differences in the stimuli, what amount of training would be necessary to achieve categorical perception? In order to answer these questions, two studies were run. In the first, listeners labeled the stimuli using English orthography and in the second subjects were briefly trained to categorize the stimuli.

2.3.4 English orthographic labeling

To ascertain English speakers' judgments concerning the stimuli, five native speakers of American English labeled a subset of the stimuli using English orthography. The subset consisted of 16 equally spaced stimuli from the larger set of 100. The stimuli in this subset sample the full range of the fricative and vowel transition dimensions separating Polish [ʂ] and [ɕ] (see fig. 1). Listeners labeled each token five times. Tokens were presented in random order (five repetitions of the list of 16 tokens or one randomization of the 16*5 trials) and the listeners entered their responses on a computer keyboard. Each response was presented presented back to the subject on the subject's computer screen. After two seconds a new sound was presented along with a blank screen.

Table 2.2 shows the labels used by each subject and their frequency of use. The label *sha* was the most commonly used label for all subjects. The second most common labels were *shya* and *shia,* and only subject 101 did not use one of these two labels. Additionally, subject 100 used a large number of different labels (9), although only a handful of these were used frequently. The other subjects used fewer labels, with 104 only using 3.

Label	**s100**	**s101**	**s102**	**s103**	**s104**	**Total**
sha	*44*	*47*	*39*	*40*	*53*	**223**
shya				*26*	*25*	**51**
shia	*20*		*25*	*1*		**46**
shot		*20*	*1*			**21**
schia	*6*			*10*		**16**
ssha		*10*				**10**
sa			*8*			**8**
scha	*5*			*2*		**7**
tza			*5*			**5**
sha3					*2*	**2**
shaw			*1*	*1*		**2**
shz			*1*			**1**
tsha		*1*				**1**
zha	*1*					**1**
shja	*1*					**1**
shout		*1*				**1**
chia	*1*					**1**
s	*1*					**1**
sch	*1*					**1**
sja		*1*				**1**

Table 2.2: Labeling responses from the five subjects.

All subjects used *sha* as a label for all stimuli, except for subject 104, who never labeled stimulus f0-v0 (alveopalatal fricative + alveopalatal vowel) as *sha*, but otherwise used it extensively. Four of the five subjects indicated a distinction along the vowel continuum using the label *shya* or *shia*, typically used for stimuli with vowels v0 and v3 (i.e. the alveopalatal end of the continuum). Other labels were used extensively, but none showed a clear pattern of use. The lone exception to this is a single subject (101) who used *ssha* as a label for f0 and f3 tokens and did not use any label to indicate a distinction along the vocalic dimension. This subject also used the label *shot* frequently (20 times), though there was no discernible pattern Further exceptional behavior from this subject will be discussed in the next section. Table 2.3 shows the use of *sha* and *shia, shya,* or *ssha* by all subjects.

Subject 100									
"sha"	v0	v3	v6	v9	"shia"	v0	v3	v6	v9
f0	*2*	*1*	*3*	*5*	f0	*2*	*2*	*1*	*0*
f3	*1*	*2*	*4*	*4*	f3	*3*	*0*	*1*	*1*
f6	*2*	*1*	*4*	*5*	f6	*1*	*2*	*0*	*0*
f9	*1*	*1*	*2*	*6*	f9	*3*	*2*	*2*	*0*

Subject 101									
"sha"	v0	v3	v6	v9	"ssha"	v0	v3	v6	v9
f0	*2*	*2*	*1*	*2*	f0	*2*	*0*	*1*	*1*
f3	*2*	*2*	*3*	*4*	f3	*3*	*2*	*1*	*0*
f6	*3*	*4*	*4*	*4*	f6	*0*	*0*	*0*	*0*
f9	*4*	*4*	*3*	*3*	f9	*0*	*0*	*0*	*0*

Subject 102									
"sha"	v0	v3	v6	v9	"shia"	v0	v3	v6	v9
f0	*2*	*2*	*2*	*4*	f0	*3*	*3*	*2*	*0*
f3	*2*	*1*	*3*	*2*	f3	*2*	*3*	*1*	*0*
f6	*3*	*1*	*4*	*2*	f6	*2*	*3*	*0*	*0*
f9	*2*	*1*	*4*	*4*	f9	*3*	*3*	*0*	*0*

Subject 103									
"sha"	v0	v3	v6	v9	"shya"	v0	v3	v6	v9
f0	*2*	*2*	*2*	*4*	f0	*3*	*3*	*0*	*0*
f3	*2*	*1*	*3*	*3*	f3	*4*	*2*	*0*	*1*
f6	*1*	*2*	*4*	*3*	f6	*3*	*4*	*1*	*0*
f9	*2*	*2*	*3*	*4*	f9	*2*	*3*	*0*	*0*

Subject 104									
"sha"	v0	v3	v6	v9	"shya"	v0	v3	v6	v9
f0	*0*	*2*	*4*	*5*	f0	*5*	*3*	*0*	*0*
f3	*2*	*4*	*5*	*5*	f3	*3*	*1*	*0*	*0*
f6	*1*	*3*	*5*	*5*	f6	*4*	*2*	*0*	*0*
f9	*1*	*2*	*4*	*5*	f9	*4*	*3*	*0*	*0*

Table 2.3: Response tallies for each subject for the label "sha" and any other label consistently used.

Overall, these subjects are using labels indicating the English palatal glide for stimuli having the alveopalatal formant transitions, although all stimuli may be labeled as *sha*. However, subjects, with one exception, are not making a distinction along the fricative dimension. This would seem to indicate that most English listeners are able to perceive a categorical difference in only the vocalic dimension. Nevertheless, it is also quite possible that limitations in English orthographic representations make it possible to represent the alveopalatal transitions, but not the fricative noise.

Note that the one subject who did label the alveopalatal fricatives differently resorted to a novel representation, *ssha*. This subject's other unique label, *shot,* was not used consistently. Interestingly, this subject has significant experience with Turkish, speaking with native proficiency. The extent to which this accounts for the unique labeling pattern is unknown[3].

In order to discern whether the variation in fricative noise can be used by English listeners further exploration is necessary. Consequently. the following experiment is designed to train listeners to associate abstract labels with specific regions of the stimulus set.

2.3.5 English labeling with brief training

In this experiment subjects were given brief training with the Polish labels for these sounds and then asked to label the entire stimulus set. This avoids the restrictive English orthographic labels and allows for a test of listeners' abilities to learn distinctions in the stimulus set. Training sets were designed to push learners to either use both the fricative and vowel dimension as the Polish native speaker did, or to use only the fricative dimension which most American English listeners did not do in experiment 1.

Ten subjects (three of whom participated in the labeling described above, subjects 101, 103, 104) were given brief training on specific labels, *sz* and *ś*, and then labeled the entire stimuli set (five repetitions of each stimulus, random order.) Subjects were told that they would be learning two sounds in Polish that are very similar to the English sound *sh*. Train-

3 Two additional Turkish listeners were run through a similar experiment to explore the hypothesis that Turkish listeners are sensitive to fricative variation and ignore formant transitions for this contrast. The results were inconsistent, leaving the issue unresolved. However, one subject during debriefing indicated that he (incorrectly) believed the focus of the experiment was to test his ability to discriminate final unreleased stops and that some of the stimuli (same as this experiment) had final stops.

ing consisted of passive listening to 18 tokens where each token was presented with a simultaneous visual presentation of the desired label. The training procedure continued with a session in which listeners labeled the same tokens with accuracy feedback. In the training phase of the experiment, each token was presented five times for a total of 90 trials (about 7mins); the correct response was presented if a subject responded incorrectly.

Subjects were divided into groups differing by the distribution of training stimuli (Figure 2.5). In one group, five subjects were trained on the 18 tokens closest to the natural tokens (9 alveopalatals and 9 retroflexes), "natural category learners" (NC). The second group consisted of a total of five subjects who were trained on tokens varying maximally on the fricative dimension and minimally on the vowel dimension, "fricative categorizers" (FC). The FC subjects were further divided into two groups, three subjects training on tokens from the alveopalatal end of the vowel continuum, and two from the retroflex end. Training duration was identical for all groups, only the stimulus set for training varied.

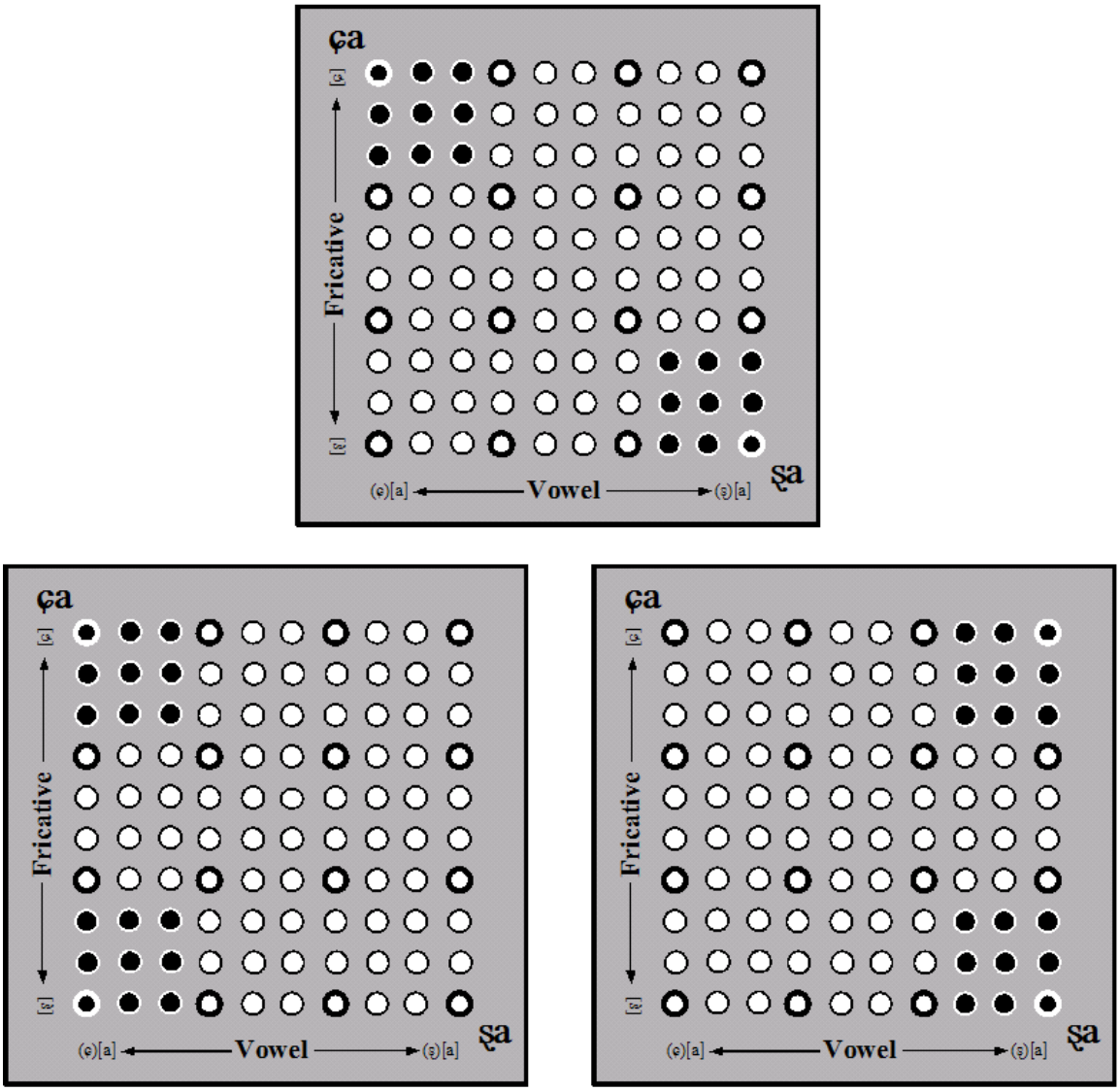

Figure 2.5: Pilot training sets; dark filled circles represent tokens used for training, white filled tokens were used for testing (along with the training tokens.) Heavier lines indicate tokens used for orthographic labeling. Clockwise from top: 1) natural category group, 2) fricative categorizers (retroflex transitions), 3) fricative categorizers (alveopalatal transitions)

2.3.6 Results

With a single exception, all subjects in the natural category group ignored variation along the fricative dimension and instead relied on the vocalic dimension for determining category membership (see Figure 2.6). This is in contrast to the Polish listener who used both dimension for categorization. One subject, however, used the fricative dimension only and ignored the vocalic dimension. This is the same subject mentioned above as using *sha* and *ssha* labels and is further unique in being a fluent speaker of Turkish in addition to English. All subjects had a training accuracy greater than 85%.

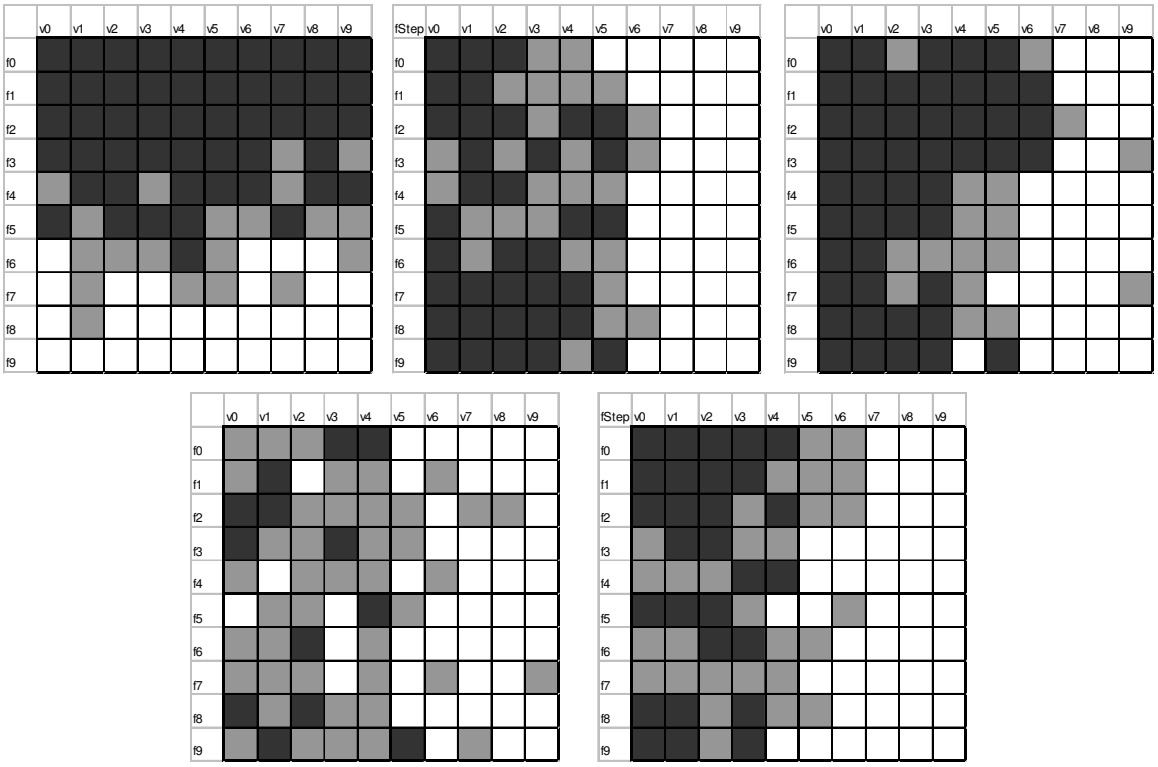

Figure 2.6: Results from the natural category training subjects. Dark shading indicates the stimulus was labeled as alveopalatal >66%; gray shading indicates between 33% and 66% alveopalatal labeling; white indicates <33% alveopalatal labeling. The top left panel is the subject that categorized along the fricative dimension.

The fricative categorizers show a similar pattern (see Figure 2.7), despite different training. Only one subject showed overall categorization along the fricative dimension, and a second subject reversed the labels. Performance in training was poorer than that for the natural category subjects and much more variable, from near chance (50%) to 75% accuracy.

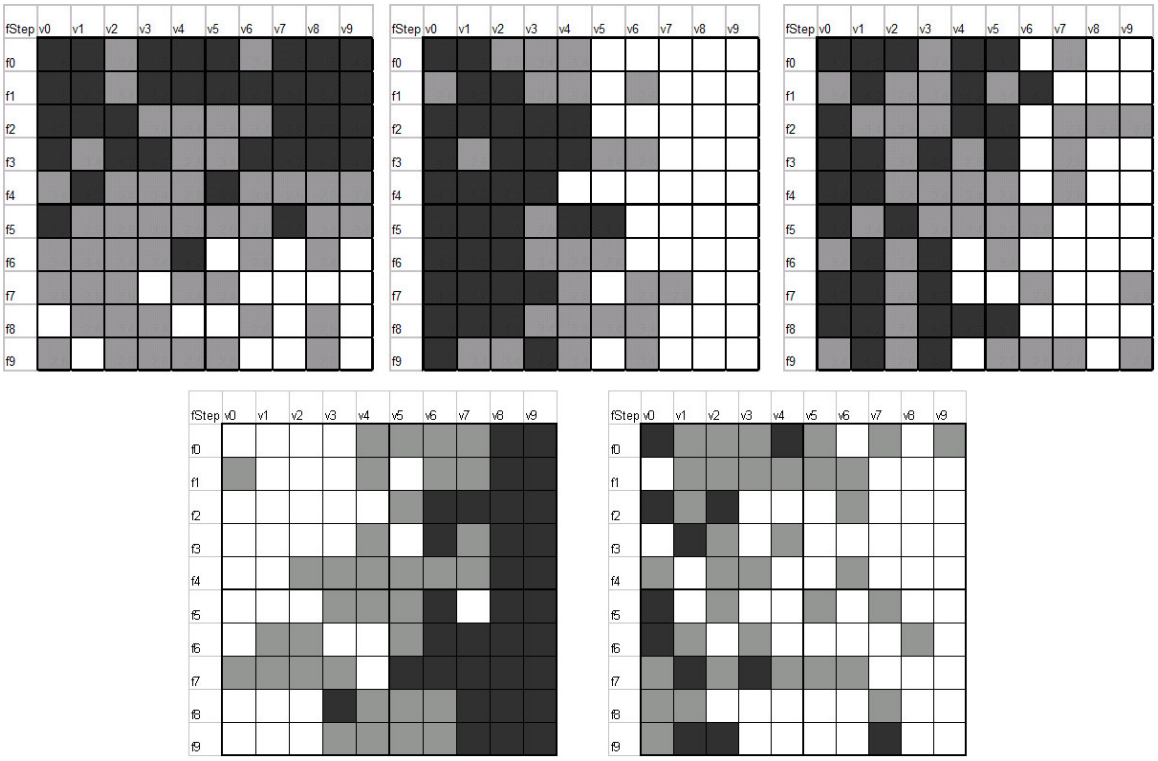

Figure 2.7: Results of fricative categorizers. Top row shows subjects trained on alveopalatal vocalic tokens and bottom row shows retroflex.

Interestingly, three of these subjects show a change in strategy during the labeling testing, switching from an initial categorization based on the fricative continuum and later switching to one based on the vocalic cues. The subjects who performed better in training showed this pattern. This is most dramatically demonstrated by subject 303, a subject with high accuracy during the training phase from the FC training group (Figure 2.8).

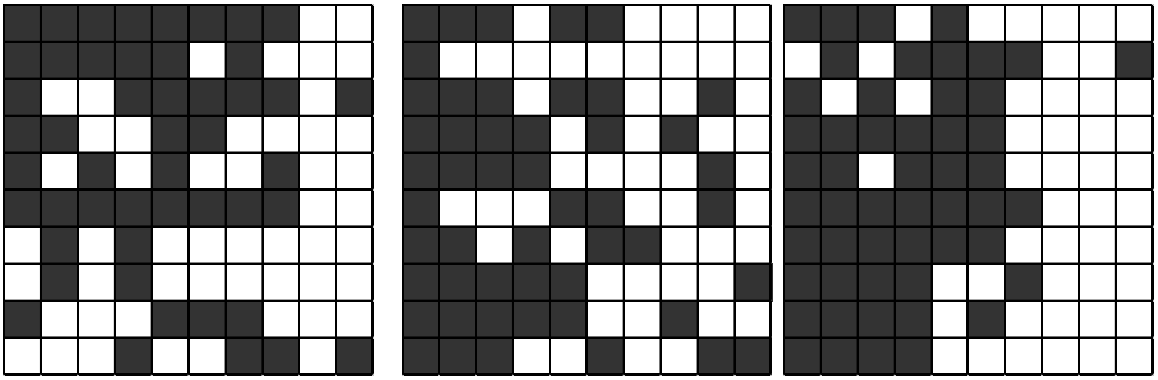

Figure 2.8: Times series results from subject 303. Left panel represents block 1/5, central panel shows block 3/5, and right panel shows block 5/5. Black indicates stimulus labeled as alveopalatal, white as retroflex.

Overall the results demonstrate that the vocalic cues to the perception of this contrast are more robust than the fricative cues for these subjects and that little training is necessary to achieve reliable categorization along this dimension. With brief training on categorization along the fricative dimension subjects will either ignore those cues outright or gradually shift to using the vocalic cues. However, the fact that some listeners were able to use the fricative dimension during training and over the first block of test indicates that additional training may make the fricative dimension cues more robust and override the vocalic cues.

The role of linguistic experience is tantalizingly hinted at in these results as well. The lone subject with significant experience in another language, Turkish, showed consistently different results from the more monolingual English listeners. Further, none of the listeners used both dimensions to categorize the stimuli as the native Polish speaker did.

2.4 General discussion

In general, these results show that the perceptual differences among the Polish post-alveolars are only partially available to English listeners without training. Although both categories can be perceived by English listeners as /ʃ/ and labeled as such, subjects can reliably perceive a difference between them, labeling the alveopalatal's transition as a palatal glide.

Especially interesting in the results is the ability of English listeners to categorize using one dimension (vocalic) while showing extreme difficulty in using the other dimension (fricative noise). In a training experiment, this allows for a comparison between cues that are easily attended to and those that are difficult to attend to. Further, the initial success some subjects had with training in the fricative dimension suggests that with more extensive training, American English listeners may be able to use this cue reliably.

Also of interest in these results is a contradiction of the Lisker (2001) study. In those experiments English listeners could not reliably identify the alveopalatal and retroflex sibilants as different in the context of full syllables. However, these results show that English listeners can differentiate these sounds, but only using vocalic information. This discrepancy is likely due to differences in experimental design. In this experiment, listeners only had to identify the alveopalatal and retroflex voiceless sibilants as distinct. In Lisker's experiment, subjects also had to identify the Polish dental sibilant, making the three-way place distinction. The subjects were quite reliable in identifying the sibilant as different from alveopalatal and retroflex, which follows Best et al. (2001) as the dental sibilant is quite similar to English /s/

(Lisker 2001, Nowak 2006). It is possible that this additional perceptual demand severely restricted subjects' ability to focus on the relevant distinguishing characteristics of the post-alveolars.

Generally, these stimuli appear to be suitable for a perceptual training study. Training to reliably use fricative noise cue information should be relatively minimal and can be contrasted with vocalic information, which is highly robust. The stimulus set is sufficiently natural with a full specification of information available in natural speech, yet has variation that can be used to explore changes in perceptual space. Although listeners do not initially attend to the fricative noise difference between Polish [ʂ] and [ɕ], the data here suggest that American English listeners can be trained to use this subtle acoustic cue. The vowel formant difference between Polish [ʂ] and [ɕ] seems to be more salient to American English listeners and thus should form the basis for a strong category if training enforces this tendency.

CHAPTER 3

PERCEPTION OF THE POLISH ALVEOPALATAL ~ RETROFLEX SIBILANT CONTRAST BY ENGLISH AND MANDARIN LISTENERS

3.1 Introduction

The previous chapter demonstrated that English listeners can perceive a distinction in the Polish alveopalatal ~ retroflex distinction, but that they tend to rely on vocalic cues to do so. This is unlike a Polish listener who used both fricative and vocalic cues to make the distinction. More generally, this demonstrates that listeners may not use all of the cues available to them for a non-native contrast, even though such information is used to make other contrasts in the language (i.e. /s/ ~ /ʃ/). This chapter explores in more detail English listeners' perception of the Polish contrast and compares that perception with listeners of a different language group that has similar categories to Polish, i.e. Mandarin.

The previous chapter left a few questions unanswered. First, the extent to which subjects are aware of the fricative variation present in the stimuli is unknown. One subject demonstrated an ability to use that dimension, while all other subjects ignored the fricative

dimension unless their attention was directed to it through training. Even under such training conditions, however, subjects showed a strong bias towards reliance on the vocalic cues available.

Similarly, this shows a lack of integration of the two cue sources. This is somewhat unexpected as consonant - vowel co-articulation significantly affects the perception of place for fricatives (LaRiviere et al. 1975, Whalen 1981, 1983), suggesting a high degree of integration (e.g. Grau and Kemler Nelson 1988, Kemler Nelson 1993). Similarly for stops, Wood and Day (1975) demonstrated that variation in an irrelevant dimension degraded perceptual discriminability, which they attribute to perceptual integration. However, this perceptual integration effect is present only for consonants perceived as speech; sounds heard as non-speech do not show cue integration (Best et al. 1981, Tomiak et al. 1986). The results presented in Chapter 2 provide evidence that English listeners have not generalized their knowledge of the relationship between spectral information resulting from fricative noise and second formant locus and movement to this new contrast. That is, for non-native sounds listeners do not integrate the cues having no experience with them in that context. A goal of this chapter will be to explore this issue in more detail.

Finally, the limited nature of the experiments in Chapter 2 precluded any predictive statistical analysis. Moreover, the lone Polish speaker used to demonstrate the naturalness and clear categorization of the stimulus set was not a naive subject. In order for more reliable and predictive results, a larger and broader subject population is necessary.

The two following experiments attempt to answer some of these issues using the same stimulus set described in Chapter 2. The first experiment uses a discrimination paradigm to compare subjects' perception of correlated and conflicting cue stimuli (a standard methodology used to explore perceptual cue integration – a la Best et al. 1981), while the second experiment examines labeling in more detail. For both experiments native speakers of English and Mandarin participated as subjects, in order to compare performance by listeners who do not control the retroflex/alveopalatal distinction (English) with listeners who do control the distinction (Mandarin). Mandarin, especially certain dialects (see e.g. Li 2006), has a three-way sibilant distinction that is usually described as being identical to that of Polish (Ladefoged and Maddieson 1996, Stevens et al. 2004), thus allowing a comparison of non-native and near-native perceptual differences.

3.2 Discrimination experiment

This experiment uses a speeded AX discrimination paradigm (same-different). This design has been shown to minimize effects of native language and tap sensory-trace rather than context coding memory (Pison 1973, Fox 1984, Johnson 2004, Johnson and Babel 2007). It is used here to explore perceptual cue integration, following Best et al. 1981, and is similar to a Garner paradigm experiment (Garner 1974). In this experiment the four extreme stimuli, the two correlated cue endpoints and two conflicting cue endpoints, are contrasted such that there are six different pairs: two pairs differing in both dimensions (correlated~correlated, conflicting~conflicting), and four pairs differing in one dimension (two fricative, two vocalic). From an information theoretic standpoint, pairs differing in two

dimensions should be easier to discriminate than pairs differing in a single dimension. However, if the speech perception mechanism is not by-passed by the task, then an effect of integrality, i.e. slower reaction times caused by the unnaturalness of the conflicting cue stimuli, should be present.

3.2.1 Materials

The stimuli for this experiment are drawn from the stimulus set described in Chapter 2. In this experiment, only the four "corner" stimuli are used, that is the alveopalatal fricative and vowel endpoint stimulus (f0v0), the retroflex fricative and vowel endpoints (f9v9), the alveopalatal fricative endpoint with retroflex vowel endpoint (f0v9), and the retroflex fricative endpoint with the alveopalatal vowel endpoint (f9v0).

3.2.2 Subjects

Two groups of subjects participated. Twenty-four native speakers of English between the ages of 18 and 23 were recruited from the University of California at Berkeley. Twenty-two native speakers of Mandarin who varied in age from 20-32 were also recruited from the University of California at Berkeley. Only subjects from northern provinces and Beijing were run in order to increase the likelihood that all subjects had the full three-way distinction in sibilants. Specifically, subjects were from Jilin, Heilongjiang, and Liaoning provinces as well as the municipality of Beijing. As all of the subjects were students at the University of California, Berkeley, they all spoke English proficiently.

3.2.3 Procedure

The experiment used a speeded AX (roving) same-different discrimination paradigm. Subjects heard two sounds and then responded by indicating whether the two sounds were exactly identical or slightly different by pressing a button on a five button box. The leftmost button was labeled "same" and the rightmost "different". Subjects were instructed to use the index finger of each respective hand to press the two buttons as well as to leave each finger on the button between trials. The interstimulus interval was 100ms and the inter-trial interval was 1s. Subjects had up to 2s to respond before the message "no response detected" was shown on the screen and a new trial initiated. After each response, the accuracy and speed of that response was reported on the subject's computer screen for 1s. Subjects were instructed to respond "different" to any difference that they heard and to respond in less than 500ms with greater than 75% accuracy.

There were twelve different pairs (see Figure 3.1), each possible combination of the four sounds (six pairs X two orders), and four same pairs. In each block subjects participated in 24 trials consisting of each different pair balanced by 12 same pairs (4 X 3). There were seven such blocks for a total of 168 trials (84 "different", 84 "same"). At the conclusion of each block the cumulative accuracy (percent correct) and cumulative reaction time were presented on the screen along with the expected values (i.e. 500ms, 75% correct). Prior to the seven blocks, a single practice block of five random pairs was presented.

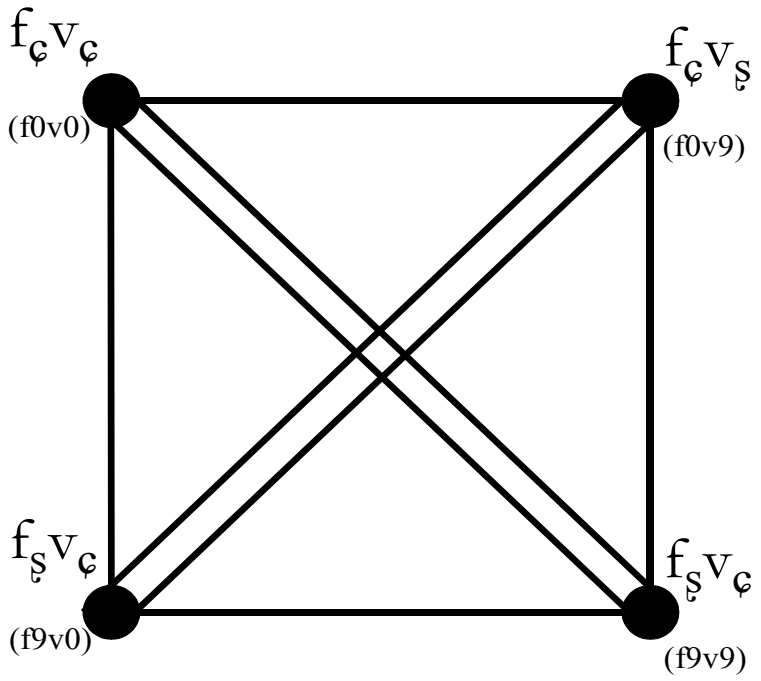

Figure 3.1: Pairs used in discrimination experiment. Each circle represents one stimulus. Lines connect pairs; single lines connect pairs differing in a single dimension, double lines connect pairs differing in two dimensions.

3.2.4 Results and discussion

Both accuracy and reaction time were analyzed. Accuracies were converted to *d'*, a bias free measure of sensitivity (differencing model; Macmillan and Creelman 2005), such that a *hit* is defined as a correct response to a "different" pair and a *false alarm* is defined as responding "different" to a "same" pair. A repeated measures ANOVA was run using the factors Pair and Language Group. There were no significant differences in language group or by pair. However, for all pairs for both groups sensitivity was significantly above 0 (Eng: mean = 1.35, σ = 1.06; Mand: mean = 1.65, σ = 0.93). Bias, *c*, was also analyzed in the same

way. Using the definitions of hits and false alarms given above, a positive bias indicates a tendency to respond "same" and a negative bias indicates a tendency to respond "different". This analysis also found no significant differences. The mean bias of 0.09 was not significantly different from 0.

A second sensitivity analysis comparing performance across blocks was conducted to examine differences potentially caused by accuracy feedback. A repeated measures ANOVA was run using the factors Block (1-7) and Language Group (Figure 3.2[4]). A significant effect was found for Block ($F[6,258] = 4.84$, $p < 0.001$) and a significant interaction between Block and Language ($F[6,256] = 3.22$, $p < 0.01$). This interaction was explored using two separate ANOVAs run within each language group. Only the English group showed a significant effect for Block ($F[6,132] = 6.92$, $p < 0.001$). Tukey post-hoc tests reveal a significant difference between blocks 1 and 6 ($p < 0.05$) and 1 and 7 ($p < 0.05$) for the English listeners. Moreover, the two language groups were different only for block 1 ($p < 0.05$).

4 The error bars in this and all subsequent graphs display the 95% confidence interval calculated using a *t*-distribution

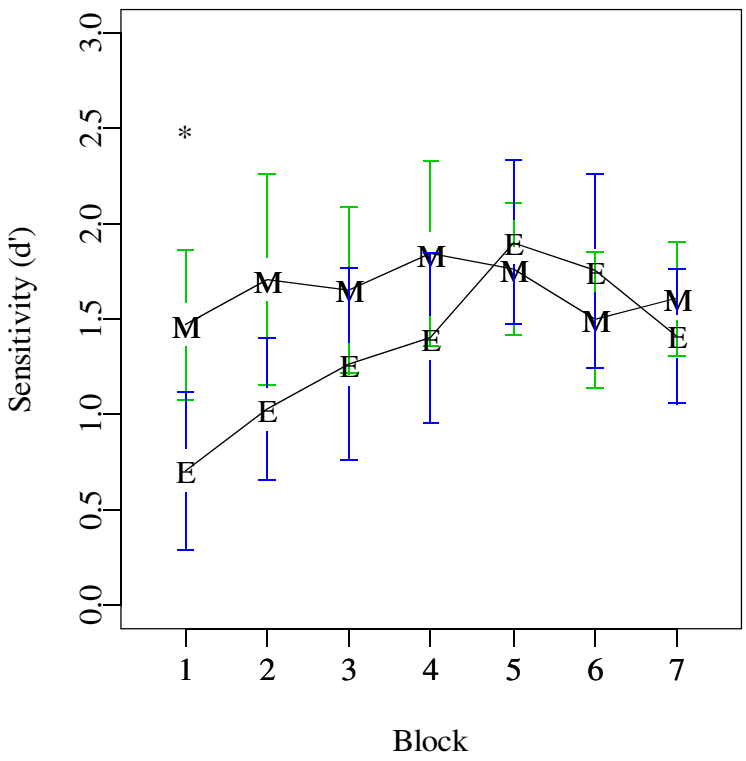

Figure 3.2: Sensitivity by block. "M" and "E" indicate Mandarin and English, respectively.

Unlike the Mandarin listeners, the English listeners show improvement over the first half of the experiment, eventually achieving the same performance level as the Mandarin listeners. Presumably this is an effect of the accuracy feedback. However, an analysis of the change from block 1 to block 7 in English listeners' sensitivities to each pair showed no significant differences, only the overall difference between blocks reported above. Therefore the changes in sensitivity seen over the course of the experiment are general and not localized to specific pairs or dimensions of difference.

An analysis of reaction times of correct responses greater than 100ms and less than 1500ms was also conducted. These reaction times (log transformed) were analyzed using a repeated measures analysis of variance having the within subjects factor Pair and between subjects factor Language. Pairs were collapsed by order and only the six "different" pairs were analyzed. A significant main effect was found for Pair ($F [9,387] = 10.6$, $p < 0.001$) as well as a significant interaction between Pair and Language ($F [9,387] = 3.5$, $p < 0.001$; see fig. 3.2). This interaction was explored through two further ANOVAs for each language group. This analysis found a significant effect of Pair for only the Mandarin group ($F [5,105] = 13.7$, $p < 0.001$). A subsequent series of Tukey HSD post-hoc tests found significant ($p < 0.5$) differences between pairs f0v0~f0v9 and f0v0~f9v0 (fully alveopalatal and the two conflicting stimuli), f0v0~f9v9 (fully alveopalatal and fully retroflex), and f9v0~f9v9 (the pair differing in the vocalic dimension only and having retroflex fricative noise), as well as between f0v0~f9v0 and f9v0~f9v9 (the stimulus having alveopalatal fricative noise and retroflex vocalic portion compared with both fully correlated stimuli). Further comparisons between language groups show a significant difference between English and Mandarin listeners ($p < 0.05$) for the pairs f0v0~f0v9 and f0v9~f9v0 (see Figure 3.3).

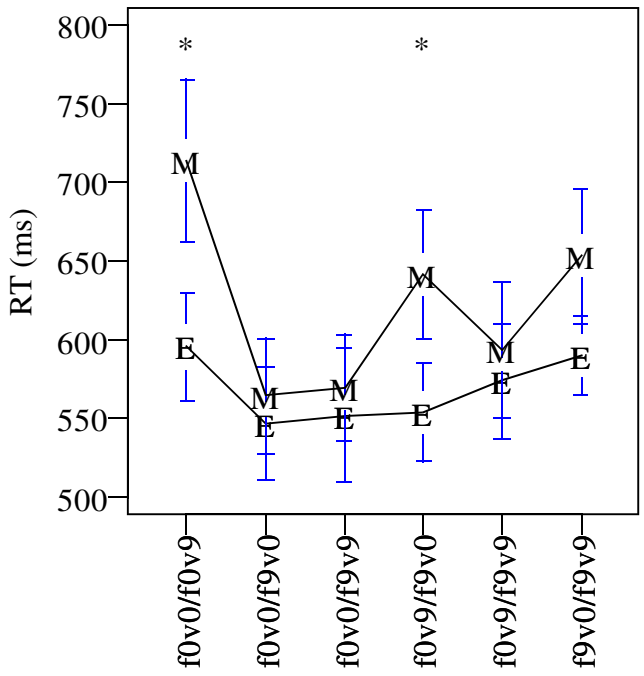

Figure 3.3: Mean reaction times for Mandarin (M) and English (E) listeners. Asterisks indicate significant differences in means between language groups.

These results show that Mandarin listeners responded most slowly to the two pairs which differed only in the vocalic dimension, especially the pair having alveopalatal fricative noise (f0v0~f0v9). Mandarin listeners were also slower than their English listener counterparts in discriminating the two conflicting cue stimuli from each other..

In order to get a better understanding of these data, the pairs were recoded and further collapsed by dimension, i.e. whether the pair varied by a single dimension and which dimension that was. A repeated measures analysis of variance with the factors Dimension

(fricative, vocalic, both) and Language was run finding a significant main effect for Dimension ($F [2,86] = 26.0$, $p < 0.001$) and a significant interaction between Dimension and Language ($F [2,86] = 4.60$, $p < 0.05$). Tukey post-hoc tests showed significant differences between the language groups for pairs differing along the vocalic dimension ($p <0.001$) and in both dimensions ($p < 0.05$), see Figure 3.4. The dimensions did not differ significantly within the English listeners, but did within the Mandarin listeners where the vocalic dimension was significantly different from the fricative dimension ($p < 0.001$) and from pairs differing in both dimensions ($p < 0.01$).

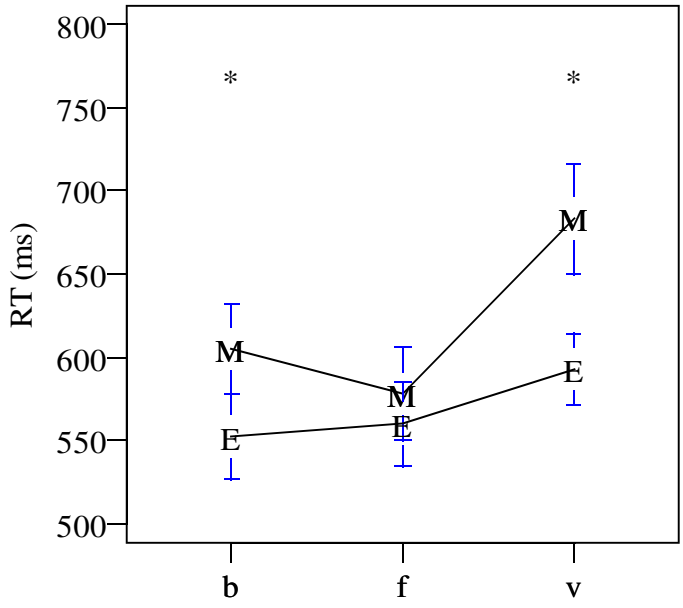

Figure 3.4: Mean reaction times for English listeners (E) and Mandarin listeners (M) by dimension. B = both dimensions differ, f = fricative dimension differs only, v = vocalic dimension differs only. The "*" indicates a significant difference in means ($p < 0.5$).

These results generally show that Mandarin listeners are slower than English listeners when the stimuli vary only in the vocalic dimension, suggesting a lower degree of reliance on this cue for discrimination. English listeners were also a little slower on average when discriminating pairs solely on the basis of the vowel transition; however this trend for English listeners was smaller (and non-significant) than it was for Mandarin listeners. Unlike English listeners, Mandarin listeners are also slower when both cues differ. This result is explained by

the fact that Mandarin listeners were slowed when discriminating the conflicting cue stimuli, an effect not present for the English listeners, who perform equally well for correlated and conflicting cue stimuli.

A third analysis was run with the pairs recoded and collapsed by their status as correlated or conflicting. Specifically, the fully correlated cue pair f0v0~f9v9 (the natural pair), the fully conflicting cue pair (f0v9~f9v0), and the mixed pairs that differ in only one dimension (e.g. f0v9~f9v9) were all contrasted as three separate groups. The log transformed reaction times were analyzed in a repeated measures ANOVA having the between subjects factor Language (Mandarin, English), and the within subjects factor Cor (Correlated, Conflicting, Mixed). This analysis resulted in a significant main effect for Cor ($F[2,86]=8.33$, $p<0.001$) and a significant interaction between Cor and Language ($F[2,86]=3.61$, $p<0.05$). A subsequent Tukey post-hoc test showed significant differences (see Figure 3.5) in means between Mandarin and English listeners for the conflicting ($p<0.05$) and mixed cue conditions ($p<0.01$) and between both language groups RTs overall ($p<0.001$). The interaction was investigated further by separate repeated measures ANOVAs for each language group using the factor Cor. A significant effect was found only within the Mandarin listeners ($F[2,42]=10.43$, $p<0.001$) and Tukey post-hoc tests demonstrated a significant difference between subjects' reaction times for the correlated stimuli and the conflicting cue stimuli ($p<0.05$) as well as the correlated stimuli and the mixed stimuli ($p<0.05$).

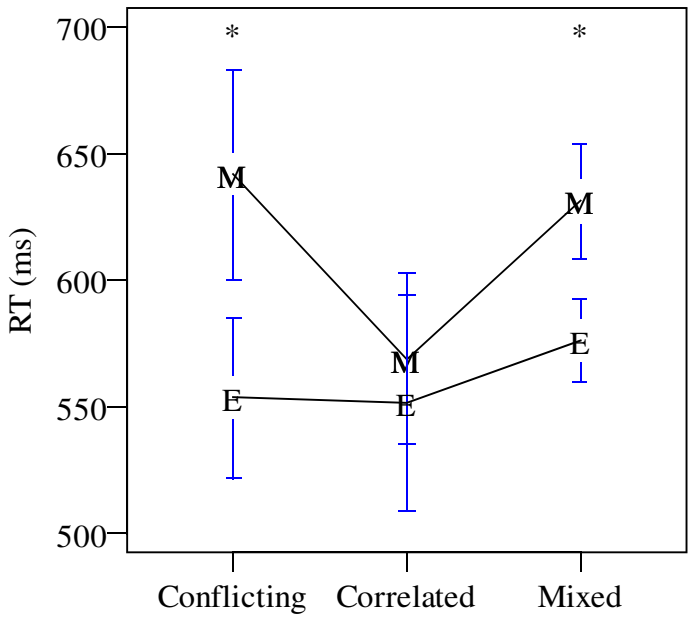

Figure 3.5: Reaction times for discriminating conflicting cue, correlated cue, and mixed conflicting/correlated cue pairs. M= Mandarin listeners, E= English listeners.

The result of this analysis confirms the trends found in the earlier analyses. Mandarin listeners are significantly affected by conflicting cues. They are considerably slower in discriminating pairs that have conflicting cues, whether both cues are conflicting or only one is conflicting, an effect not found with the English listeners. Moreover, they demonstrated similar reaction times to the English listeners only for the correlated cue condition; all suggesting that they are reacting to the conflicting place of articulation information rather than the "different" information present in the signal.

In these results there are clear indications of an effect of native language in this experiment. First, Mandarin listeners show an initial ability to perform well in the discriminations experiment. English listeners, on the other hand, required several blocks before they were able to discriminate at the same level as the Mandarin listeners. Secondly, Mandarin listeners, unlike English ones, were much slower to respond to the pair that differs in both dimensions using conflicting cue tokens. As mentioned earlier, it might be assumed that having both cues differ would speed discrimination as more "different" information is present. This was not the case. English listeners showed no such effect and Mandarin listeners respond similarly to English listeners only when both cues differ and have correlated cues as their reaction times slow to the conflicting information. A second language effect was the reaction time difference between the two groups for pairs varying only in the vocalic dimension. This was somewhat surprising as Mandarin listeners presumably must also use this dimension for categorization and apparently do not ignore it, otherwise they would not be affected in the previously mentioned conflicting cue phenomena as a "different" response can easily be made upon hearing the fricative noise differences. The exact reasons for this effect is unknown.

These language specific results contradict the assumption that a speeded discrimination task bypasses lexical and language effects because it is a low-memory load task (Pisoni 1973, Fox 1984, Johnson 2003, Johnson and Babel 2007). Instead, it supports the view that language specific knowledge is available in trace-coding memory, a view supported by recent studies involving tone (Huang 2004, Krishnan et al. 2005). However, it should also be noted

that English listeners can discriminate using the fricative dimension well above chance (see also Lisker 2001) but apparently have difficultly using this dimension for categorization as demonstrated in Chapter 2, suggesting that the speeded same-different task is tapping information that is not easily accessed at higher memory levels.

3.3 Labeling experiment

This experiment was designed to specifically test categorization and native language effects. In it subjects assigned labels to the entire stimulus grid under limited time pressure. Given the results above and in Chapter 2, it is expected that English listeners will categorize using the vocalic dimension while Mandarin subjects will use both dimensions to categorize.

3.3.1 Materials

All 100 stimuli in the two dimensional [ɕ]/[ʂ] grid (described in Chapter 2) were used in this experiment.

3.3.2 Subjects

The subjects that participated in the discrimination experiment also participated in this experiment. However the data from two English listeners were lost, thus the results of this experiment are based on data from twenty-two English subjects and twenty-two Mandarin subjects.

3.3.3 Procedure

Subjects were asked to label individual sounds from the stimulus set by pressing buttons on a five-button box. For English listeners the leftmost button was labeled "sha" and the rightmost button was labeled "shya", while for Mandarin listeners the leftmost button was labeled "sha" (Pinyin orthography for the voiceless retroflex fricative) and the rightmost button was labeled "xia" (Pinyin orthography for the voiceless alveopalatal fricative.) Each trial consisted of a presentation of a single sound from the stimulus set and the inter-trial interval was 2s. Subjects had up to 2s to respond[5]. Feedback consisted of only reaction time for that trial and the cumulative reaction time up to that point. Subjects were asked to respond in less than 1s. Each block consisted of 100 trials, i.e. a presentation of each stimulus. Subjects participated in 5 blocks for a total of 500 trials. The five main blocks were preceded by five practice trials of random presentations of stimuli.

Following the experiment the Mandarin listeners were asked to fill out a brief questionnaire asking questions about their judgement of the naturalness of the stimuli in comparison to their native sounds. The first question asked whether the person who spoke the sounds was a) "a native speaker of Mandarin", b) "a native speaker of another dialect of Chinese", c) "a native speaker of another language (such as English) trying to speak Mandarin", or d) "a native speaker of another language speaking sounds from their own language". The second and third questions asked them to rate on a scale from 1 to 5 how native-like the "xia" and "shia" sounds were where 1 = "sounded very much like Mandarin" and 5 = "sounded foreign".

5 This was lengthened from the previous experiment to further encourage a more categorical, explicit memory response.

3.3.4 Results and discussion

The English listeners, as in the pilot work, did not categorize using the fricative dimension, only the vocalic dimension, and labeled a larger area of the stimulus space as “sha” than “shya” (Figure 3.6). Generally, subjects were inconsistent in their categorization, with only the extreme ends labeled greater than 66% as “sha” or “shya”.

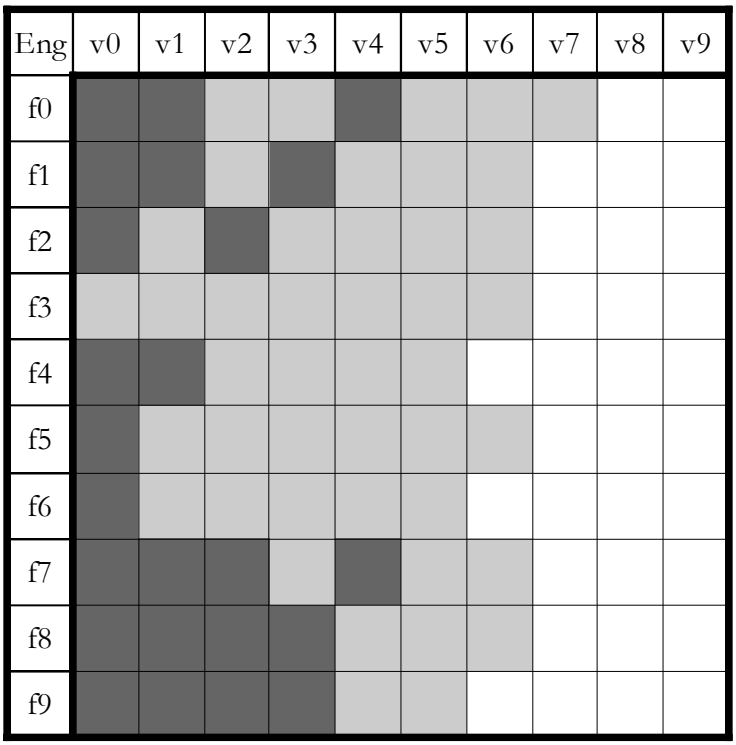

Figure 3.6: Labeling results for English listeners. White squares indicate > 66% labeling as "sha", light gray squares indicate 33%-66% labeling as "sha", and dark squares indicate < 33% labeling as "sha".

However, the Mandarin listeners, like the Polish listener in Chapter 2, did use both dimensions (see Figure 3.7). Like the English listeners, these subjects also labeled a larger area as “sha” rather than the alveopalatal “xia”; however, the area that was inconsistently labeled (33%-66% “sha”) is considerably smaller than that for the English listeners. Further, the regions for each category are much more clearly defined than those for the English listeners.

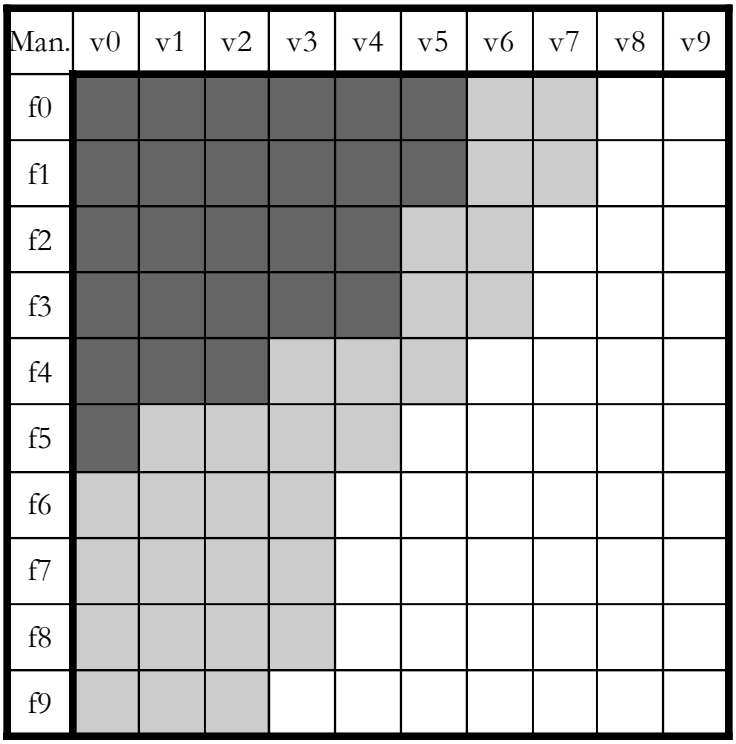

Figure 3.7: Labeling results for Mandarin listeners. White squares indicate > 66% labeling as "sha", light gray squares indicate 33%-66% labeling as "sha", and dark squares indicate < 33% labeling as "sha".

Reaction times greater than 100ms and less than 2000ms were also analyzed. Overall the English listeners were slower than their Mandarin counterparts, having a mean RT of 665ms (σ = 120ms) compared to the Mandarin listeners' mean of 606ms (σ = 154ms). The log transformed reaction times were further analyzed in a repeated measures ANOVA having the factors Language (Mandarin, English) and Cor (correlated cue stimuli, conflicting cue stimuli). Correlated cue stimuli were defined as those in the diagonal connecting f0v0 (the natural alveopalatal) and f9v9 (the natural retroflex) and stimuli in the off diagonal regions of the stimulus space (near the f0v9 and f9v0 corners) were considered to have conflicting cues (see Figure 3.8).

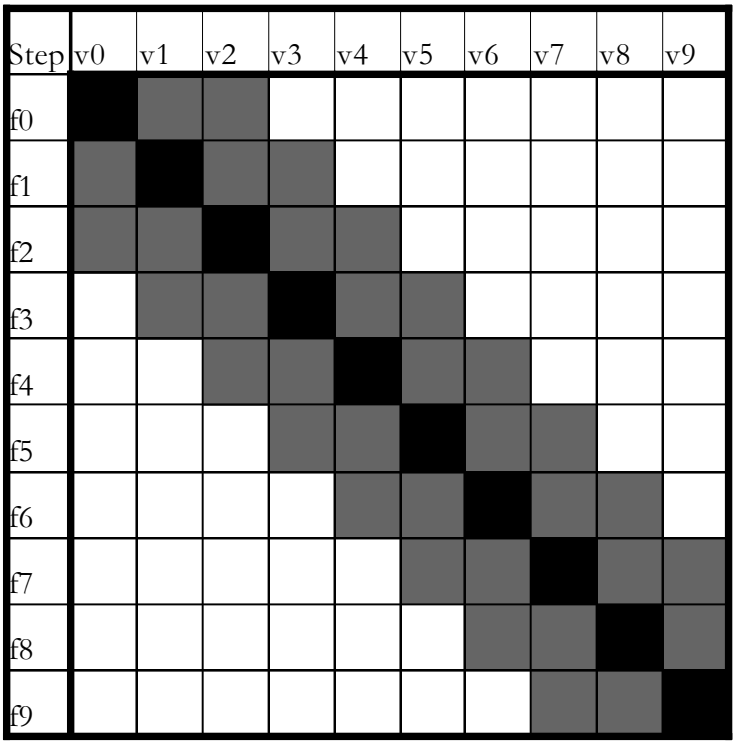

Figure 3.8: Correlated and conflicting cue definitions for the analysis. Stimuli that are considered correlated are in black and gray; black stimuli were fully correlated. White squares were considered conflicting for the analysis.

This analysis yielded a significant main effect for Cor ($F [1,41] = 32.3$, $p < 0.001$) and a significant interaction for Cor and Language ($F [1,41] = 22.1$, $p < 0.001$). Tukey post-hoc tests show significant differences between correlated and conflicting cues only for the Mandarin listeners ($p < 0.001$), Figure 3.9.

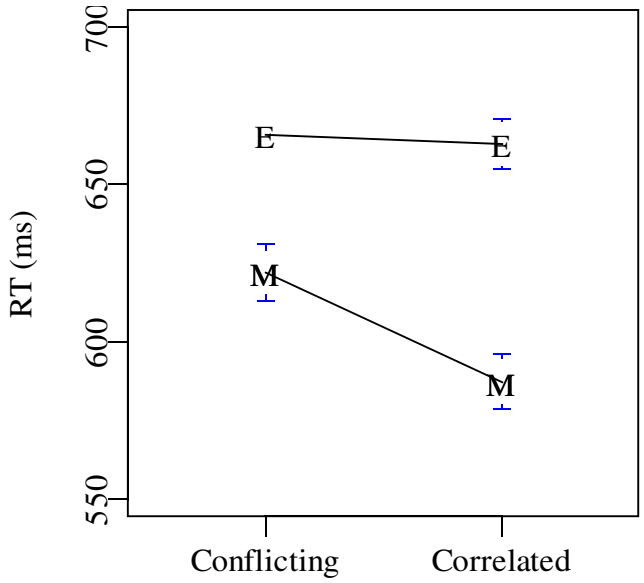

Figure 3.9: Comparison of reaction times for conflicting and correlated cue stimuli for Mandarin (M) and English (E) listeners.

The English listeners show no effect of cue correlation for the stimuli while the Mandarin listeners show a clear effect where the correlated cue stimuli are responded to more quickly. This is most certainly an effect of native language and more evidence that the Mandarin listeners are relying on both cues to categorize these modified syllables.

The post-experiment questionnaire demonstrated no difference between the two categories in rating the naturalness of the sounds as both had a mean value of 2.6. Because raw means may obscure differences in how each subject used the scale, a different measure was calculated. For each subject the difference between the rating for each sound was calculated

by subtracting the "shia" rating from the "xia" rating such that a negative difference indicates "shia" was rated as being more native-like and a positive difference indicates "xia" was rated more native-like. The result of this analysis was an average difference of 0.05, confirming the previous analysis and suggesting that subjects did not as a group rate one sound as more or less native-like than the other. Moreover, the questions about the talker's language resulted in a majority of responses (n=16) in favor of the speaker being a native speaker of another language (not a Chinese dialect) trying to speak Mandarin. A further four subjects responded that the talker was a native speaker of Mandarin and two responded that the speaker was a native speaker of some other dialect of Chinese. No subjects thought these were sounds from a different language. These results generally support the idea that these particular stimuli are easily assimilated to the native Mandarin categories, but that they are not identical to the equivalent Mandarin categories. The exact reason for the differences may be in tone or articulation or both.

Both groups label larger areas as the retroflex "sha". For English listeners a likely explanation can be found in English phonology. The sequence /ʃɑ/ is a legal, though somewhat rare, word in English "Shah", and is present in the onset of many words, e.g. "shot", "shop", "shock". However, the sequence /ʃjɑ/ is unattested as an onset. Thus, the labeling results can be interpreted as a kind of Ganong effect (Ganong 1980). Additionally, the retroflex sound is more like English /ʃ/. For the Mandarin listeners an explanation of the difference is less obvious. As these are not sounds produced by a Mandarin speaker and so are likely subtly different from the phonetic realizations in Mandarin, it is possible that the

retroflex endpoint token was more similar to the native one and therefore more natural (though note that Ladefoged and Maddieson 1996 describe them as being less similar than the two alveopalatals). An alternate possibility is that Mandarin listeners allow for more variability in the retroflex category in natural speech due to lexical or other linguistic factors, shifting the alveopalatal / retroflex boundary. A final possibility is that the manipulation of the fricative duration toward the retroflex may have affected the perception of the two sounds leading to more retroflex responses. Nowak (2006 and p.c.) reported no significant differences in duration between retroflex and alveopalatal and therefore ruled it out as a cue for Polish listeners. However, it is possible that in Mandarin there are consistent differences in fricative duration for this contrast that listeners use to identify the sounds.

3.4 General discussion

Both of the experiments presented in this chapter found language specific patterns in speech perception. Primarily, Mandarin listeners use both fricative noise information and vocalic information to categorize the contrast between retroflex [ʂ] and alveopalatal [ɕ] fricatives.. However, despite using fricative noise information to discriminate sibilants, and having a clear ability to discriminate using the fricative dimension, English listeners do not use this information to categorize this non-native contrast. This manifests itself most clearly in listeners' labeling of the two-dimensional stimulus region; English listeners show a one-dimensional categorization while Mandarin listeners show a two-dimensional one. Further, this difference between the two language groups is also demonstrated in their responses to correlated and conflicting cue stimuli. Specifically, in the discrimination experiment Mandarin are

much slower in discriminating the two conflicting cue stimuli from each other than discriminating the two correlated cue stimuli, unlike the English listeners who are equally fast at discriminating both pairs. This is despite the fact that two pieces of information differ in both pairs, the Mandarin listeners are slowed by the "unnatural" conflicting cue stimuli. Similarly, in the labeling experiment, Mandarin listeners show a much slower reaction time for conflicting cue stimuli as compared to correlated cue stimuli; English listeners show no such effect.

It is not surprising that Mandarin listeners show this correlated~conflicting cue phenomenon. Because these Polish sounds map easily on to the native Mandarin categories they are essentially treated as native by the subjects. As native sounds, Mandarin listeners have a lifetime of experience with the co-articulatory effects that result in the the cues being present on both the consonant and vowel portions of the syllable and are slowed when confronted with unusual combinations of these. However, it is somewhat surprising that English listeners also do not show this effect. Even though English listeners do have experience with sibilant fricative ~ vowel co-articulation effects (e.g. /s/~/ʃ/), this data suggests that they do not generalize that knowledge beyond the contexts in which they experience it natively.

The reasons for English listeners bias towards listening to the vocalic information rather than the fricative information is unknown, though there are several possibilities. One is that the formant transition information is psychoacoustically more robust than the fricative noise information and thus easier for English listeners to use. A second, similar explanation is found in Nittrouer's Developmental Weight Shifting Hypothesis. Under this hypothe-

sis, children focus more on large scale changes in formants resulting from large scale movements in articulators. This view can be expanded to be a more general theory of perceptual acquisition and applied to the English adult listeners described here. Finally, an alternate possibility is found in the results of Wagner et al. (2006) where it was found that languages (like English) which have spectrally similar fricatives in their inventories are more readily able to use formant transition information for identification.

Ultimately, though, it is still an open question whether English listeners can in fact categorize using the fricative dimension. This would involve bringing the knowledge available to discriminate the fricative noise to bear for categorization. Of particular interest is whether this dimension could be used independently of the vocalic dimension for categorization. Training subjects to use this dimension allows for a unique study into cue acquisition, as laid out in Chapter 1.

CHAPTER 4

ENGLISH LISTENERS' PERCEPTUAL LEARNING OF THE POLISH POST-ALVEOLAR SIBILANT CONTRAST

4.1 Introduction

The previous chapter established that English listeners primarily use transition information to differentiate the Polish /ʂ/~/ɕ/ contrast, unlike Mandarin listeners who use transition information as well as fricative noise information to differentiate the two sounds. This difference in behavior seems to be due to the two groups' different language backgrounds; English listeners have no experience with this contrast while Mandarin listeners have extensive experience with an identical or at least extremely similar contrast. These different experiences presumably drive Mandarin listeners to use all available cues in the signal to categorize the two sounds. Such results prompts us to ask whether English listeners can be trained to use fricative information reliably and how training on the new contrast will affect the perceptual categories and their relative space. Essentially, such questions ask whether English listeners can be trained to selectively attend to different information in the signal and how such changes in attention can affect sensitivity to those dimensions.

As discussed in Chapter 1, a key part of perceptual learning is learning to attend to the relevant cues. Children seem particularly sensitive to details that can be used for perception (see e.g. Werker and Tees, 1984; Mareschal et al. 2003) and differing attention to cues affects their categorization (Maye, Werker, and Gerken 2002; French et al. 2004). Specific to speech perception it has been shown that children acquiring their native language shift which cues they rely upon to make relevant distinctions. Specifically a series of studies by Nittrouer and colleagues (e.g. Nittrouer, 1992; Nittrouer and Miller, 1997; Nittrouer, 2002) found that young children generally weight transition cues heavily when discriminating place in sibilant fricatives (/s/ and /ʃ/), gradually reaching adult-like weighting of fricative noise as more informative than transition around 7-8 years old. Nittrouer (2002) demonstrated that for the /f/ ~ /θ/ distinction children predictably used cues like adults, who weightformant transition much more heavily than fricative noise. Nittrouer's general conclusion is that children initially give more attention to the more transparent large-scale dynamic cues rather than static ones.

Although adults show different cue weighting for different phonetic contrasts as a result of the acquisition of these contrasts, this can be modified by training (Francis et al. 2000, Francis and Nusbaum 2002, Francis and Nusbaum 2007). First, Francis et al. (2000) used synthetic stimuli having correlated or conflicting stop burst and formant transitions and trained English listeners with each cue separately. Their subjects demonstrated increased performance with trained cues and decreased performance with untrained cues. A second study by Francis and Nusbaum (2002) trained English listeners to identify the Korean stop con-

trasts and examining subjects' attention to various cues before and after. This study found that subjects' changed their attention to specific cues as necessary; results showed both acquired equivalence and acquired distinctiveness, with the effect determined by the categories. Specifically, they found evidence that contrasts having a high degree of distinctiveness before training show within-category acquired equivalence while difficult contrasts show increases in sensitivity (acquired distinctiveness) to the relevant dimensions.

Similarly, in visual perception, Goldstone (1994) demonstrated that subjects' attention could be directed through simple labeling training to different perceptual dimensions. This study explicitly trained subjects to categorize a two dimensional stimulus set such that one group of subjects had to categorize using a single dimension while ignoring variation in the other, a second group categorized using the previous group's irrelevant dimension, and a third group used both dimensions separately. There were two stimulus sets used for the experiment, one consisted of squares varying by the highly integral dimensions brightness and saturation and the second consisted of the highly separable dimensions size and brightness. Goldstone found that subjects could adequately categorize both stimulus sets using one or both dimensions independently. Importantly, he also found a significant increase in sensitivity to the trained dimension for both stimulus sets and a decrease in sensitivity to the irrelevant dimension, but only for the perceptually separable set.

Although Goldstone (1994) primarily found acquired distinctiveness, other visual studies, notably Livingston, Andrews, and Harnad (1998), have found within-category acquired equivalences. This study used various visual stimuli (e.g. chick genitalia, simulated sin-

gle-celled organisms) that were associated with nonce words and its authors hypothesize that the difference between their results and Goldstone's arises from the comparative ease in discrimination of their stimuli, i.e. it was simpler for subjects to ignore differences rather than heighten them. This argument is similar to that advanced by Francis and Nusbaum to explain different examples of acquired equivalence and distinctiveness in their 2002 study.

The Polish post-alveolar contrast described in this dissertation provides an excellent opportunity to further explore this aspect of perceptual learning. It offers two different dimensions of categorization where one dimension is easy for a specific population (English listeners) to use for categorization and the other is considerably more difficult for the same group. Given the results described above, it is hypothesized that training English listeners to use the fricative noise dimension of the stimulus set used in Chapters 1-3 should result in an increase in sensitivity to that dimension, while training to use the vocalic dimension, which is already highly salient for English listeners, should result in within category compression.

The experiments in this chapter address these issues through an extension of the training experiment described in Chapter 2. Specifically the design of Goldstone (1994) is combined with the stimuli used in the previous experiments to separately train each dimension of the stimulus space. In experiment 1 English listeners were trained to categorize using only the variation in fricative noise spectrum. In experiment 2 English listeners were trained to categorize using only the variation in the vocalic cues. In experiment 3, following Gold-

stone, subjects were trained to use both dimensions separately. A fourth experiment, acting as a control, uses only the testing conditions without training to compare the relative benefit of training to the previous experiments.

4.2 Experiment 1: Fricative dimension categorizers

This experiment was designed to train English listeners to categorize a two-dimensional [ʂa] to [ɕa] continuum using only the fricative noise information. In this experiment the two-dimension stimulus space was divided into two categories along the fricative dimension. Although subjects hear the full range of variation in training, only the fricative dimension was relevant to correct classification.

4.2.1 Materials

The stimulus set is the same as that described in Chapter 2. Discrimination testing used the smaller subset of stimuli (see Figure 4.1).

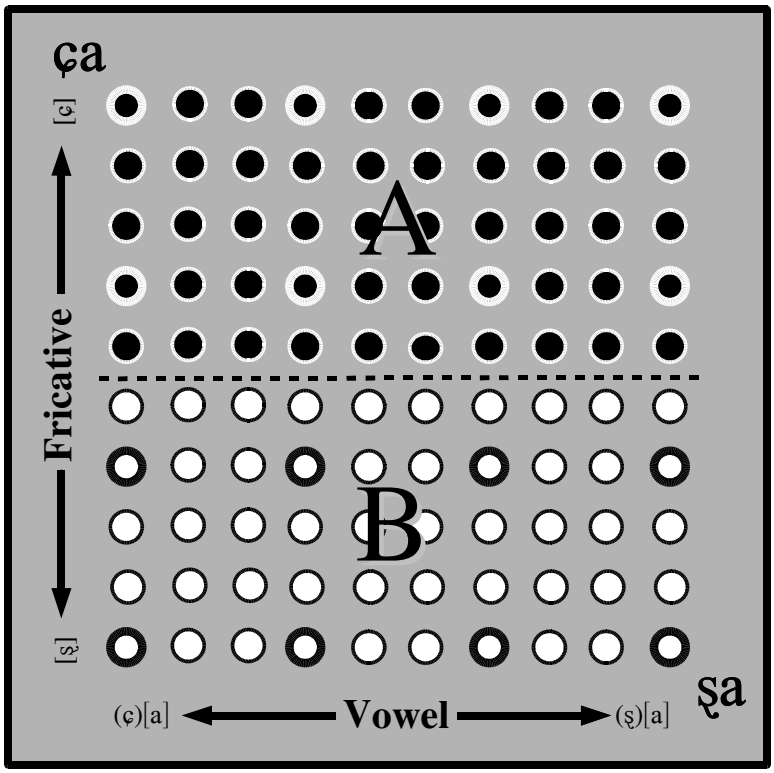

Figure 4.1: Category assignments for the stimuli in experiment 1. Heavy circles denote discrimination stimuli.

4.2.2 Subjects

Twenty-nine male and female subjects between the ages of 18 and 45 participated in the experiment. All were native English speakers with no experience with Polish or Mandarin Chinese. All reported normal hearing and were paid $10 per session.

4.2.3 Procedure

Subjects participated in three to five one-hour sessions held on consecutive days. The first session consisted of a discrimination test followed by labeling familiarization and a labeling test. The stimuli for the discriminating task were drawn from a 4 X 4 subset of the larger stimulus set (see Figure 4.1) where adjacent pairs were separated by three steps in either dimension. Pairs in the center of the fricative distribution were considered cross-boundary tokens and other pairs were considered within-category for the analysis below. During each discrimination trial subjects were presented two pairs of sounds. One pair, the "different" pair, differed along a single dimension and the second pair, the "same" pair consisted of sounds identical to either the first or second sound of the "different" pair. Only adjacent "different" pairs were tested such that the distance between each pair was three steps in a single dimension; larger and smaller distances were not compared. Subjects were asked to press a button on a five-button box corresponding to the "different" pair, either the first pair/leftmost button or the second pair/rightmost button. The two stimuli making up each pair were separated by 100ms while the two pairs were separated from each other by 500ms.

Labeling consisted of a single random presentation of a syllable from the full set of stimuli. Subjects were instructed to categorize each sound as category "A" or "B" by pressing the leftmost or rightmost buttons (respectively). All stimuli with alveopalatal fricatives (fricative steps 0-4) were defined as category "A" while retroflex stimuli (fricative steps 5-9) were category "B". Thus only the fricative dimension was relevant for categorization. Subjects participated in six blocks of labeling and were presented the entire stimulus set during each block. During the first block subjects received accuracy feedback; this was the labeling famil-

iarization period. Positive feedback consisted of the cumulative accuracy score for that session while negative feedback consisted of the correct category label and cumulative accuracy for that session. Subsequent blocks comprised the labeling test period and no feedback was given other than whether or not a response was detected.

Sessions 2-4 comprised the training sessions. These consisted of six labeling blocks with accuracy feedback. Subjects who achieved an overall session accuracy of greater than 85% or completed four sessions moved onto post-testing in the following session. Subjects who showed no major improvement by the middle of the third session were asked about their categorization strategy and encouraged to listen to the "first part" of the sound.

Post-testing consisted of one block of labeling training, five blocks of the labeling testing, and finished with the discrimination test. Subjects were then debriefed and paid for their participation.

4.2.4 Results and discussion

Two subjects dropped out of the study prematurely. The remaining subjects were divided into three groups based upon their labeling performance in the pre- and post-test labeling. The labeling accuracy data for the extreme stimuli in the pre- and post-test labeling, i.e. all stimuli having a fricative component within three steps of each endpoint, were converted in to d' units (Macmillan & Creelman, 2005). Six subjects had a pre-test d' greater than 2.0 and so were not considered to be potential learners[6]. Of the remaining twenty-one subjects, seven did not demonstrate substantial improvement, having a labeling $d' < 1$ on the

6 All of these subjects were asked to return and categorize using the vocalic dimension (see experiment 2). Three agreed and showed equal or even greater accuracy using that dimension.

final session and were classified as "non-learners"[7]. Theses subjects displayed two different patterns, subjects who showed no consistency in categorization and subjects who categorized using the vocalic dimension but were unable to switch to using the fricative dimension. The fourteen remaining subjects all achieved a final labeling $d' > 2.0$, except four who all had a d' score > 1.0. These fourteen subjects were classified as "learners" and their results are analyzed in detail below.

Generally, the learners show a consistent initial categorization using the vocalic dimension but were able to switch to relying solely on the fricative dimension (Figure 4.2), showing a significant improvement in accuracy from pre-test to post-test in a paired *t*-test ($t[13] = -7.6$, $p<0.001$; mean pre-test $d'=0.35$, mean post-test $d'=1.62$).

7 It is possible that these subjects could improve with more training or a different training paradigm. This was not tested.

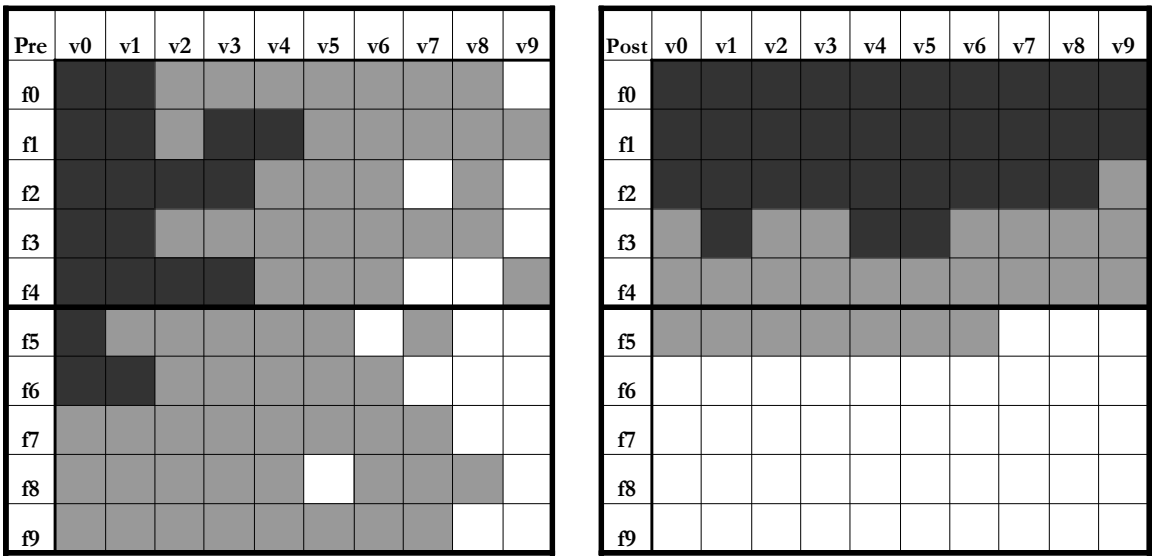

Figure 4.2: Cumulative labeling in pre- (left panel) and post-test (right panel) for all learners. Horizontal dimension represents vocalic continuum and vertical dimension represents fricative dimension. Black filled squares indicate > 67% labeling as "A". Gray squares indicate 33%-67% labeling as "A". White squares represent < 33% labeling as "A".

The discrimination results for learners were also converted to *d′* values. The overall mean d-prime was 0.59, the pre-test mean was 0.46, and the post-test mean was 0.72. These values were considerably lower than those values from the labeling task. This resulted from the task and is due to the small differences in step size necessary to examine within category changes in perception. These were analyzed in a repeated measures ANOVA having the factors Test (pre-, post-), Dimension (fricative, vowel), and Boundary (cross-category, within category). Significant main effects were found for Test ($F[1,14]=15.89$, $p < 0.01$; pre-test d'=0.46, post-test d'=0.72) and Boundary ($F[1,14]=9.43$, $p < 0.01$; cross-category d'=0.70, within category d'=0.54). A significant interaction was found between Test and Dimension

(F[1,14]=6.62, p<0.05, see Figure 4.3). The test and dimension interaction was explored further with Tukey post hoc tests. A significant difference between pre- and post-test *d*-prime scores was found for the fricative dimension only ($p < 0.001$). Additionally, only the post-test d-primes between the two dimensions were significantly different ($p < 0.05$).

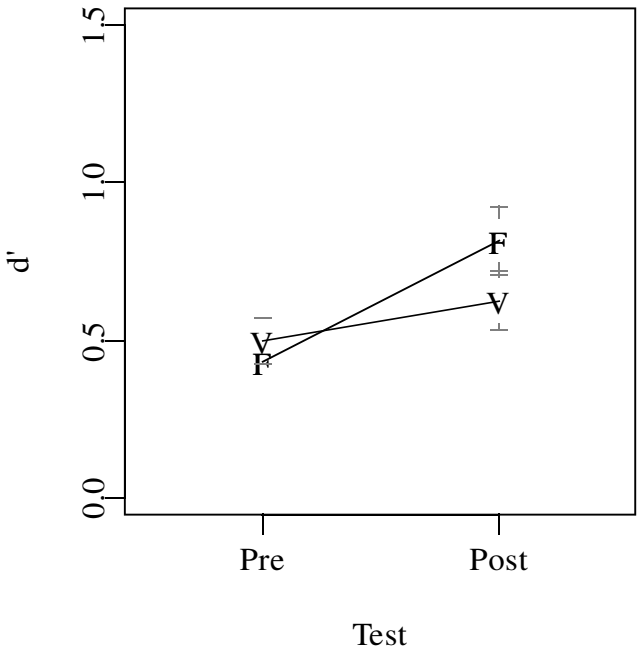

Figure 4.3: Fricative dimension learner's change in sensitivity from pre-test to post-test for the fricative "F" and vocalic "V" dimensions.

These results demonstrate that subjects could learn to rely solely on the fricative noise information for categorization, despite their initial reliance on the vocalic cues. Further, this resulted in a significant increase in sensitivity to the new dimension. There was also a high degree of variability in subject performance, with nearly a quarter of subjects failing to learn and another 20% able to perform at the task immediately with minimal training. The reason for these differences is not immediately clear. No evidence of within-category compression or significant change to irrelevant dimension was found. However, due to the already low sensitivities values, it is possible that a floor effect is present and no further reduction in sensitivity is possible.

These results support those of Goldstone (1994) and Francis and Nusbaum (2002) in showing a significant increase to the trained dimension. If these dimensions are viewed as integral, then the behavior of the irrelevant dimension is as predicted by Goldstone (1994). However, if seen as separable, as the previous chapters suggest, then these results contradict those of Goldstone (1994) who found acquired equivalence for an irrelevant, highly separable dimension. Of course, such a comparison is complicated by the differences in the two experiments. First, this experiment uses speech categories rather than visual ones, so differences may be due to differences in these two modes of perception. Secondly, the two types of stimuli have a different status; in Goldstone's work the stimuli were squares having no particular meaning beyond the experiment, unlike this experiment where the stimuli, especially the retroflex tokens, may have special status as suggested in Chapter 2. Third, and relatedly, Goldstone's stimuli were manipulated such that each pair of adjacent squares was

equally discriminable in each dimension. As the current experiment used "natural" speech categories, this manipulation was not done. In fact, the experiments from previous chapters demonstrate that the vocalic dimension is privileged for English listeners.

4.3 Experiment 2: Vocalic dimension categorizers

This experiment was conducted to direct subjects to categorize along the vocalic dimension. Because American English subject seem to be generally predisposed to categorize this dimension, within-category acquired equivalence might be expected as suggested by the results of previous experiments (Livingston, Andrews, and Harnad, 1998; Goldstone, 1998; Francis and Nusbaum, 2002). Further, as subjects seem to be biased against using the fricative dimension for categorization, subjects should show no change in sensitivity to that dimension or possibly even a decrease in sensitivity.

4.3.1 Methods

The stimuli for this experiment are the same as for the previous experiments. The procedure was also identical to the previous experiment except that the category boundary was defined along the vocalic dimension. Stimuli with vocalic parts closer to the alveopalatal end (v0-v4) were considered category "A" while those closer to the retroflex end (v5-v9) were category "B". Only differences in the vocalic dimension were relevant to categorization (see Figure 4.4).

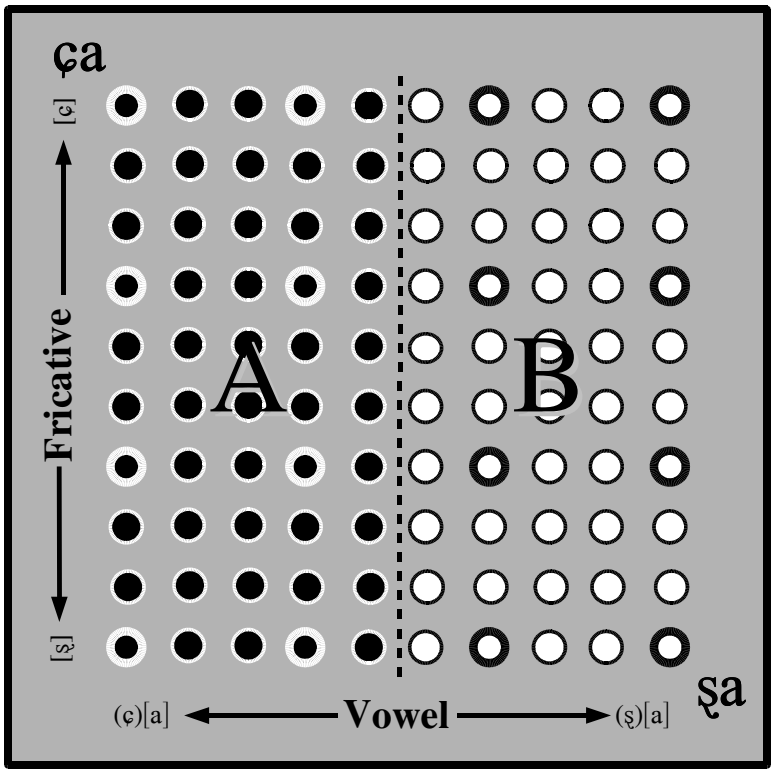

Figure 4.4: Category assignments of the stimuli for experiment 2, vocalic dimension categorizers. Stimuli marked with heavy circles were used in the discrimination task

Fourteen male and female participants between the ages of 18 and 22 participated in the experiment. All reported normal hearing and were paid $10 per session. All were native speakers of English and had no exposure to Polish or Mandarin.

4.3.2 Results and discussion

The labeling results were analyzed as in the previous experiment. For this experiment, twelve of the fourteen subjects had final *d'* values > 2.0 while the remaining two participants were below *d'* = 0.5. Improvement through training for the learners was significant, though small (t [11]= -2.53, $p<0.05$; mean pre-test d'=1.73, mean post-test d'=2.23). As in the first experiment, subjects generally categorized consistently along the vocalic dimension. Here the primary effect of training was a sharpening of the category boundary (Figure 4.5).

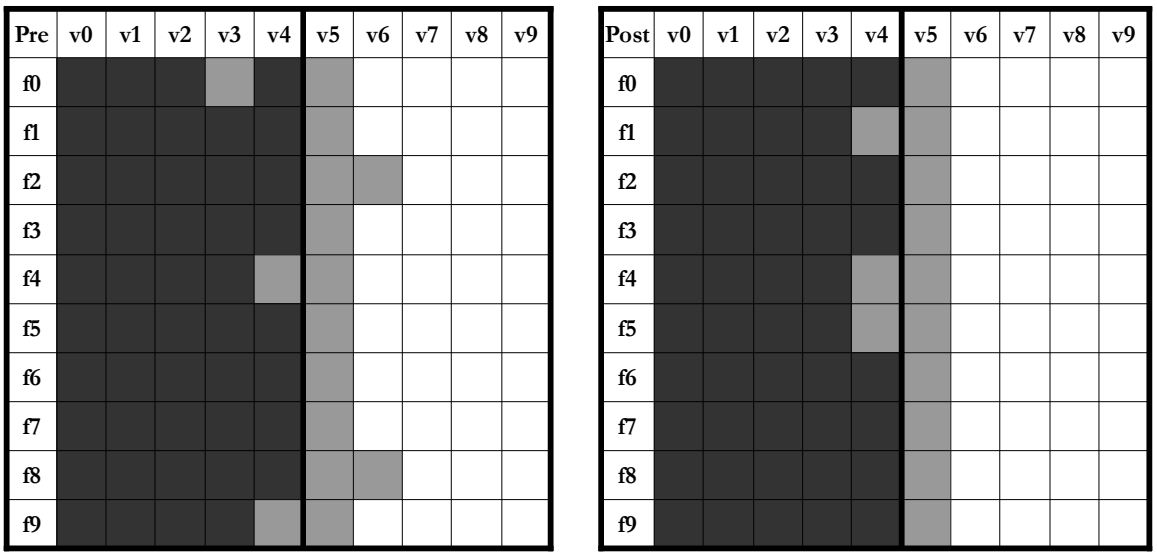

Figure 4.5: Cumulative labeling in pre- (left panel) and post-test (right panel) for all vocalic dimension learners. Horizontal dimension represents vocalic continuum and vertical dimension represents fricative dimension. Black filled squares indicate > 67% labeling as "A". Gray squares indicate 33%-67% labeling as "A". White squares represent < 33% labeling as "A".

The discrimination results were analyzed as in the previous experiment.. The Test factor (see Figure 4.6) did not reach significance ($F[1,11]=4.1$, $p = 0.09$; pre-test $d'=0.59$, post-test $d'=0.76$); the only significant main effect found was for Boundary ($F[1,11]=32.68$, $p < 0.001$; cross-category $d'=0.97$, within-category $d'=0.53$) indicating that subjects were overall more sensitive to cross-boundary tokens. Further, a two-way interaction between Boundary and Test (see Figure 4.7) was found to be significant ($F[1,11]=4.40$, $p < 0.05$). This interaction was explored further with with Tukey post hoc tests which demonstrated significant differences in means between pre-test and post-test for cross-boundary pairs ($p<0.01$) and between cross-boundary and within-boundary pairs for both pre-test ($p<0.05$) and post-test ($p<0.001$).

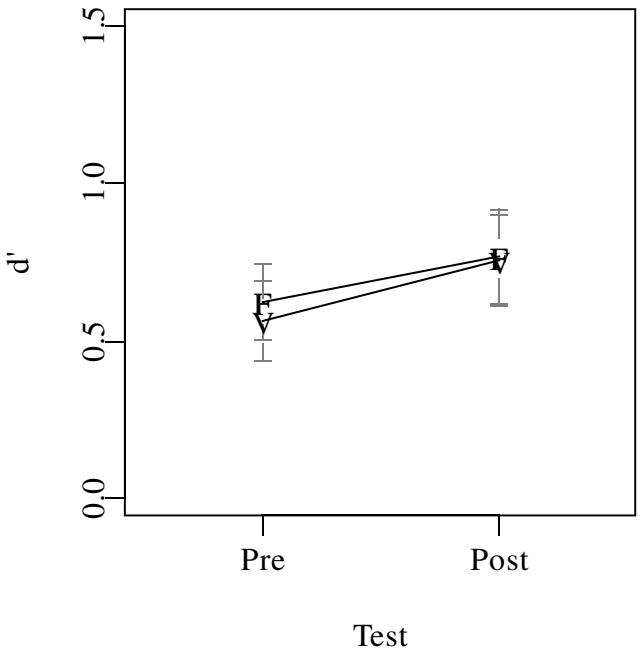

Figure 4.6: Vocalic dimension learner's change in sensitivity from pre-test to post-test for the fricative "F" and vocalic "V" dimensions.

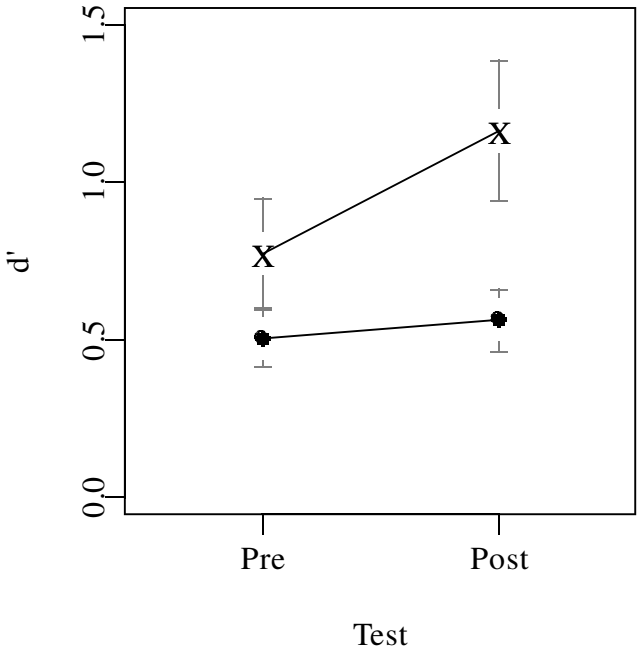

Figure 4.7: Vocalic dimension learner's change in sensitivity from pre-test to post-test cross-category "X" and within category "•".

These results do not support the initial hypothesis as there is no within-category compression. Indeed, there are no significant differences between the two dimensions. Instead, the primary effect of training seems to be a sharpening of the category boundary through heightened sensitivity to the boundary tokens for both dimensions.

4.4 Experiment 3: Two-dimensional categorizers

This experiment is designed so that subjects were required to use both the fricative noise and formant transition dimensions independently. In this categorization task, subjects are asked to divide the stimulus space into four regions with the mid-point along each dimension acting as a category boundary (i.e. each quadrant.) Thus, subjects must use both fricative and vocalic information to make the categorization.

4.4.1 Methods

The stimuli were identical to those used in previous experiments. The procedure is identical to the previous experiment except that the stimulus space was divided into four categories such that participants had to use each dimension independently (see Figure 4.8). Category "A" consisted of alveopalatal fricative and vocalic steps (f0-f4 + v0-v4). Category "B" consisted of alveopalatal fricative steps with retroflex vocalic steps (f0-f4 + v5-v9). Category "C" consisted of retroflex fricative steps with alveopalatal retroflex steps (f5-f9 + v0-v4). Category "D" consisted of retroflex fricative and vocalic steps (f5-f9 + v5-v9). Thus, "B" and "C" are "conflicting cue" categories, while "A" and "D" are "cooperating cue" categories. The four leftmost buttons of the box were labeled "A", "B", "C", and "D" respectively for labeling. Sixteen male and female native speakers of English between the ages of 18 and 25 participated in the experiment. All reported normal hearing and were paid $10 per session. None had exposure to Polish or Mandarin.

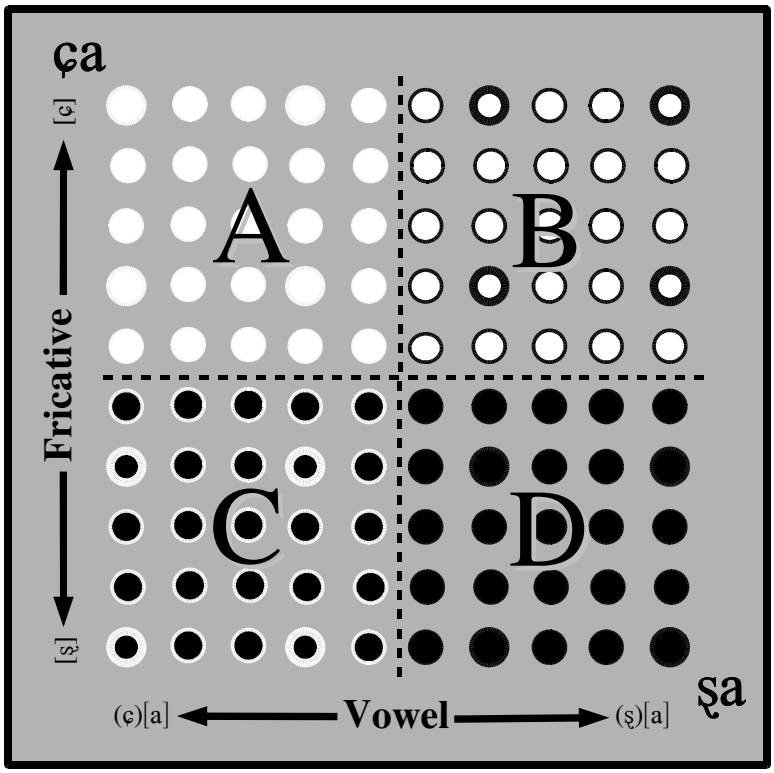

Figure 4.8: Category assignments of the stimuli for experiment 3, two-dimension categorizers.

4.4.2 Results and discussion

Eleven of the sixteen subjects consistently used the category labels in the post-test labeling, defined as an overall estimated $d' > 1$. As would be expected given the previous results, most subjects had difficulty initially with categories divided by the fricative dimension, i.e. A/C and B/D and the primary improvement was in learning to use the fricative dimension for categorization.. However, pooled across subjects, there was some initial consistency in use of the labels with just brief familiarization (see Figure 4.9). Training resulted in a general ability to use both dimensions independently for labeling as indicated by significant im-

provement in labeling from pre-test to post-test (t [10]=-7.42, p<0.001; mean pre-test d'=0.54, mean post-test d'=1.35). Pre-test accuracies for each category were: A=46%, B=46%, C=43%, D=36%. Post-test accuracies for each category were: A=63%, B=71%, C=75%, D=54%.

Pre	v0	v1	v2	v3	v4	v5	v6	v7	v8	v9
f0	AC	A	A	A	AB	AB	AB	B	B	BD
f1	A	A	AB	AB	A	AB	AB	B	B	BD
f2	AB	ABC	A	AC	AC	ABC	B	B	ABD	B
f3	AC	A	A	AC	A	AB	AB	BD	BD	BD
f4	A	AC	AC	AC	ABC	AD	AB	B	BD	BD
f5	C	C	AC	AC	AC	ACD	AC	BD	BD	ABD
f6	AC	AC	AC	C	AC	BC	BD	BD	BD	D
f7	AC	AC	C	C	AC	C	CD	D	D	D
f8	C	C	C	C	C	CD	CD	BD	BCD	D
f9	CD	CD	CD	CD	CD	D	D	D	D	D

Post	v0	v1	v2	v3	v4	v5	v6	v7	v8	v9
f0	A	A	A	A	A	AB	B	B	B	B
f1	A	A	A	A	AB	AB	B	B	B	B
f2	A	A	A	A	A	AB	B	B	B	B
f3	AC	A	A	A	A	AB	B	B	B	B
f4	AC	AC	AC	A	A	AB	B	B	BD	B
f5	AC	AC	AC	AC	AC	A	AB	BD	BD	BD
f6	AC	AC	C	C	C	C	D	BD	BD	BD
f7	C	C	C	C	C	C	D	D	D	D
f8	C	C	C	C	C	C	CD	D	D	D
f9	C	C	C	C	C	C	CD	D	D	D

Figure 4.9: Cumulative labeling in pre- (left panel) and post-test (right panel) for all learners. Horizontal dimension represents vocalic continuum and vertical dimension represents fricative dimension. Each square represents one stimulus and each letter indicates that for a particular stimulus > 25% of responses were for the corresponding category.

There were many different categorization patterns found among the subjects. The non-learners generally demonstrated a two-category division along vocalic dimension, conflating categories “A” and “C” against “B” and “D”. However, the learners also demonstrated a va-

riety of four-category representations (e.g. subjects 203 and 208; see Figure 4.10) and some subjects showed consistent use of only three categories (e.g. subjects 211, 212; see Figure 4.10). The large number of different categorization schemes prohibited further division of the learners into smaller groups.

203	v0	v1	v2	v3	v4	v5	v6	v7	v8	v9
f0	C	C	C	A	AC	AB	B	B	BC	B
f1	C	C	A	C	A	B	B	B	B	B
f2	C	C	AC	C	A	B	B	AB	B	B
f3	C	C	A	A	A	AB	B	B	B	AB
f4	C	A	AC	A	A	AB	B	B	B	B
f5	C	C	C	C	A	A	D	D	B	D
f6	C	C	C	C	C	AC	D	D	D	D
f7	C	C	C	C	C	A	D	D	D	D
f8	C	C	C	C	C	C	D	D	D	D
f9	C	C	C	C	C	CD	D	D	D	D

208	v0	v1	v2	v3	v4	v5	v6	v7	v8	v9
f0	A	A	A	A	A	A	C	B	B	B
f1	A	A	A	A	C	B	B	B	B	B
f2	A	A	A	A	A	A	C	B	B	B
f3	A	A	A	AC	A	A	C	B	B	B
f4	A	A	A	A	D	C	C	B	B	B
f5	AC	C	C	C	C	C	C	BC	BD	B
f6	C	C	C	C	C	C	C	C	D	D
f7	C	C	C	C	C	C	C	D	D	D
f8	C	C	C	CD	C	C	CD	C	D	D
f9	C	C	C	C	C	C	D	CD	D	D

212	v0	v1	v2	v3	v4	v5	v6	v7	v8	v9
f0	C	C	C	C	C	BC	BC	B	D	BD
f1	C	C	C	C	BC	BD	BC	B	B	B
f2	C	C	C	C	C	BC	BD	D	D	BD
f3	C	C	C	C	C	CD	BD	BD	B	B
f4	C	C	C	C	C	BDC	B	D	BD	BD
f5	C	C	C	CD	D	C	B	B	BD	BD
f6	C	C	C	C	C	C	BCD	D	D	CD
f7	C	C	CD	D	C	CD	D	D	D	CD
f8	C	C	C	C	C	D	CD	D	CD	CD
f9	C	C	C	C	CD	D	D	D	D	CD

211	v0	v1	v2	v3	v4	v5	v6	v7	v8	v9
f0	A	A	A	B	A	B	B	B	B	B
f1	A	A	A	A	A	B	B	B	B	B
f2	A	A	A	A	A	B	A	B	B	B
f3	A	A	A	B	A	B	B	B	B	B
f4	A	A	A	A	AB	A	B	B	B	B
f5	A	AC	A	A	A	A	AB	B	B	B
f6	A	C	C	C	AC	A	AC	A	A	B
f7	C	C	C	AC	C	C	AC	A	AC	A
f8	C	C	C	C	C	C	C	C	C	C
f9	C	C	C	C	C	C	C	C	C	C

Figure 4.10: Examples of post-test labeling spaces demonstrating a variety of categorizations of the stimulus space. Subject number is indicated in the upper left corner of each panel. See Figure 4.9 for an explanation of this graphic representation.

The discrimination results (analyzed as before) show significant main effects for Test ($F[1,9] = 10.86$, $p<0.01$; pre-test d'=0.58, post-test d'=1.03, see Figure 4.11) and Boundary ($F[1,9] = 39.95$, $p<0.001$; cross-category d'=1.16, within-category d'=0.63), as well as a significant interaction between Test and Boundary ($F[1,9] = 5.54$, $p<0.05$; see Figure 4.12). Tukey post-hoc tests showed a significant increase in sensitivity from pre-test to post-test for both within-category pairs ($p<0.05$) and cross-boundary pairs ($p<0.001$). The two pair types were significantly different from each other in the post-test ($p <0.001$).

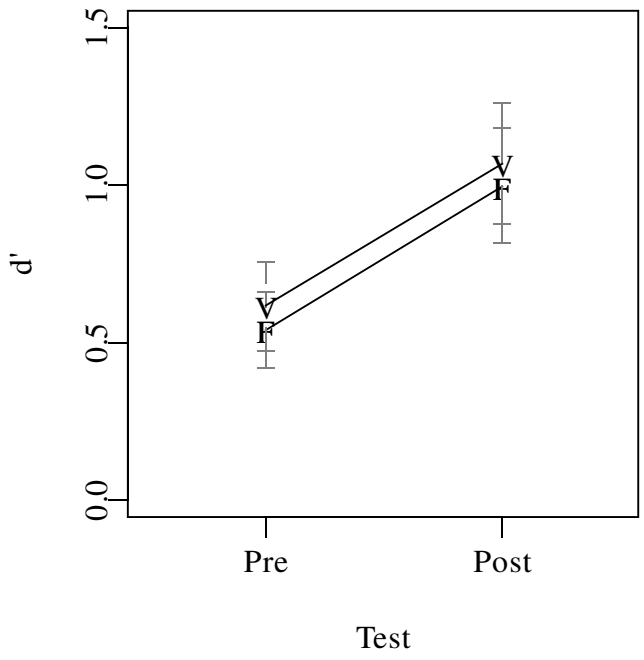

Figure 4.11: Two-dimension learner's change in sensitivity from pre-test to post-test for the fricative "F" and vocalic "V" dimensions.

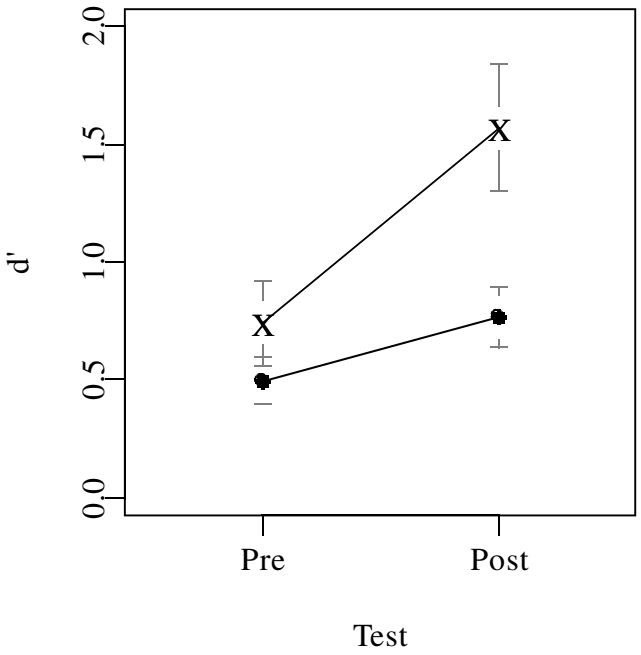

Figure 4.12: Two-dimension learner's change in sensitivity from pre-test to post-test cross-category "X" and within category "●".

Overall, a majority of subjects were able to attend to both dimensions separately and use both for categorization. These results further show that training to use each dimension independently results in a general improvement in sensitivity to both dimensions even

though the primary difficulty in categorization is learning to differentiate along the fricative dimension. Further, this improvement affects cross-category sensitivity and to a smaller extent within-category sensitivity; there is no within-category compression.

These results support the findings of Goldstone (1994) for the two-dimensional learners in that increases in sensitivity were found for both dimensions. However, unlike Goldstone who found the least improvement for two-dimensional learners, this experiment had the largest increases in d' from pre-test to post-test. This discrepancy can be attributed to training length differences in the two studies. In Goldstone's study all subjects, regardless of categorization group, received the same amount of training. In this training regime, on average, listeners in Experiment 3 heard 1800 of training trials, listeners in Experiment 1 heard 600-1800, and listeners in Experiment 2 heard 200. This also may account for the significant improvement in the vocalic dimension in this experiment (and unlike Experiment 2) because subjects trained for a longer period. These effects of training can be further confirmed by examining the performance of subjects who receive no training.

4.5 Experiment 4

This control experiment examined how the exposure to the stimuli in the discrimination tasks alone affected sensitivity, if at all. In this experiment subjects were tested using the same procedure as the previous experiments, but no training or labeling experience of any kind was given. Thus any effects can be attributed to participating in the discrimination tests, which can be compared against performance in the previously described experiments.

4.5.1 Methods

The stimuli were the same as used in previous experiments. This experiment consisted solely of the two discrimination tests, administered on separate consecutive days. No labeling was conducted. Fourteen male and female subjects between the ages of 18 and 23 participated. All reported normal hearing and were native speakers of English with no exposure to Polish or Mandarin. They were paid $10 for their participation.

4.5.2 Results and discussion

The discrimination results were analyzed as in previous experiments. There were no significant effects (Test factor: $F[1,12]=1.84$, $p = 0.2$; pre-test $d'=0.41$, post-test $d'=0.50$). The lack of a test effect suggests that the improvement seen in the previous experiments from pre- to post-test can be safely assumed to be due to the training, rather than simply exposure to the stimuli.

4.6 Post-hoc analysis

As discussed earlier, there was no attempt to standardize the perceptual distances between the discrimination pairs; instead the natural distances were maintained. This is in contrast to Goldstone's study in which a separate experiment was conducted to establish uniform sensitivities with and across each dimension before training. Because of this, it is useful to know the extent to which subject's initial sensitivities to the Polish stimuli used in these experiments was biased or “warped”. In order to explore this, the following post-hoc analysis of the pre-test data was conducted to explore subjects' perception of the stimulus space.

4.6.1 Methods

The pre-test discrimination data from the previous four experiments was combined and analyzed as a whole. All subjects (n=79) were included regardless of performance on later training and post-test conditions. The stimulus pairs were classified as before by dimension, but also by region within the stimulus space and whether the two dimensions were primarily correlated or conflicting.

4.6.2 Results

The results were converted to *d′* units as before; however the data was recoded. In this analysis each dimension was divided into three regions with the step 0-3, 3-6, and 6-9 pairs collapsed across each dimension. Specifically for each dimension it is assumed for the analysis that 0-3 represents alveopalatal within category discrimination, 3-6 represents cross-category discrimination, and 6-9 represents within category discrimination for retroflex cues. Further all pairs were divided into two groups, those that had correlated fricative and vocalic cues and those where the two cues were in conflict. A repeated measures ANOVA with the factors Dimension (fricative, vocalic), Region (alveopalatal, boundary, retroflex), and Correlation (correlated, conflicting) found a significant main effect for Region ($F[2]= 32.83$, $p < 0.001$) and a significant interaction between Dimension and Region ($F[2]=5.18$, $p < 0.01$). The Dimension by Region interaction is shown in Figure 4.13. Subsequent Tukey tests of means showed that the regions differ by dimension only for the retroflex cue ($p <0.01$). The alveopalatal and boundary regions differed significantly for both dimensions (fricative:

p<0.001; vocalic: p<0.001) and the retroflex region differed from the boundary only for the fricative dimension (p<0.01). The difference between the alveopalatal and retroflex regions was significant for both dimensions (fricative: p < 0.05; vocalic: p < 0.001).

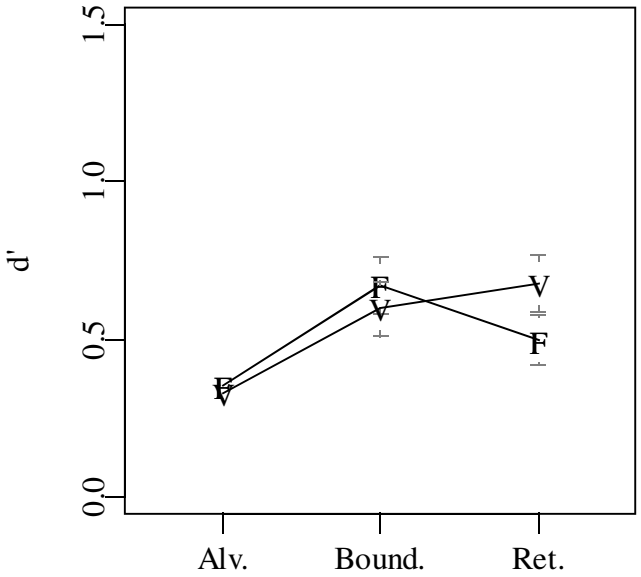

Figure 4.13: Sensitivity to each region of the stimulus space for the fricative "F" dimension and vocalic "V" dimension for all subjects' discrimination pre-tests.

This analysis confirms that English listeners do not perceive the stimulus space uniformly. They are least sensitive to differences within the alveopalatal region for both dimensions. For the fricative dimension, subjects are most sensitive to cross-boundary pairs while the retroflex pairs are the most easily discriminated in the vocalic dimension. No difference between discrimination of stimuli having correlated or conflicting cues was found, as expected given the results detailed in Chapter 3.

For both dimensions the high degree of sensitivity to retroflex tokens is the likely cause of the larger proportion of responses as retroflex found in the previous chapter. It is therefore possibly due to the language effects discussed in the previous chapters, i.e. subjects are behaving as if the retroflexes are English /ʃ/, or some general property of the stimuli. Additionally, even though the previous experiments demonstrate that English listeners show a bias towards using the vocalic dimension for categorization, this does not seem to be borne out by the results in this analysis generally as the listeners are more sensitive to the vocalic dimension only for the retroflex ends of the continua.

4.7 General discussion

The experiments in this chapter demonstrate that English listeners could learn to differentiate the Polish alveopalatal ~ retroflex sibilant contrast and do so using either fricative noise cues or vocalic cues. The first experiment demonstrated that English listeners were able to overcome their initial tendency to notice the vowel formant dimension and attend to the fricative noise information. This supports the finding of other researchers that subjects'

attention can be directed differentially to facilitate categorization of a new contrast (Goldstone, 1994; Francis and Nusbaum 2002). This direction of attention heightened subject's sensitivities to the dimension of categorization for all but the vocalic learners.

Overall, there is only evidence for acquired distinctiveness. For fricative learners and two-dimensional learners there are increases in sensitivity to the dimension of categorization without corresponding within category compression. In fact, for two-dimensional learners there is clear within category expansion in addition to cross-category expansion. Such results generally support the findings of previous research (e.g. Goldstone 1994; Francis and Nusbaum 2002). However, the extreme difficulty of the discrimination task left little room for a reduction in sensitivity and it is possible the floor effects negate any acquired equivalence that may be present. In fact, it was hypothesized that the vocalic learners would show compression due to the ease of categorization using that dimension; instead there was no significant change. Similarly, it might be assumed that for the fricative categorizers the irrelevant vocalic dimension would show a loss in sensitivity. Though, again there was no effect.

Moreover, although listeners could easily label using the vocalic dimension and had difficulty using the fricative dimension, this pattern was not fully present in the discrimination data; only for retroflex cues was this true. This is a puzzling result. Quite possibly, however, this is a task related effect as the discrimination paradigm was necessarily highly sensitive, having short interstimulus intervals, in order to assess fine-grained within-category changes rather than broad category level changes in sensitivity (as might be shown by an ABX paradigm, for example.) This is supported by the results of Guenther, Husain, Cohen,

and Shinn-Cunningham (1999) where a within category compression effect was found only when a long ISI, categorical discrimination task was used, as opposed to a short ISI discrimination task which revealed no effect of training.

However, it is also of interest that warping of the perceptual space was found using such a highly sensitive psychoacoustic task. Like the results found using a similar discrimination task in Chapter 3, these subjects showed a warping of the perceptual space at a very low level of processing. This suggests that exposure to a contrast and training using a contrast has an effect at very low as well as the higher cognitive levels demanded by the labeling task.

As described in the previous chapters, English listeners do not seem to be affected by the degree of correlation between the fricative and vocalic cues. The results of these experiments confirm this. Most directly, subjects were equally sensitive to correlated and conflicting cue stimuli as demonstrated by the post-hoc analysis. Additionally, in the two-dimensional categorization experiment, subjects demonstrated no more difficulty in learning the conflicting cue categories composed of physically different chunks ("B" and "C") than the "natural" correlated categories ("A" and "D"). In fact subjects showed the least accuracy in labeling the correlated retroflex category "D" and assigned it the least area in the stimulus space. That subjects relatively easily learned to categorize speech stimuli that could be impossible to produce, and not physically generated by this particular talker's vocal tract, calls into question theories of speech perception that rely on individuals perceiving gestures (e.g. Liberman et al. 1967, Fowler 1986, Best 1995).

CHAPTER 5

CONCLUSION

5.1 Summary of results and discussion

It was argued in the introduction that during acquisition layers of experience and knowledge resulted in the perceptual abilities seen in adults and that general perceptual learning mechanisms are operant both in children and adults. As shown with the Mandarin listeners, their experience with their native language affects and warps perception at even very low levels and guides what they attend to in the speech signal. Similarly, the native English listeners, having a different body of experience, showed a distinctly different pattern of warping of the perceptual space and cue use. The English listeners' perceptual space and cue use was subsequently modified through training such that perceptual space was warped in a new way. This resulted in a heightened sensitivity (even at a very low cognitive level) to the stimuli and dimensions of categorization.

More specifically, the experiments detailed in this dissertation demonstrate several other important findings. First, English listeners can distinguish the Polish post-alveolar sibilant contrast. They do this primarily by relying on formant transition information and show

no use of the variation in fricative noise that native listeners use and that English listeners themselves use to distinguish native English sibilants. This is contrasted against Mandarin listeners who can also differentiate these sounds but demonstrate a more robust categorization using both fricative noise information as well as vocalic information. In the case of Mandarin listeners, they assimilate the two sounds to their own native categories which are extremely similar. In Chapter 3 it was further argued that Mandarin listeners' native experience with these sounds allows them to use both of these pieces of information. A ramification of this is that Mandarin listeners also show and effect of integrality of the two cues that English listeners do not. This is despite the fact that English listeners do have experience with co-articulation effects in sibilants in their native language. This suggests that perceptual integration effects are learned from experience for specific native contrasts and not generalized to new cases. This integrality is a unitization phenomenon that presumably originates early in acquisition and offers evidence that much perceptual knowledge is heavily tied to very specific linguistic contexts. Moreover, as with the results of Nittrouer and colleagues' work with English learning children, the Mandarin listeners have developed an ability to use the most relevant cues for identifying their native categories. American English listeners, not having this sibilant contrast, have not learned to use the full range of cues available to differentiate it.

The results from Chapter 3 also pose a methodological question. In that chapter, the discrimination experiment found an effect of native language. This is in contrast to other experiments where native language effects are not present in short ISI AX (same-different) dis-

crimination experiments, but often are in more categorical, high-level knowledge tasks. The results from this study imply that some language-specific knowledge is present at very low-level auditory processing, in this case knowledge of co-articulation. Moreover, the 4IAX task used in the training experiment in Chapter 4 is also considered a highly sensitive method similar to short ISI AX discrimination and effects of relatively brief training were present in that data, also.

The results from these studies comparing Mandarin and English listeners were followed up by a training study which demonstrated that English listeners can use the fricative dimension for categorization if given sufficient training and directed attention. Labeling training was generally sufficient to direct attention to the fricative dimension and away from the vocalic dimension. Such results are supportive of several studies on selective attention, e.g. Goldstone (1994) and Francis and Nusbaum (2002).

This training study also found evidence for acquired distinctiveness for stimuli in a newly-acquired dimension of categorization, but was not able to find any acquired equivalence as a result of category learning. The lack of acquired equivalence, even for irrelevant variation (which subjects had previously found to be useful), is surprising. One possibility is that the difficulty of the task resulted in a floor effect, masking any sensitivity reduction. Similarly, it is also possibility that subjects never acquired enough skill at making the distinction and, as suggested by other studies, acquired equivalence was never in operation. Another possibility is that the flat distribution of the training stimuli rather than a multi-modal dis-

tribution (a la Maye et al. 2002) and use of discrimination for testing served to only highlight differences in the stimuli, rather than similarities. Future work may hold the key to further elucidating this question.

5.2 Future work

The work presented in this dissertation provide several questions that can be addressed in future work. One clear question is whether English listeners can learn to use both cues of Polish contrast integrally and exhibit behavior like the Mandarin listeners in the labeling experiment in Chapter 2. Research from perceptual learning (e.g. Livingston et al.1998, Yamada and Tohkura 1992, Francis and Nusbaum 2002) indicates that indeed, subjects will attend to new dimensions when necessary. A question for this particular contrast is whether English listeners will be able to learn to use the fricative dimension when the vocalic cues are already so robust for them. That is, can listeners learn to use largely redundant cues? Similarly, the extent to which Polish and/or Mandarin listeners can use the dimensions separately is also unknown, however evidence from the selective attention literature (e.g. Francis et al. 2000) suggests that it is quite possible for listeners to attend to one cue and ignore other, competing cues.

The training study attempted to explore the change in perceptual space resulting from perceptual learning. That study only found expansion of the perceptual space, leaving the question open as to whether acquired equivalence can be induced. Several modifications of the study can address some of the possible shortcomings of that experiment. One option would be to make the discrimination task easier by increasing the step size of "different"

pairs. This would not only make the task easier and result in more robust results, but also remove the potential floor effect. Similarly, the discrimination task could be changed to a less sensitive and more categorical measure, such as ABX discrimination or a rating task.. A third modification would be to change the distribution of the stimuli in training such that subjects maximally hear endpoints of a category and minimally hear ambiguous central tokens. Such an experiment would provide a replication and extension of Maye 2000.

Another issue unanswered by the training task is the extent to which learning in the experiment is generalizable to new situations. For example, would subjects who learn to use the fricative cues also be able to use fricative noise cues to categorize the voiced counterparts of the sibilants in question? Could such knowledge even be extended to the Polish affricates made at the same place of articulation?

Finally, one aspect that has been sorely neglected is the extent to which production affects perception and how that knowledge fits in to this research. For example, it is possible that the wide variety of sibilant production strategies available to speakers and variation in production across languages with different inventories affects perception (see Toda and Honda 2005). The extent to which such variation affects the subjects in these experiments (including the wide variety in performance of English listeners in the training experiment) is unknown.

LIST OF REFERENCES

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

Bertoncini, J., Floccia, C., Nazzi, T., & Mehler, J. (1995). Morae and syllables: Rhythmical basis of speech representations in neonates. *Language and Speech*, *38*, 311-329.

Best, C. T., Morrongiello, B., & Robson, R. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception & Psychophysics*, *29*, 191-211.

Best, C. T., McRoberts, G. W., & Sithole, N. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human perception and performance, 14,* 345-360.

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America, 109* (2), 775-794.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999) Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics, 61*(5), 977-985.

Case, P., Tuller, B., Kelso, J. A. S. (2003, November 17). The dynamics of learning to hear new speech sounds. *Speech Pathology*.

Dell, G. S., Reed, K. D., Adams, D. R., & Meyer, A. S. (2000). Speech errors, phonotactic constraints, and implicit learning: a study of the role of experience in language production. *Journal of Experimental Psychology: Learning, Memory and Cognition, 26*(6), 1355-1367.

Edwards, J., Fox, R., & Rogers, C. (2002). Final consonant discrimination in children: Effects of phonological disorders, vocabulary size and phonetic inventory size. *Journal of Speech, Language and Hearing Research, 45*, 231-242.

Eimas, P. D., Siqueland, E. R., Jusczyk, P. W., & Vigorito, J. (1971). Speech perception in infants. *Science, 171*, 303-306.

Eisenberg, L. S., Shannon, R. V., Martinez, A. S., Wygonski, J., & Boothroyd, A. (2000). Speech recognition with reduced spectral cues as a function of age. *The Journal of the Acoustical Society of America, 107(5)*, 2704-2710.

Fox, R. A. (1984). Effect of lexical status on phonetic categorization. *Journal of Experimental Psychology: Human perception and performance*, *10*(4), 526-540.

Francis, A. L., Baldwin, K., & Nusbaum, H. C. (2000). Effects of training on attention to acoustic cues. *Perception & Psychophysics, 62*(8), 1668-1680.

Francis, A. L., & Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance, 28*(2), 349-366.

French, R. M., Mareschal, D., Mermillod, M., & Quinn, P. C. (2004). The role of bottom-up processing in perceptual categorization by 3- to 4-month-old infants: Simulations and data. *Journal of Experimental Psychology: General, 133*(3), 382–397.

Garner, W. R. (1974). *The processing of information and structure*. Hillsdale, NJ: Erlbaum.

Gibson E.J. (1991). *An Odyssey in Learning and Perception.* Cambridge, MA: MIT Press

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning Memory and Cognition, 22*(5), 1106-1183

Goldstone, R. L. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General, 123*(2),178-200.

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology, 49*, 585-612.

Grau, J. W., & Kemler Nelson, D. G. (1988). The distinction between integral and separable dimensions: Evidence for the integrality of pitch and loudness. *Journal of Experimental Psychology: General, 117*(4), 347-370.

Guenther, F. H., Husain, F. T., Cohen, M. A., & 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.

Guion, S.G. (1998). The role of perception in the sound change of velar palatalization. *Phonetica,* 55, 18-52.

Guion, S. G., & Pederson, E. (2007). Investigating the role of attention in phonetic learning. In O. S. Bohn & M. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 57-77). Amsterdam: John Benjamins.

Harris, K. S. (1958). Cues for the discrimination of American English fricatives in spoken syllables. *Language and Speech, 1*, 1-7.

Harnsberger, J.D. (2001). The perception of Malayalam nasal consonants by Marathi, Punjabi, Tamil, Oriya, Bengali, and American English listeners.: A multidimensional scaling analysis. *Journal of Phonetics*, 29, 303-27.

Hawkins, S., & Smith, R. (2001). Polysp: a polysystemic, phonetically-rich approach to speech understanding. *Italian Journal of Linguistics-Rivista di Linguistica*, *13*(1), 99-188.

Hazan V., & Rosen S. (1991). Individual variability in the perception of cues to place contrasts in initial stops. *Perception & Psychophysics, 49*(2), 187-200.

Hollich, G., Jusczyk, P. W., & Luce, P. A. (2002). Lexical neighborhood effects in 17-month-old word learning. *Proceedings of the 26th Annual Boston University Conference on Language Development* (pp. 314-23). Boston, MA: Cascadilla Press.

Huang, T. (2001). The interplay of perception and phonology in Tone 3 in Chinese Putonghua. In E. Hume and K. Johnson (eds.) *Studies on the Interplay of Speech Perception and Phonology, Ohio State University Working Papers in Linguistics.* 55, 23-42.

Jamieson, D. G., & Morosan, D. E. (1986). Training non-native speech contrasts in adults: Acquisition of the English /θ/-/ð/ contrast by francophones. *Perception & Psychophysics*, *40*(4), 205–215.

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson & J. W. Mullennix (Eds.), *Talker Variability in Speech Processing* (pp. 145-165). San Diego, CA: Academic Press.

Johnson, K. (2004). Cross-linguistic perceptual differences emerge from the lexicon. In *Proceedings from the 8th Annual Texas Linguistics Society, (TLS-03).* Somerville, MA: Cascadilla Press.

Johnson, K. & Babel, M. (2007). Perception of Fricatives by Dutch and English Speakers. *UC Berkeley Phonology Lab Annual Report*, 302-325.

Kemler Nelson, D. G. (1993). Processing integral dimensions: The whole view. *Journal of Experimental Psychology: Human Perception and Performance, 19*(5), 1105-1113.

Kuhl, P. K. (1991). Human adults and infants show a 'perceptual magnet effect' for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics, 50*(2), 93-97.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science, 255*, 606-608.

Ladefoged, P. & Wu, Z. (1984). Places of articulation: An investigation of Pekingese fricatives and affricates. *Journal of Phonetics, 11*, 291-302.

Ladefoged, P. & Maddieson, I. (1996). *The Sounds of the World's Languages*. Oxford: Blackwell.

Lambacher, S., Martens, W., Nelson, B., and Berman, J. (2001). Identification of English voiceless fricatives by Japanese listeners: the influence of vowel context on sensitivity and response bias. *Acoustic Science & Technology*, 22, 334-43.

Li, F. (2005, June). An Acoustic Study on Feminine Accent in Beijing Dialect --- Through a Comparison between Beijing Dialect and Songyuan Dialect. Paper presented in *The 17th North American Conference on Chinese Linguistics*, Monterey, CA.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54*(5), 358-368.

Lisker, L. (2001). Hearing the Polish sibilants [s š ś]: Phonetic and auditory judgements. In N. Grønnum, & J. Rischel (Eds.), Travaux du Cercle Linguistique de Copenhague XXXI. To honour Eli Fischer-Jørgensen (pp. 226–238). Copenhagen: C.A. Reitzel.

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y. & Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/: III. Long term retention of new phonetic categories. *Journal of the Acoustical Society of America, 96*(4), 1242-1255.

Livingston, K.R., Andrews, J., & Harnad, S. (1998). Categorical perception effects induced by category learning. Journal of Experimental Psychology: Learning, Memory, and Cognition. 24:3, 732-753.

Logan GD. 1988. Toward an instance theory of automatization. *Psychological Review*, 95:492-527

Logan, J.S., S.E. Lively, & D.B. Pisoni (1991). Training Japanese listeners to identify English

/r/ and /l/: a first report. The Journal of the Acoustical Society of America. 94, 1242-1255.

Lotto, A.J., K.R. Kluender, & L.L. Holt (1988). Depolarizing the perceptual magnet effect. Journal of the Acoustical Society of America, *103*, 3648-3655.

McCandliss, B., Fiez, J., Protopapas, A., Conway, M., & McClelland, J. L. (2002). Success and failure in teaching the [r]-[l] contrast to Japanese adults: Tests of a Hebbian model of plasticity and stabilization in spoken language perception. *Cognitive, Affective, & Behavioral Neuroscience, 2*(2), 89-108.

Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology, 38*, 465-94.

Mareschal, D., Powell, D., & Volein, A. (2003). Basic-level category discriminations by 7- and 9-month-olds in an object examination task. *Journal of Experimental Child Psychology, 86*, 87-107.

Mayo, C., & Turk, A. (2005). The influence of spectral distinctiveness on acoustic cue weighting in children's and adults' speech perception. *The Journal of the Acoustical Society of America, 118*(3), 1730-1741.

Maye, J. (2000). *Learning Speech Sound Categories From Statistical Information*. Unpublished doctoral dissertation, Arizona State University.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82*, B101-B111.

Mielke, J. (2003). The interplay of speech perception and phonology: Experimental evidence from Turkish. *Phonetica, 60*, 208-229.

Miller, G. A. and Nicely, P. A. (1955). An analysis of perceptual confusions among some English consonants. *The Journal of the Acoustical Society of America. 27*, 338-352.

Mills, D. L., Prat, C., Zangl, R., Stager, C. L., Neville, H. J., & Werker, J. F. (2004). Language experience and the organization of brain activity to phonetically similar words: ERP evidence from 14- and 20-month-olds. *Journal of Cognitive Neuroscience, 16*, 1452-1464.

Nazzi, T., Jusczyk, P. W., & Johnson, E. K. (2000) Language discrimination by English-learning 5-month-olds: Effects of rhythm and familiarity. *Journal of Memory and Language, 43*, 1-19.

Nittrouer, S. (1992). Age-related differences in perceptual effects for formant transistions

within syllable and across syllables. *Journal of Phonetics, 20*, 1-32.

Nittrouer, S. (2002). Learning to perceive speech: How fricative perception changes, and how it stays the same. *The Journal of the Acoustical Society of America. 112*(2), 711-719.

Nittrouer, S. (2006). Children hear the forest. *The Journal of the Acoustical Society of America. 120*(4), 1799-1802.

Nittrouer, S. & Miller, M. E. (1997). Predicting developmental shifts in perceptual weighting schemes. *The Journal of the Acoustical Society of America, 101(*4), 2253-2266.

Nittrouer, S. & Studdert-Kennedy, M (1987). The role of coarticulatory effects in the perception of fricatives by children and adults. *Journal of Speech and Hearing Research, 30*, 319-329.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115*(1), 39-57.

Nowak, P. (2006). The role of vowel transitions and frication noise in the perception of Polish sibilants. *Journal of Phonetics, 34*, 139-152.

Onishi, K. H., Chambers, K. E. & Fisher, C. (2002). Learning phonotactic constraints from brief auditory experience. *Cognition, 83*, B13-B23.

Papouek, M., & Papouek, H. (1989). Forms and functions of vocal matching in interactions between mothers and their precanonical infants. *First Language, 9*, 137-158.

Patterson, M. L., & Werker, J. F. (2002). Infants' ability to match dynamic phonetic and gender information in face and voice. *Journal of Experimental Child Psychology, 81*, 93-115.

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perceptual Psychophysics*, *13*, 253-260.

Pisoni, D. B., Aslin, R. N., Perey, A. J., & Hennessy, B. L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stopc onsonants. *Journal of Experimental Psychology: Human Perception and Performance, 8*, 297-314.

Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 421-435.

Pruitt, J. S., Strange, W., Polka, L., & Aguilar, M. C. (1990). Effects of category knowledge and syllable truncation during auditory training on American's discrimination of Hindi retroflex - dental contrasts. *The Journal of the Acoustical Society of America, 87*(Supplement

1), S72.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1926-1928.

Schacter, D.L. (1987). Implicit memory: history and current status. *The Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 501-18.

Stevens, K. N., Li, Z., Lee, C.-Y., & Keyser, S. J. (2004). A note on Mandarin fricatives and enhancement. In G. Fant, H. Fujisaki, J. Cao, and Y. Xu (Eds.), *From Traditional Phonology to Modern Speech Processing* (pp. 393-403). Beijing: Foreign Language Teaching and Research Press.

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review, 64*, 153–181.

Strange, W., & Dittman, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception & Psychophysics, 36*(2), 131-145.

Terbeek, D. (1977). A cross-language multidimensional scaling analysis of vowel perception. *UCLA Working Papers in Phonetics,* 37, 1-271.

Tomiak, G. R., Mullenix, J. W., & Sawusch, J. R. (1987). Integral perception of phonemes: Evidence for a phonetic mode of perception. *The Journal of the Acoustical Society of America 81*(3), 755-781.

Tsao, F.-M., Liu, H.-M., & Kuhl, P. (2004). Speech perception in infancy predicts language development in the second year of life: A longitudinal study. *Child Development, 75*, 1067-1084.

Wagner, E., & Cutler, A. (2006). Formant transitions in fricative identification: The role of native fricative inventory. *The Journal of the Acoustical Society of America, 120*(4), 2267-2277.

Werker, J. F. & Curtain, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development, 1*(2), 197-234.

Werker, J. F., & Tees, R. C. (1984a). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7*, 49-63.

Werker, J. F., & Tees, R. C. (1984b). Phonemic and phonetic factors in adult cross-language speech perception. *The Journal of the Acoustical Society of America, 75*, 1866-1878.

Werker, J. F., Fennell, C. T., Corcoran, K. M., & Stager, C. L. (2002). Infants' ability to learn

phonetically similar words: Effects of age and vocabulary size. *Infancy, 3*, 1-30.

Walley, A. C. (1988). Spoken word recognition by young children and adults. *Cognitive development, 3*, 137-165.

Whalen, D. H. (1984). Subcategorical phonetic mismatches slow phonetic judgments. *Perception & Psychophysics, 35*, 49-64.

Whalen, D. H. (1981). Effects of vocalic formant transitions and vowel quality on the English [s] / [ʃ] boundary. *The Journal of the Acoustical Society of America, 69*, 275-282.

Wood, C. C. & Day, R. S. (1975). Failure of selective attention to phonetic segments in consonant-vowel syllables. *Perception & Psychophysics, 17,* 346-335.

Yamada, R. A. & Tohkura, Y. (1992). The effects of experimental variables on the perception of American English /r/ and /l/ by Japanese listeners. *Journal of Perception & Psychophysics, 52*(4), 376-392.

Zwicker, E. (1961). Subdivision of the audible frequency range into critical bands (*Frequenzgruppen). The Journal of the Acoustical Society of America, 33*, 248.

Zygis, M. & Hamann, S. (2003). Perceptual and acoustic cues of Polish coronal fricatives. *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, 395-398.