

UNIVERSITY OF CALIFORNIA, SAN DIEGO

**Inductive Inference in Non-Native Speech Processing and Learning**

A dissertation submitted in partial satisfaction of the  
requirements for the degree  
Doctor of Philosophy

in

Linguistics

by

Bozena Pajak

Committee in charge:

Professor Eric Baković, Co-Chair  
Professor Roger Levy, Co-Chair  
Professor Sarah C. Creel  
Professor Victor Ferreira  
Professor Tamar Gollan  
Professor Sharon Rose

2012

Copyright  
Bozena Pajak, 2012  
All rights reserved.

The dissertation of Bozena Pajak is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

---

---

---

---

---

Co-Chair

---

Co-Chair

University of California, San Diego

2012

## TABLE OF CONTENTS

Signature Page . . . . .		iii
Table of Contents . . . . .		iv
List of Figures . . . . .		vi
List of Tables . . . . .		vii
Acknowledgements . . . . .		viii
Vita and Publications . . . . .		xii
Abstract of the Dissertation . . . . .		xiii
Chapter 1	Introduction . . . . .	1
	1.1 Significance . . . . .	1
	1.2 Place in the Literature . . . . .	3
	1.3 Outline of the Dissertation . . . . .	4
Chapter 2	A Hierarchical Inductive Theory of Multiple Language Learning . . . . .	6
	2.1 General Framework . . . . .	6
	2.2 Background . . . . .	11
	2.2.1 Speech Perception . . . . .	12
	2.2.2 Phonetic Categorization . . . . .	14
	2.2.3 Classic Theories . . . . .	15
	2.3 Hierarchical Inductive Theory in Non-Native Speech Processing . . . . .	19
	2.3.1 Informativity of Phonetic Dimensions . . . . .	20
	2.3.2 Perceptual Learning in Adulthood . . . . .	23
	2.3.3 Phonetic Category Induction . . . . .	23
	2.3.4 The Hierarchical Inductive Model . . . . .	24
	2.4 Predictions of the Hierarchical Inductive Model . . . . .	28
	2.4.1 General Predictions . . . . .	29
	2.4.2 Phonetic Dimensions Studied . . . . .	31
	2.4.3 Properties of the Languages Studied . . . . .	33
	2.4.4 Specific Predictions . . . . .	36
Chapter 3	L1 Perceptual Biases in Non-Native Speech Perception . . . . .	41
	3.1 Introduction . . . . .	41
	3.2 Experiment . . . . .	42
	3.2.1 Method . . . . .	42

	3.2.2 Results . . . . .	45
	3.3 Discussion . . . . .	47
	3.4 Chapter Appendix . . . . .	49
	3.5 Acknowledgements . . . . .	53
Chapter 4	L1 Perceptual Biases in Word Learning . . . . .	54
	4.1 Introduction . . . . .	54
	4.2 Experiment . . . . .	57
	4.2.1 Method . . . . .	58
	4.2.2 Results . . . . .	62
	4.3 Discussion . . . . .	72
	4.4 Acknowledgements . . . . .	73
Chapter 5	Overcoming L1 Perceptual Biases in Phonetic Category Learning . . . . .	74
	5.1 Introduction . . . . .	74
	5.2 Experiment 1 . . . . .	75
	5.2.1 Method . . . . .	75
	5.2.2 Results . . . . .	78
	5.3 Interim Discussion . . . . .	81
	5.4 Experiment 2 . . . . .	83
	5.4.1 Method . . . . .	84
	5.4.2 Results . . . . .	85
	5.5 Discussion . . . . .	86
	5.6 Acknowledgements . . . . .	88
Chapter 6	Combining L1 Perceptual Biases and Distributional Cues in Phonetic Category Learning . . . . .	89
	6.1 Introduction . . . . .	89
	6.2 Experiment . . . . .	89
	6.2.1 Method . . . . .	92
	6.2.2 Results . . . . .	97
	6.3 Discussion . . . . .	102
	6.4 Acknowledgements . . . . .	103
Chapter 7	Conclusion . . . . .	105
	7.1 Summary of Main Findings . . . . .	106
	7.2 Implications . . . . .	108
	7.2.1 Language Learning . . . . .	108
	7.2.2 Phonological Representations . . . . .	111
	7.3 Future Directions . . . . .	112
References	. . . . .	115

## LIST OF FIGURES

Figure 2.1: A hierarchical model of linguistic knowledge in acquisition of multiple languages. . . . .	9
Figure 2.2: A hierarchical model of phonetic category knowledge in acquisition of multiple languages . . . . .	25
Figure 3.1: Overall results. . . . .	46
Figure 3.2: Results from the critical <i>length</i> trials split by segment. . . . .	46
Figure 4.1: Example of a screen shot from the word-learning task. . . . .	61
Figure 4.2: Overall results. . . . .	63
Figure 4.3: Results from the critical <i>length</i> and <i>place</i> trials. . . . .	64
Figure 4.4: Critical results from the perceptual-discrimination task by block. . . . .	66
Figure 4.5: Critical results from the word-learning task by block. . . . .	67
Figure 4.6: Results from the word-learning task: faster learners. . . . .	69
Figure 4.7: Results from the word-learning task: slower learners. . . . .	70
Figure 5.1: Critical training stimuli in Experiment 1. . . . .	77
Figure 5.2: Experiment 1: proportion of ‘different’ responses on ‘different’ trials by participants trained on the sonorant segment class. . . . .	79
Figure 5.3: Experiment 1: proportion of ‘different’ responses on ‘different’ trials by participants trained on the fricative segment class. . . . .	80
Figure 5.4: Critical training stimuli in Experiments 1 and 2. . . . .	85
Figure 5.5: Experiment 2: proportion of ‘different’ responses on ‘different’ trials. . . . .	86
Figure 6.1: Schematic representation of segment statistics in two languages constructed for the purpose of this study. <b>Top:</b> strongly bimodal <i>place</i> and weakly bimodal <i>length</i> . <b>Bottom:</b> strongly bimodal <i>place</i> and unimodal <i>length</i> . . . . .	90
Figure 6.2: Critical training stimuli. . . . .	95
Figure 6.3: Critical ‘different’ word pairs in testing. . . . .	96
Figure 6.4: Proportion of ‘different’ responses on ‘different’ <i>place</i> trials. . . . .	98
Figure 6.5: Proportion of ‘different’ responses on ‘different’ <i>length</i> trials. . . . .	99

## LIST OF TABLES

Table 2.1:	Summary of the surface segmental inventories in English, Korean, Vietnamese, Cantonese, and Mandarin: length and alveolo-palatal vs. retroflex place of articulation. . . . .	36
Table 2.2:	Hypothetical language inventories. . . . .	37
Table 2.3:	Example predictions for discrimination performance on non-native length contrasts by native speakers of $L_A$ , $L_B$ , and $L_C$ . . . . .	37
Table 2.4:	Predictions for discrimination of length and place contrasts by speakers of different native languages. . . . .	38
Table 3.1:	Sound segments used in stimuli, and the occurrences of corresponding sounds in Korean, Vietnamese, Cantonese, and Mandarin. . . . .	43
Table 3.2:	Mean and standard deviation of self-reported participant characteristics: Korean speakers. . . . .	49
Table 3.3:	Mean and standard deviation of self-reported participant characteristics: Vietnamese speakers. . . . .	50
Table 3.4:	Mean and standard deviation of self-reported participant characteristics: Cantonese speakers. . . . .	51
Table 3.5:	Mean and standard deviation of self-reported participant characteristics: Mandarin speakers. . . . .	52
Table 4.1:	Stimuli. . . . .	59
Table 4.2:	Trial types in perceptual discrimination (AB=words presented auditorily) and word learning – testing (AB=labels for visually presented pictures). . . . .	61
Table 4.3:	Results from the word-learning filler trials. . . . .	68
Table 6.1:	Proportion of ‘different’ responses on ‘same’ trials. . . . .	97

## ACKNOWLEDGEMENTS

First, I would like to thank my advisors, Eric Baković and Roger Levy, who not only contributed their insights at all stages of the work reported in this dissertation, but also are the ones responsible for my transformation from an unfledged student and a linguistics novice into a “grown-up” academic.

Eric advised me throughout graduate school, which I would not have survived without his unwavering support and encouragement. He was there to guide my first steps in different areas of the academic life: my first conference, my first journal paper, my first journal review, my first teaching. . . His writing style—clear, concise, well-structured—has become the standard I have tried to emulate in my own work, facilitated by Eric’s consistent and thorough feedback. He made me push the boundaries of what I thought I could achieve, yet somehow he always knew what my true limits were. He inspired my interest in languages’ phonological patterns, which defined all my later research. He always encouraged me to pursue the questions I was passionate about, and his clever insights are reflected in all my research, from first year projects to dissertation work. His integrity, respect for other researchers and their work, and his remarkable ability to navigate the social world of academia have left a lasting impression on my professional life.

Roger became my official advisor later in graduate school, but his feedback has influenced my linguistics formation from early on. He introduced me to the fields of experimental psycholinguistics and cognitive psychology, and shared insights that profoundly impacted my dissertation work and the rudiments of my whole research program. He showed me how to strive for precision and coherence in the thought process, how to critically examine my own ideas and underlying assumptions, and—very generally—what it means to be a good scientist. His encouragement has given me confidence to pursue bigger and more complex questions, and his guidance has continuously led me to deeper understanding of my own work. His contagious enthusiasm for science in general has constantly reminded me of the real goal of being a researcher, the joy of discovery, and the sense of community among all scientists.

I would also like to express my deepest gratitude to the rest of my dissertation committee: to Sarah Creel, for sharing her passion for research on sound perception, for teaching me how to be a better experimental researcher, and for noticing little things that would inspire me to think for hours; to Victor Ferreira, for incredible instantaneous insights into my work, and helping me see the broader picture of where my research fits in; to Tamar Gollan, for sharing her knowledge and insights related to bilingualism, warning me about potential experimental pitfalls, and for encouraging me to pursue high-level theoretical questions; to Sharon Rose, for first introducing me to experimental research, for constant support and guidance throughout graduate school, and for always reminding me about the fine-grained cross-linguistic details relevant to my investigations.

I am grateful to the rest of the linguistics faculty who all in different ways contributed to my graduate training. Special thanks to Andrew Kehler for advice and feedback at different points in graduate school, and especially for sharing his insights about different aspects of academic life in the form of the hilarious and incredibly useful Linguistics 293 class. I would also like to thank the whole CRL community, members of the San Diego Phonology Interest Group, Computational Psycholinguistics Lab, and Language Acquisition & Sound Recognition Lab for creating such a vibrant and supportive environment that helped me thrive intellectually. Special thanks to fellow graduate students in my cohort: Kathryn Davidson Zaremba, Gabriel Doyle, J. Grant Loomis, and Mohammad Salihie, whose company, support, and friendship made my cultural adjustment at the beginning of graduate school a seamless and fun experience.

I would not have been able to do this research without the help of many undergraduate research assistants: Hannah Byers-Straus, Annabelle Cadang, Rafi Feliciano, Christopher Gaudiot, Rachel O'Sullivan, Nicole Pyon, Ryan Short. Special thanks are due to K. Michael Brooks whose initiative and willingness to learn have made him not only my longest-term research assistant, but also a true collaborator. I could always count on the support of all the staff, among which I am especially grateful to Margaret Paulson, Alycia Randol, Ezra Van Everbroeck, and Lucie Wiseman for always being there to solve any administra-

tive or technical difficulties I encountered. Finally, I would like to acknowledge my funding both from the Linguistics Language Program and the Center for Research in Language.

My acknowledgements would not be complete without thanking those who helped me on my way to graduate school. I am grateful to Ivan Dihoff for inspiring me to pursue linguistics and for encouraging me to apply to graduate schools in the U.S. Thanks to Manuel Rubio Sánchez for helping me get through the application process that seemed so daunting and overwhelming from a foreigner's perspective.

I am grateful to my family in Kraków for raising me to be independent and resourceful, and for fully supporting all my career choices despite my inability to explain what it is that I do. Thanks to all my in-laws in Texas and Louisiana for their support and encouragement, especially at the very last stages of writing this dissertation during family vacation.

Finally, the greatest thanks go to my husband who found the secret to keeping me sane throughout the ups and downs of graduate school (which will likely also come in handy later on). His love, support, and invariant optimism lighten every day of my life.

\* \* \*

This research was supported by NIH Training Grant T32-DC000041 from the Center for Research in Language at UC San Diego to the dissertation author, and also by a UC San Diego Academic Senate Grant to Roger Levy.

Chapter 3 is a revised version of Pajak (2010) ["Perceptual advantage from generalized linguistic knowledge". In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 369–374). Austin, TX: Cognitive Science Society], and it includes material submitted for publication [Pajak & Levy. "Inductive generalization over phonetic categories in perceptual reorganization"]. The dissertation author was the primary investigator and author of these papers. In addition to being presented to the Cognitive Science Society, this work was also presented at the 84th Annual

Meeting of the Linguistic Society of America, and at the 51st Annual Meeting of the Psychonomic Society.

Chapter 4 is a revised version of Pajak, Creel, & Levy (2012) [“Can native-language perceptual bias facilitate learning words in a new language?” In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 2174–2179). Austin, TX: Cognitive Science Society], and it includes material currently being prepared for submission for publication [Pajak, Creel, & Levy. “Can native-language perceptual bias facilitate learning similar words in a new language?”]. The dissertation author was the primary investigator and author of these papers. In addition to planned presentation to the Cognitive Science Society in August 2012, this work was also presented at the 86th Annual Meeting of the Linguistic Society of America.

Chapter 5 is a revised version of Pajak & Levy (2011) [“Phonological generalization from distributional evidence”. In L. Carlson, C. Hölscher & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 2673–2678). Austin, TX: Cognitive Science Society] and Pajak & Levy (to appear) [“How abstract are phonological representations? Evidence from distributional perceptual learning”. In *Proceedings of the 47th Annual Meeting of the Chicago Linguistic Society*. Chicago, IL: University of Chicago]. The dissertation author was the primary investigator and author of these papers. This work was presented to the Cognitive Science Society and the Chicago Linguistic Society.

Chapter 6 is a revised version of Pajak & Levy (2012) [“Distributional learning of L2 phonological categories by listeners with different language backgrounds”. In A. K. Biller, E. Y. Chung, & A. E. Kimball (Eds.), *Proceedings of the 36th Boston University Conference on Language Development* (pp. 400–413). Somerville, MA: Cascadilla Press]. The dissertation author was the primary investigator and author of these papers. This work was presented at the 36th Boston University Conference on Language Development.

## VITA

- 2005 M.A. in International Relations  
Jagiellonian University, Krakow, Poland
- 2006 M.A. in Sociology  
Jagiellonian University, Krakow, Poland
- 2008 M.A. in Linguistics  
University of California, San Diego
- 2012 Ph.D. in Linguistics  
University of California, San Diego

ABSTRACT OF THE DISSERTATION

**Inductive Inference in Non-Native Speech Processing and Learning**

by

Bozena Pajak

Doctor of Philosophy in Linguistics

University of California, San Diego, 2012

Professor Eric Baković, Co-Chair

Professor Roger Levy, Co-Chair

Despite extensive research on language acquisition, our understanding of how people learn abstract linguistic structures remains limited. In the phonological domain, we know that perceptual reorganization in infancy results in attuning to native language (L1) phonetic categories and, consequently, in difficulty discriminating and learning non-native categories. This difficulty has been proposed to originate from novel sounds being perceptually mapped onto L1 phonetic categories, leading to massive L1 interference. However, ample evidence that the adult speech processing system preserves a considerable degree of plasticity suggests that more complex learning mechanisms might be in place. In this dissertation I propose an alternative theory in which non-native speech

processing is guided by principles of hierarchical inductive inference regarding how likely a given phonetic dimension is to be phonologically informative in any novel language. This theory differs crucially from mapping theories in predicting that when a phonetic dimension is informative (e.g., phonologically contrastive) in one's native language, discriminations involving that dimension should be enhanced even among classes of sounds for which the dimension is not informative in the native language. I provide experimental evidence supporting the inductive theory, demonstrating that language learning goes beyond the acquisition of specific phonetic categories, and includes higher-order generalizations regarding the relative importance of phonetic dimensions in the language as a whole. I argue that this theory can be extended beyond phonetic category learning to other domains of language acquisition, and that it suggests that adults and infants recruit the same domain-general learning mechanisms when acquiring novel languages.

# Chapter 1

## Introduction

### 1.1 Significance

Languages are highly complex systems with abstract multi-level structures. Yet, humans are capable of learning languages without explicit instruction, through mere exposure to speech input. Furthermore, multiple languages can be learned by a single individual throughout their lifetime. How do people achieve this daunting task? And how is this intricate language knowledge represented in the mind? These are the high-level questions guiding the research pursued in this dissertation.

Language acquisition has generally been understood as a transition from an initial state of knowledge, in which the entire range of language potentialities is available, to a later (fluent-speaker) state of knowledge of a specific language. Acquisition of any additional languages in adulthood has been viewed as limited by cognitive maturation, as well as by native-language (L1) knowledge that appears to interfere with acquiring novel linguistic structures.

In this dissertation I propose an alternative to this classic approach: a new theory of language acquisition based on the general hierarchical inductive approach to learning. The basic premise of this view is that learning involves making simultaneous inductive inferences at multiple levels of abstraction: not only about particular categories, but also about higher-level category structure. Following this main line of thinking, I posit that through L1 exposure learners not

only acquire the abstract representations of that specific language—as identified in prior work—but also form higher-order generalizations regarding possible structures of other languages, which then affect acquisition of any additional language. I claim that these generalizations, initially formed in early childhood, continue evolving throughout life span with each additional language learned.

I develop this general approach to language acquisition in the area of phonological acquisition, including perception and learning of speech sounds. The theory reconceptualizes perceptual “tuning” to native-language sounds in the first year of life as a problem of hierarchical inductive inference about the structure of the language’s sound system: learners are proposed to not only acquire the specific L1 phonetic category inventory, but also draw higher-order generalizations about the properties of the L1 sound system as a whole.

The theory also offers a novel approach to non-native sound perception and second language (L2) phonetic category learning. In particular, I argue that the constraints on non-native sound perception by naive monolingual listeners are imposed by the higher-order generalizations established during perceptual development in infancy, and not just by the specific inventory of L1 sounds learned. I further claim that the process of phonetic category learning in L2 is guided by similar inductive inferences as L1 acquisition. However, the initial inferences are in this case made based on both the properties of L2 input and the higher-order generalizations about the likely properties of other languages.

Finally, the theory provides a framework that generates specific predictions about the patterns of phonetic category acquisition in each additional language – patterns that depend on the kind of linguistic properties that had been acquired through previous language exposure.

In addition to the theoretical contribution, this dissertation includes a set of experiments, which were designed to test some empirical predictions of the hierarchical inductive theory. The experiments reported here examine how an individual’s previous linguistic experience affects perception of non-native contrasts, and what kinds of inferences learners make regarding phonetic categories underlying input from a novel language. The results are consistent with the initial predictions of the theory, supporting the view of phonological acquisition

as guided by principles of hierarchical inductive inference.

The proposed theory and experimental results have strong implications for both theoretical and applied linguistics. Investigating the patterns of non-native sound perception and categorization provides insights into questions that are relevant for any theory of language: how native-language knowledge is represented in the mind and how it changes over time. By examining all the subtle ways in which language knowledge affects learning of novel languages, we can gain a better understanding of what that initial language knowledge was in the first place, thus shedding some light onto the nature of linguistic representations. Furthermore, investigating how learning novel languages progresses allows us to understand how overall linguistic knowledge continues to develop with exposure to new language properties, and how much plasticity remains in linguistic representations that are formed in adulthood. Finally, this research can have implications for applied research in language teaching. Specifically, understanding how learners represent and use previously acquired linguistic knowledge can help develop better teaching methodologies that would guide learners to take advantage of their previous knowledge while minimizing potential interference.

## **1.2 Place in the Literature**

The hierarchical inductive theory of language learning proposed in this dissertation fits within the general approach to learning as a process of rational hypothesis construction and testing, in which learners infer the underlying structure of their input by generalizing beyond the specific surface properties that they are exposed to (e.g., Tenenbaum & Griffiths, 2001; Xu & Tenenbaum, 2007; Gerken, 2010; Tenenbaum, Kemp, Griffiths, & Goodman, 2011). In the specific domain of phonological acquisition, the theory builds on insights from previous work on induction in speech category learning (de Boer & Kuhl, 2003; Vallabha, McClelland, Pons, Werker, & Amano, 2007; McMurray, Aslin, & Toscano, 2009; Feldman, Griffiths, & Morgan, 2009), also drawing from the general literature on perceptual categorization (Posner & Keele, 1968; Nosofsky, 1986; Kr-

uschke, 1992), as well as findings regarding perceptual reorganization and phonetic category learning in infancy (Eimas, 1978; Jusczyk, 1985, 1992; Kuhl, 1991; Kuhl et al., 2008) .

At the same time, the proposed theory is radically different from other models of non-native speech perception and L2 phonological acquisition, where perception and learning of novel sounds have been assumed to rely on the process of mapping of L2 sounds onto L1 phonetic categories (Best, 1995; Best & Tyler, 2007; Flege, 1995; Hancin-Bhatt, 1994; Kuhl & Iverson, 1995; Kuhl, 2000a). Under these views, L2 learners—instead of making implicit rational predictions about the L2 phonetic categories—try to establish conceptual links between L2 sounds and their most similar L1 counterparts, so as to process the unfamiliar sounds directly through their L1 phonological system. I propose, in contrast, that learners do not directly filter the L2 speech input through their L1 phonetic categories, but rather that they make the best possible guesses about how individual novel sounds are grouped into categories by relying on the same mechanisms that are used in general categorization processes for many types of perceptual stimuli.

This dissertation is thus intended to bring together insights from multiple research areas that study speech processing and learning from varying perspectives, the main goal being to contribute to our understanding of the fundamental nature of language acquisition—whether native or non-native—of mental linguistic representations, and of language development across the life span.

### **1.3 Outline of the Dissertation**

This dissertation is organized into seven chapters. Chapter 2 introduces the inductive theory of learning multiple languages, situates the proposal within the relevant literatures, as briefly overviewed in the previous section, and outlines specific predictions of the theory in the domain of phonological acquisition. Chapters 3-6 describe experimental studies that test these predictions. Chapter 3 shows how native-language perceptual biases can facilitate percep-

tion of non-native contrasts, even when it requires generalization across dissimilar classes of sounds such as vowels and consonants. Chapter 4 extends these findings to learning novel minimal-pair words with the same contrasts. Chapter 5 provides evidence that native-language perceptual biases can be overcome through a short distributional training on novel contrasts, and that these newly learned contrasts generalize across dissimilar sound classes. Chapter 6 presents data showing how the two sources of information—native-language biases and distributional cues—interact in the early stages of learning a new language. Finally, Chapter 7 summarizes the main findings and outlines possibilities for future research.

## Chapter 2

# A Hierarchical Inductive Theory of Multiple Language Learning

### 2.1 General Framework

Infants are born with an ability to learn any of the world's languages that are present in their environment, but through exposure to speech in a particular language they transition from this initial, unconstrained state of knowledge to a fluent-speaker state of knowledge of that language. Language learning is not, however, restricted to the early stages of life. Languages can be acquired throughout the life span (Cenoz & Genesee, 1998; Cenoz, Hufeisen, & Jessner, 2001; De Angelis, 2007), although the ability to achieve native-like proficiency in a new language gradually declines with age (Flege, Yeni-Komshian, & Liu, 1999; Stevens, 1999; Hakuta, Bialystok, & Wiley, 2003).

In the classic approaches to L2 acquisition (see, for example, MacWhinney, 1987; Towell & Hawkins, 1994; Mitchell & Myles, 1998; MacWhinney, 2008; Ritchie & Bhatia, 2009) the common notion is that learning a new language relies strongly on L1 knowledge, with possible access to innate linguistic biases (Universal Grammar; Chomsky, 1965). The role of L1 knowledge is generally viewed as a source of interference (Selinker, 1969; Odlin, 1989): that is, the principles governing L1 grammar are transferred directly to L2, and interfere with learning the L2 properties that differ from L1. In the most extreme view,

the Representational Deficit Hypothesis (e.g., Hawkins & Chan, 1997; Hawkins & Liszka, 2003; Tsimpli, 2003), linguistic properties not found in L1 are believed to be unacquirable by adult learners. This means that L2 representations must remain non-target-like regardless of the amount of L2 exposure or proficiency, even though learners might learn to “mimic” the surface L2 properties (Towell & Hawkins, 1994). Other authors argue that L2 learning is only constrained by L1 representations in an initial stage of learning, and with more exposure learners are able to revise their L2 representations in line with the input they are receiving (Schwartz & Sprouse, 1994), or that only some aspects of L2 structure remain unlearnable (Goad & White, 2006).

In this dissertation I suggest a different approach, in which learners—instead of being strictly constrained by their L1 representations—use their L1 knowledge as one of the sources of information as they make inferences about the underlying structure of an L2 and as they build new L2 representations. This idea follows from the general approach to learning as a process of rational hypothesis construction and testing (e.g., Tenenbaum & Griffiths, 2001; Xu & Tenenbaum, 2007; Gerken, 2010; Tenenbaum et al., 2011). Under these views, learning is an inductive inference problem: learners have access to some observed data (e.g., speech), based on which they try to recover the underlying structure of a model that would reproduce the data they have observed (e.g., a model of language grammar). Learners consider multiple possible models at the same time, and incrementally update the inferred probability of each model as they are exposed to more data (Gerken, 2010). I propose that L1 knowledge—together with knowledge of any other previously acquired languages—should be considered one of the biases that guide this inductive inference process in language acquisition.

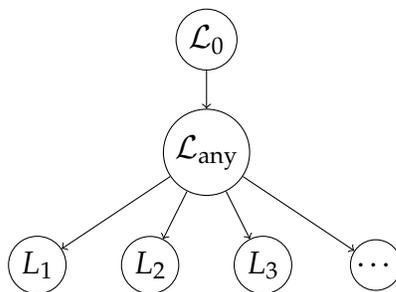
Additionally, I adopt a hierarchical view of the inductive learning process, which is informed by recent progress in the computational modeling of the acquisition of abstract knowledge (Chater & Manning, 2006; Tenenbaum et al., 2011). A common theme underlying the success of these models is that learning occurs not only at a single, flat level of representation but rather hierarchically, with the learner making simultaneous inductive inferences not only

about particular categories, but also about higher-level category structure. To take one example from the language domain: when acquiring individual verbs in a language, learners also infer general properties of verbal classes that they then extend to any novel verbs they encounter (Perfors, Tenenbaum, & Wonnacott, 2010).

An intuitive illustration of hierarchical inductive inference is Goodman's (1955) example of colored marbles (see also Kemp, Perfors, & Tenenbaum, 2007). Imagine a stack of bags, where each bag is filled with colored marbles. Drawing a single marble out of a bag provides you with little information about other marbles in that bag. However, if you had previously emptied several bags, you might be able to formulate hypotheses about the content of a new bag based on a single draw from that bag. For example, suppose that you had previously emptied two bags and discovered that the first bag was filled with black marbles while the second one was filled with white marbles. Now suppose that you draw a single blue marble out of a new bag. Based on your experience with previous bags you are likely to hypothesize that the third bag contains only blue marbles. Note, however, that your inferences are strictly dependent on the previously observed data: if the first two bags contained instead marbles of different colors, you would be unlikely to hypothesize that the third bag contains only blue marbles based on a single blue-marble draw.

How do people gain these intuitions based on such limited amount of observed data? In the marble example, through experience with the stack of bags you did not only gain knowledge about marbles in each specific bag you had emptied, but you also made higher-level inferences about the stack in general: for example, that each bag contains marbles that are uniform in color, or that each bag has mixed-color marbles (Goodman, 1955; Kemp et al., 2007). Thus, we can explain the intuitions that people have about any unknown properties of the world by saying that people are able to infer abstract multi-level structures of how the world is constructed based on previously gained experience with observable world data.

I propose to view the process of learning multiple languages by following a similar line of reasoning. Let me illustrate this proposal by making an analogy



**Figure 2.1:** A hierarchical model of linguistic knowledge in acquisition of multiple languages.

with the marble example. Imagine that each language is like an individual bag, and marbles correspond to speech that is produced in that language. If learners are biased to expect that languages share structural similarities, then learning about the properties of each language should lead them not only to make inferences about that language in particular, but also to make higher-level inferences about the properties of any language.

More formally, I propose a new framework to investigate acquisition of multiple languages, which is a hierarchical model of an individual’s total linguistic knowledge—that is, knowledge of all languages learned—schematically illustrated in Figure 2.1. The model reflects my proposal that linguistic knowledge of an individual is, at any stage of acquisition, a hierarchical structure with multiple levels of representation. The bottom level corresponds to the knowledge of single learned languages:  $L_1$ ,  $L_2$ , and so on. The top level represents the initial learning bias ( $\mathcal{L}_0$ ), which can be viewed as Universal Grammar (Chomsky, 1965), innate domain-general cognitive abilities and learning predispositions (Elman et al., 1996), or a combination of both. The main novel contribution of the proposed model is the intermediate level  $\mathcal{L}_{\text{any}}$ , which represents more abstract beliefs regarding possible languages that learners obtain alongside knowledge of their native language.

The core of my  $\mathcal{L}_{\text{any}}$  proposal is that the outcome of acquisition of a language— $L_1$  or otherwise—is not only the knowledge of the specific language in question, but also beliefs regarding the structure of any other language. I view learning a language—whether in infancy or adulthood—as a process of

inductive inference, where learners continually make implicit predictions about the possible underlying structures of the novel language they are learning as they are exposed to more input.<sup>1</sup> Learners form these predictions by combining two main sources of information: (1) the statistical properties of the novel language input, and (2)  $\mathcal{L}_{\text{any}}$  biases, which are informed by all previously learned languages and the initial learning bias  $\mathcal{L}_0$ . Therefore, previous linguistic knowledge serves as an inductive bias guiding learners in their inferences about the novel language by affecting their interpretation of language input statistics.

At a high level, the hierarchical inductive theory immediately accounts for basic phenomena that other theories fail to account for. For example, there is a common intuition that while learning an L2 is relatively hard, learning each additional language ( $L_n$ ) becomes easier. There is some research suggesting that this intuition might be correct. In particular, it has been found that early bilinguals learning an L3 perform better than monolinguals learning the same language as their L2 on various standard measures of language proficiency (e.g., vocabulary and grammar tests, listening and reading comprehension tasks, essay writing, oral interviews) (Thomas, 1988; Valencia & Cenoz, 1992; Cenoz & Valencia, 1994; Sanz, 2000). Similar results were also reported for bi-/multilinguals and monolinguals learning an artificial language in the laboratory (Nation & McLaughlin, 1986; Kaushanskaya & Marian, 2009; Kaushanskaya, 2012).

Under the inductive theory, learning an L2 should indeed be relatively hard due to the strong bias resulting from learning the native language. However, learning each additional language should be facilitated because existing knowledge of multiple languages should sharpen learner's abstract knowledge regarding the likely properties of any language. Other theories can explain the difficulty in L2 learning as a result of direct interference from L1 representations, but they have no straightforward way of predicting facilitation for each  $L_n$  other than in cases when the properties of the target language (i.e., the new language being learned) coincide exactly with the properties of a known language, as is

---

<sup>1</sup>For the moment I abstract away from any maturational differences between child and adult language acquisition. This issue is discussed in more detail in sections 2.2 and 2.3.

for example the case of learning cognates (Tréville, 1996).

In the rest of this chapter I spell out the details of this general hierarchical inductive framework with specific reference to phonological acquisition. I first discuss some background literature on speech perception development and phonetic category learning in both native and non-native languages, and I review other theories of perception and learning of non-native speech sounds. Then, I describe in more detail how the inductive theory can be applied to the problem of learning novel phonetic categories, and how it differs from other theories. Finally, I outline specific predictions of the inductive theory, which I test in experiments presented in subsequent chapters.

## 2.2 Background

Phonological acquisition refers to learning the sound system of the target language. In this dissertation I am only concerned with the acquisition of individual sound categories, but other levels of phonological structure, such as syllable or prosody, may also be analyzed within the inductive framework.

Phonological acquisition is thus intimately related to auditory processing of speech, which is a complex multi-stage process. The highly variable acoustic speech signal must first undergo low-level perceptual processing of acoustic-phonetic cues. Based on this analysis, the signal is segmented into phonetic units, which can be conceptualized as points in a multidimensional space representing values across multiple acoustic-phonetic dimensions (such as fundamental frequency, formant frequency, etc.). These phonetic units are then mapped to abstract linguistic representations, such as phonemes, and matched to words in the mental lexicon (Klatt, 1989; Pisoni & Remez, 2005). Learning the sound inventory of a language involves, thus, learning how to efficiently process the acoustic-phonetic cues, and how to group acoustically variable phonetic units into functionally equivalent phonemic categories.

In this section I review some relevant literature that bears on both speech perception and phonetic category learning. First, I discuss the development of speech perception in both native and non-native languages. Then, I turn into

statistical learning of phonetic categories. Finally, I discuss the classic theories that have been proposed to account for perception and learning of non-native sounds.

### 2.2.1 Speech Perception

The development of speech perception in the first year of life provides a critical foundation for future language learning. Infants undergo profound *perceptual reorganization* (Eimas, 1978): they transition from discriminating almost any speech sound distinction (including those absent from their ambient language) to a state of enhanced sensitivity to native-language distinctions, accompanied by a decline in sensitivity to many non-native distinctions (Werker & Tees, 1984a; for reviews, see Werker, 1989; Kuhl, 2004).

These results reflect the process of tuning to the native-language sound system, through which the processing of the native-language speech signal becomes largely automatized (Jusczyk, 1992). This has led to the development of theories in which perceptual reorganization is understood as resulting from the acquisition of the specific inventory of native-language phonetic categories. In the Native Language Magnet model (Kuhl, 1994, 2000a, 2004), exposure to the native language leads to the formation of a language-specific perceptual filter, where the perceptual space becomes reconfigured (“warped”): innate perceptual sensitivity along natural auditory boundaries is replaced by sensitivity along boundaries of phonetic categories in the learner’s native language. Furthermore, the warping of perceptual space produces gradience in perceptual sensitivity – decreased near category modes and increased near the boundaries between categories (the *perceptual magnet* effect; Kuhl, 1991).

Perceptual reorganization in the first year of life involves certain *neural commitment* (Kuhl, 2000b, 2004): the brain’s neural networks commit to the variation patterns found in the native language, which may interfere with the processing of information that does not conform to those patterns. This means that perceptual reorganization fundamentally affects perception of non-native sounds (Burnham, 1986; Strange, 1987, 1995; Bohn & Munro, 2007; Hansen Ed-

wards & Zampini, 2008). Specifically, adults often have difficulty discriminating between speech sounds that do not exist in their native language (Abramson & Lisker, 1970; Werker, Gilbert, Humphrey, & Tees, 1981; Flege & Eefting, 1986, 1987; Jamieson & Morosan, 1986; Polka, 1991, 1992, among others), and recruit significantly greater brain resources when processing subtle non-native sound distinctions compared to processing native-language sounds (Zhang, Kuhl, Imada, Kotani, & Tohkura, 2005). A well-studied example is the case of Japanese native speakers learning the English /r/-/l/ distinction. In particular, L1-Japanese learners have difficulty distinguishing between the two speech sounds, which has been proposed to be the consequence of Japanese only having one phonetic category in the same acoustic-phonetic range (Goto, 1971; Strange & Dittmann, 1984; Miyawaki et al., 1975).<sup>2</sup>

The way sounds are perceived is not, however, completely set in early infancy. Children continue improving on discrimination of less salient native contrasts, and even 12-year-olds are not entirely adult-like (Hazan & Barrett, 2000; Narayan, Werker, & Beddor, 2010; Polka, Colantonio, & Sundara, 2001). Additionally, learning a second language in later childhood often results in native-like perception and production (L. Williams, 1979; Flege et al., 1999). Finally, these perceptual changes are not purely sensorineural since even in the adult speech processing system there is still a high degree of plasticity (Werker & Tees, 1984b): adults adapt to accented speech (Norris, McQueen, & Cutler, 2003), and show improvement on discrimination of subtle non-native distinctions in tasks with lower processing demands or after training (Logan, Lively, & Pisoni, 1991; Pisoni, Aslin, Perey, & Hennessy, 1982; Werker & Tees, 1984b; Goudbeek, Cutler, & Smits, 2008; Lim & Holt, 2011; McClaskey, Pisoni, & Carrell, 1983).

---

<sup>2</sup>Not all non-native distinctions are hard to discriminate. For example, sounds that differ considerably from any native-language category, such as click sounds for English native speakers, may be discriminated very well despite learners' lack of experience with this particular class of sounds (Best, McRoberts, & Sithole, 1988).

## 2.2.2 Phonetic Categorization

Perceptual tuning to native-language sounds in the first year of life goes hand in hand with their phonetic categorization. That is, infants learn to group acoustically variable phonetic units into functionally equivalent categories such as phonemes.<sup>3</sup>

One of the mechanisms underlying phonetic categorization is distributional learning. Frequency distribution patterns of sounds provide crucial information about the category structure of a language. Specifically, modal values of acoustic-phonetic cues indicate the center of a category, whereas low frequencies of cues indicate category boundaries. Thus, a trimodal distribution along some continuous acoustic dimension, such as voice onset time (VOT), may indicate a three-category distinction (e.g., /b/, /t/, and /t<sup>h</sup>/); a bimodal distribution may suggest a two-category distinction (e.g., /b/ vs. /p/); while a unimodal distribution along the same range of values implies a single category with greater variance. Infants extract these distributional frequencies from the speech signal (Maye, Werker, & Gerken, 2002; Maye, Weiss, & Aslin, 2008) and form category representations that encode information about modal values of categories (Kuhl, 1991; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Lacerda, 1995).

Statistical learning from frequency distributions of acoustic-phonetic cues is still relatively poorly understood. However, there is evidence suggesting that it is a powerful learning mechanism as infants are able to generalize a newly-learned category distinction to novel contexts (Maye et al., 2008). Specifically, infants exposed to a novel bimodal distribution along the VOT dimension for one place of articulation (e.g., alveolar) not only learn that novel distinction, but also generalize it to an analogous contrast for another place of articulation (e.g., velar).

Adults are also able to extract distributional information from novel speech signal (Maye & Gerken, 2000, 2001; Perfors & Dunbar, 2010), indicating that sensitivity to subtle statistics in non-native speech is preserved be-

---

<sup>3</sup>It is possible that there are separate functional groupings not just for phonemes, but also for (at least some) allophones, as discussed further in section 2.3.

yond infancy. However, there is mixed evidence on whether adults can—like infants—generalize newly-learned distinctions to novel contexts. Early studies with explicit category training showed that this type of generalization is possible for novel VOT distinctions (McClaskey et al., 1983; Tremblay, Kraus, Carrell, & McGee, 1997). On the other hand, exposing adults to distributional evidence for such novel VOT distinctions was inconclusive regarding the ability of participants to generalize to a different place of articulation: Maye and Gerken (2001) reported no generalization, but Perfors and Dunbar (2010) found some evidence of generalization by increasing the duration of training and using natural stimuli. Thus, there is some indication that adults might be using similar statistical learning mechanisms to those used in infancy when acquiring non-native phonetic categories.

### 2.2.3 Classic Theories

There are three main theories that have been used to explain the patterns of non-native speech perception and learning: the Native Language Magnet model (NLM, Kuhl, 1992, 1994; Kuhl & Iverson, 1995; Kuhl, 2000a; Kuhl et al., 2008), the Speech Learning Model (SLM, Flege, 1988, 1992, 1995), and the Perceptual Assimilation Model (PAM and PAM-L2, Best, 1993, 1994, 1995; Best & Tyler, 2007). These theories, while different in several respects, share the underlying assumptions that (1) L1 representations are the basis for building representations of a novel language, and that (2) L1-attuned perceptual space—the end-result of perceptual reorganization in infancy—acts as a direct perceptual filter when processing novel languages, which is an idea already found in Trubetzkoy (1939/1969). On these views, discrimination of any non-native sounds is determined by their *mapping* onto specific native-language phonetic categories that are acoustically or articulatorily most similar: discrimination of non-native contrasts is enhanced when the percepts in the contrast are mapped (assimilated) onto differing phonetic categories, and diminished when mapped onto the same phonetic category. Each model is described below in more detail.

The NLM model (Kuhl, 1992, 1994; Kuhl & Iverson, 1995; Kuhl, 2000a;

Kuhl et al., 2008) is—as already briefly described in section 2.2.1—a formalization of the *perceptual magnet* effect (Kuhl, 1991; Kuhl et al., 1992). The magnet effect is a result of warping the perceptual space. The theory states that due to shrunken perceptual distances around phonetic category modes, sounds that are phonetically similar to a category mode will be perceptually assimilated to that category. This process can be intuitively visualized as phonetic categories pulling all similar percepts toward them, producing degraded discriminability of sounds that fall within category boundaries.

The NLM model was initially conceived for purposes of native language acquisition, but it can be straightforwardly extended to non-native speech perception. Specifically, just like native-language sounds, any non-native percept will be pulled toward most similar native-language category modes (Kuhl & Iverson, 1995). As a consequence, if two non-native percepts are similar to only one native-language phonetic category (as is the case of English /r/-/l/ sounds for L1-Japanese speakers), then their discrimination will be degraded due to both of the sounds being assimilated onto the region of L1-attuned perceptual space that corresponds to a single category.

The SLM (Flege, 1988, 1992, 1995) focuses on explaining pronunciation errors in L2 speech production, hypothesizing that at least some of them arise from inaccurate perception of novel sounds. The basic premise of the model is that L1 and L2 sounds are perceptually linked, and learning of novel L2 sounds will be hard if phonetic differences between L1–L2 pairs of sounds are very subtle. These differences between sounds may be hard to detect because of two processes: (1) equivalence classification – mapping of acoustically similar sounds onto one phonetic category, and (2) filtering out of features (subsegmental properties) of L2 sounds that are phonologically irrelevant in L1. Sound perception is assumed to be position-sensitive, which means that discrimination between sounds may depend on their phonotactic environment. The greater the perceived phonetic similarity between two percepts, the higher the likelihood of classifying them as belonging to the same phonetic category. In this model, then, discrimination of two non-native percepts will be degraded if they are classified as equivalent to a single L1 category.

The second component of the SLM, which alludes to the filter metaphor, was later developed into the *feature hypothesis* stating that “features not used to signal phonological contrast in L1 will be difficult to perceive for the L2 learner and this difficulty will be reflected in the learner’s production of the contrast based on this feature” (McAllister, Flege, & Piske, 2002, p. 230). This proposal is based on the premise from theoretical phonology that speech sounds can be decomposed into distinctive features that are shared among natural classes of sounds (e.g., both /p/ and /b/ belong to the natural class of labial stops and can be described as [+labial], [-sonorant], [-continuant], but they differ in that /p/ is [-voice] and /b/ is [+voice]).

This is similar to Brown’s (1997, 2000) model of speech perception, where L2 phonological acquisition is argued to include transfer of distinctive features used in L1 to the novel language, which can then facilitate discrimination of novel sounds that are contrasted by those features. It is also related to the Feature Competition Model of segment transfer (Hancin-Bhatt, 1994), where it is proposed that learners use phonological distinctive features to compute similarity between L2 sounds and L1 categories, which in turn determines L2-to-L1 sound mappings.

According to SLM, the phonetic category structure established in infancy does not, however, remain stable throughout adulthood, but rather it keeps evolving across the life span. Novel phonetic categories can thus also be established later in life with sufficient L2 exposure. The only condition is that at least some phonetic differences between the L2 percept and L1 sounds must be perceptually detected to establish a new category. This means that learning of L2 sounds that have been classified as perceptually equivalent to some L1 categories is expected to be especially hard.

PAM (Best, 1993, 1994, 1995) was initially designed to account for non-native speech perception patterns by naive monolingual speakers, but was later also extended to experienced L2 learners under the name of PAM-L2 (Best & Tyler, 2007). The theory advocates the direct-realist perspective on perception (J. J. Gibson, 1966; E. J. Gibson, 1969; J. J. Gibson, 1979; E. J. Gibson, 1991), in which listeners are assumed to directly perceive articulatory gestures that gen-

erated the speech signal.

In PAM, any non-native percept assimilates onto native-language perceptual space and—if processed as a speech sound—either (1) is categorized as an instance of a native category if commonalities in articulatory gestures are detected, or (2) remains uncategorized if dissimilar from all native categories, thus falling in between specific native categories. Discrimination of a given non-native contrast depends on how each of the sounds is assimilated to the native perceptual system. Discrimination is good if the two sounds assimilate to two different native-language categories (*two-category assimilation*). Discrimination is somewhat impaired if the two sounds assimilate to the same native-language category but differ in overall fit for that category (e.g., one is acceptable, the other is deviant; *category-goodness difference*). Finally, discrimination is poor if the two sounds assimilate to a single native-language category as equally good instances of that category (*single-category assimilation*).

The theories described above—which I will refer to as *mapping theories* in the rest of the dissertation—differ in how exactly L2 percepts map onto L1 categories, but they all share the core premise that these mappings determine perception and learning of non-native sounds because any novel percepts are perceived in terms of native phonetic categories.

The mapping theories are supported by a considerable body of evidence from studies on non-native sound discrimination (for a review see Strange & Shafer, 2008), showing that the degree of similarity between native and non-native sounds—as assessed through acoustic and articulatory comparisons or direct measures of perceived similarity<sup>4</sup>—can predict performance on discrimination of non-native sound pairs. That is, if two non-native sounds are both assessed as highly similar to a single native-language category, their discrimination is predicted to be hard. On the other hand, if each sound in the non-native pair is considered highly similar to a distinct native-language category, then their discrimination is predicted to be easy.

These predictions have been repeatedly confirmed by a wide range of

---

<sup>4</sup>Measures of perceived similarity are based on auditory presentation of non-native sounds, where the task is to categorize each novel sound in terms of L1 categories, either through forced-choice or open response (Strange & Shafer, 2008).

studies (Miyawaki et al., 1975; Flege & Eefting, 1987; Best & Strange, 1992; Polka, 1991, 1992; Hallé, Best, & Levitt, 1999; Best, McRoberts, & Goodell, 2001; McAlister et al., 2002; Best & Hallé, 2010, among others), with the focus on cases of single-category and category-goodness assimilation (using the terminology from PAM). This means that research within all three frameworks has largely been restricted to explaining how L1 *interferes* with perception and learning of non-native sounds.

On the other hand, the framework proposed in this dissertation allows us to expand the focus of this general research area by exploring potential *benefits* of native-language knowledge. Crucially, the proposed model predicts facilitation in discrimination and learning of sounds that goes beyond immediate similarities between languages and that cannot be explained by straightforward two-category assimilation. Specifically, the proposed theory accommodates the possibility that learners make *generalizations* about novel sound systems based on their native-language knowledge, as described in more detail in the next section.

## 2.3 Hierarchical Inductive Theory in Non-Native Speech Processing

In the present section I follow the inductive framework, as sketched in section 2.1, and develop a more detailed hierarchical inductive theory in the area of perception and learning of non-native phonetic categories.

The inductive theory proposed here is an alternative to mapping approaches, providing theoretical unification of native and non-native phonological acquisition. First, it reconceptualizes perceptual reorganization as a problem of hierarchical inductive inference about the structure of the language's sound system: learners not only acquire the specific phonetic category inventory, but also draw higher-order generalizations about the properties of the system as a whole. Second, it offers a novel approach to non-native sound perception and L2 phonetic category learning. In particular, I argue that perception and learning

of non-native sounds is guided by the higher-order generalizations established during perceptual development in infancy, rather than sound-to-sound mappings. Third, the theory provides a way of constructing systematic predictions about the acquisition of phonetic categories in each additional language.

The basic premise of the general hierarchical inductive approach is that—as described in section 2.1—learning involves making simultaneous inductive inferences at multiple levels of abstraction: not only about particular linguistic categories, but also about higher-level category structure. More specifically to the phonological domain, I posit that in addition to learning individual phonetic categories, learners also infer general properties of the set of those categories. Crucially, these inferences occur as part of both perceptual reorganization in infancy and phonological acquisition of any additional language across the life span.

### 2.3.1 Informativity of Phonetic Dimensions

There are many possible higher-order generalizations that learners might make about their native-language sound system. In this dissertation I only examine potential inferences regarding the underlying set of informative phonetic dimensions from which the system is constructed. I begin by discussing the motivation for investigating this type of generalization in particular.

The main reason to search for dimension-based generalizations comes from the long-standing history of successful models in the general categorization literature. We know that people are able to categorize any perceptual stimuli by abstracting information about stimulus dimensions (e.g., color, shape, size, etc.) from single instances of the input (Posner & Keele, 1968; Nosofsky, 1986; Kruschke, 1992).

Within Kruschke's model, learning categories occurs by computing and attaching weights (or attention strength) to each of the stimulus dimensions. The attention strength reflects the relevance (or informativity) of any given dimension for a particular categorization task. That is, high strength will be associated with dimensions hypothesized as the most informative in distinguishing

between categories. This way, people are able to perform categorization tasks by selectively attending to dimensions that are relevant, while at the same time ignoring other dimensions (Nosofsky, 1986). For instance, with stimuli varying along three dimensions such as color, shape, and size, people are good at categorizing by just one dimension, for example color. In this situation, the perceptual space gets stretched along the color dimension—due to high attention strength assigned to this dimension (Kruschke, 1992)—and shrunk along the size and shape dimensions. This strategy is effective in categorization tasks because by attending selectively to the relevant dimension people maximize within-category similarity and between-category discriminability, thus avoiding between-category confusion due to variation along irrelevant dimensions.

As already discussed in section 2.2.1, similar principles may also underlie auditory tuning to native-language sounds: infants have been proposed to selectively attend to the properties of the speech signal that are most informative in interpreting meaningful distinctions in the language (Jusczyk, 1992). The fact that this is analogous to how categorization of any set of perceptual stimuli is understood, whether visual or auditory, is consistent with the idea that auditory perceptual reorganization in infancy emerges as part of domain-general developmental change in cognitive and perceptual competencies (Lalonde & Werker, 1995), such as a general increase in infants' ability to inhibit attention to irrelevant information (Diamond, Werker, & Lalonde, 1994). This indicates that learners—at least infants, but possibly also adults—might have the right cognitive tools to make higher-order generalizations about the informativity of phonetic dimensions, decomposing sounds into dimensions the same way they would any other perceptual stimuli.

Another reason to look at potential dimension-based generalizations comes from previous research on L2 phonological acquisition. As discussed in section 2.2.3, other authors (Hancin-Bhatt, 1994; Brown, 1997, 2000; McAllister et al., 2002) have previously suggested that L2 learners might be sensitive to subsegmental properties (or, more specifically, distinctive features) of L2 sounds. Most prominently, Brown (1997, 2000) has proposed that distinctive features used in L1 can be transferred to L2, facilitating discrimination of any L2 sounds

contrasted by those features. Thus, it has been suggested that learners might generalize features across segments. However, this idea has not been sufficiently pursued to assess whether L1-to-L2 transfer of features in fact takes place. Most importantly, other theories do not provide any underlying explanation for why this type of transfer would even take place. The hierarchical inductive theory proposed in this dissertation, on the other hand, offers such an explanation: the “transfer”—or generalization—of phonetic dimensions is a byproduct of the higher-order generalizations about the dimension informativity, as discussed further in section 2.3.4.

The question remains how exactly the informativity of a phonetic dimension is assessed. The dimensions that infants learn to attend to may include a range of phonetic cues with different degrees of informativity. The most informative dimensions are *contrastive*: that is, they uniquely distinguish between phonemic categories (e.g., VOT differentiating English /b-/p/, /t-/d/, /k-/g/) and are necessary to discriminate lexical items (e.g., *bin-pin*). However, dimensions providing cues other than those critical for pure phonological contrasts—such as secondary cues to phonemic categories, cues distinguishing between allophones, cues indicating talker-specific characteristics, etc.—are also likely encoded as being somewhat informative. This is in line with the proposal by K. C. Hall (2009) that there is gradience in phonological representations reflecting a continuum between contrastive and allophonic distinctions.

Furthermore, the inferred informativity of phonetic dimensions may be affected by other properties of the language’s sound system. One such property is the number of phonetic categories for which a given phonetic dimension is informative. It is possible that if a language has several pairs of categories distinguished by some dimension, then that dimension will be inferred as more informative than a dimension that distinguishes only one category pair.

The exact composition of phonetic dimensions that learners might be sensitive to is at this point unknown: the dimensions might range from subtle acoustic cues, such as formant frequencies, to more abstract ones, such those represented by distinctive features.

### 2.3.2 Perceptual Learning in Adulthood

Based on previous studies on statistical learning with adults I argue that L2 learners are able to extract at least some statistical information from the L2 speech signal, such as the distributional properties of sounds. As a consequence, learners' perceptual system may be modified in a way that sensitivity to phonetic dimensions informative in L2 is gradually increased through L2 exposure.

The extent to which such modifications may be possible is at this point rather unclear. We know that some L2 sound distinctions are notoriously hard to acquire in adulthood, and perceptual acuity seems overall diminished (see section 2.2). On the other hand, not all non-native distinctions are hard to learn, and we know that L2 exposure triggers almost immediate neural changes even before any behavioral improvement in L2 is observed (McLaughlin, Osterhout, & Kim, 2004), and such changes might be preserved relatively long-term, even without any further L2 exposure (Morgan-Short, Finger, Grey, & Ullman, 2012).

Therefore, while it seems possible for an adult perceptual system to adapt to some novel sounds and novel category distinctions with sufficient language exposure, there is likely some perceptual threshold for acoustically subtle contrasts beyond which the system cannot easily adapt. This would lead to general inability to perceive those subtle acoustic cues that are not linguistically informative in the learners' L1, and—as a consequence—difficulty in learning L2 categories that employ those cues. In these cases, individual characteristics of a learner—such as general perceptual abilities, phonological memory, motivation, etc.—may determine the success in learning those hard contrasts.

### 2.3.3 Phonetic Category Induction

Recent work in phonetic category induction has represented native-speaker knowledge of phonetic categories as a set of distributions over perceptual space, one distribution for each phonetic category (de Boer & Kuhl, 2003; Vallabha et al., 2007; McMurray et al., 2009; Feldman et al., 2009). For example, sound categories such as /b/ or /p/ are distinguished using the VOT dimension. Therefore, a native speaker's representation of a category like /b/

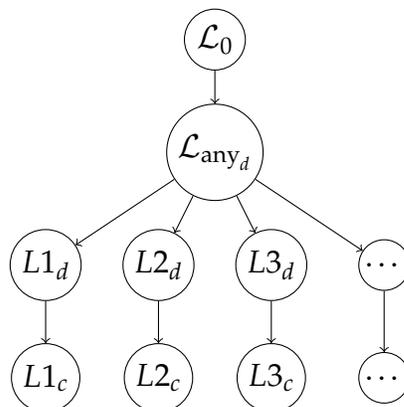
might include some distribution over VOT values, such as a Gaussian distribution with a particular mean and variance. Learners are assumed to acquire the relevant distributions for each phonetic category by tracking the statistical distribution of sounds and phonetic cues in the speech signal, as described in section 2.2.2.

However, from the hierarchical inductive perspective, as outlined in section 2.1, it is natural to expect that the knowledge a learner extracts from a language's sound system includes not only the specific distribution of each sound category but also higher-order generalizations regarding the underlying set of informative dimensions from which the system as a whole is constructed. Thus, while the model presented in this dissertation is not computationally implemented, my proposal is to essentially add a set of higher-level distributions responsible for generating specific category-level distributions.

### 2.3.4 The Hierarchical Inductive Model

In this section I describe in more detail the hierarchical inductive theory as it is applied to perception and acquisition of phonetic categories. The model is intended as a set of principles underlying perceptual development across the life span that bears on three areas: (1) perceptual reorganization in infancy, (2) non-native speech perception by naive listeners, and (3) phonological acquisition of additional languages beyond L1.

The proposed model of a learner's overall knowledge of phonetic categories in multiple languages is schematically illustrated in Figure 2.2. This model is a more specific version of the general model presented in Figure 2.1. The two lower levels now correspond to the knowledge of single learned languages ( $L_1$ ,  $L_2$ , and so on), while the two higher levels again represent the initial learning bias ( $\mathcal{L}_0$ ) and the inferences over individual languages ( $\mathcal{L}_{\text{any}}$ ). The subscripts indicate the type of acquired knowledge:  $c$  refers to the knowledge of specific *categories* in a language, while  $d$  refers to the knowledge regarding the informativity of specific phonetic *dimensions* – either at the level of individual languages or at the higher  $\mathcal{L}_{\text{any}}$  level. Analogous to Kruschke's (1992) model of



**Figure 2.2:** A hierarchical model of phonetic category knowledge in acquisition of multiple languages

category learning, one way of implementing the gradient informativity of phonetic dimensions is by using weights (or attention strength) attached to each dimension.

Let me first discuss the implications of this model for understanding perceptual reorganization in infancy. Within this model, learning phonetic categories in the native language involves not only—as others have proposed—gaining knowledge about statistical distributions of each specific sound category ( $L_{1_c}$ ), but also drawing higher-level inferences about the overall informativity of different phonetic dimensions for discriminating between L1 categories ( $L_{1_d}$ ). These higher-order generalizations are proposed to emerge as a byproduct of learning the specific native-language sound system, and, in particular, the L1 phoneme inventory. The exact nature of  $L_{1_d}$  generalizations will naturally depend on the person’s language background because the specific configurations of informative dimensions vary across languages (Ladefoged & Maddieson, 1996): for example, VOT distinguishes categories in English, but plays no role in Hawaiian; segmental length is contrastive for many Japanese categories (e.g., /p/-/pp/, /t/-/tt/), but is only a secondary cue to some English vowel distinctions (e.g., lax /ɪ/ versus tense /i/) and is relatively uninformative in languages such as Spanish.

Why would infants make such higher-order inferences during acquisi-

tion of the native-language sound system? The answer is that generalizations about phonetic dimension informativity may benefit further learning. Recall from section 2.2.1 that perceptual reorganization involves committing cognitive/neural resources to the variation patterns found in the native language, thus “tuning” to the native-language phonetic categories. I posit that the role of higher-order inductive inferences during perceptual reorganization is to guide this tuning process in order to achieve optimal allocation of resources not only to processing the specific sound patterns encountered in the language so far, but also to processing the patterns that may be encountered during further learning of that language.

Understood this way, perceptual reorganization would yield general perceptual enhancement along dimensions that are highly informative in the native language. For example, if a learner has acquired some phonetic categories distinguished along the VOT dimension (e.g., /p/-/b/, /t/-/d/), then perceptual enhancement should be observed not just for discriminating between those specific categories, but also other potential categories distinguished by the same dimension (e.g., /k/-/g/). Since languages generally reuse phonetic dimensions to distinguish between more than one pair of categories (Ladefoged & Maddieson, 1996), this kind of process would make learning more efficient: due to previously obtained perceptual acuity to a given phonetic dimension, learning a category distinction for one pair of sounds would lead to facilitation in learning analogous categories.

If perceptual reorganization in infancy were affected by these higher-order inferences about the native-language sound system, it would have clear empirical consequences for non-native speech perception by naive listeners. In particular, we should observe enhanced discrimination of non-native contrasts along dimensions inferred from one’s native-language input to be informative. Going back to the VOT example, if VOT distinguishes between some categories in a learner’s native language (e.g., /p/-/b/, /t/-/d/, /k/-/g/), then perceptual enhancement should also be observed for non-native categories distinguished by the same dimension (e.g., /q/-/G/).

My proposal does not, however, limit higher-order generalizations to just

one level. Crucially, in addition to the generalizations at the level of an individual language, learners are proposed to make inferences about the potential informativity of phonetic dimensions *in any language* ( $\mathcal{L}_{any_d}$ ), which then affects the learning of any additional languages. This proposal is based on the assumption that learners are biased to expect languages to share structural similarities, which might follow from the general expectation of similarities between elements within any concept category.

It is an open question when exactly such inferences would emerge during the development of a monolingual infant since one might expect that a learner might need to first recognize that there *are* in fact more languages beyond the one being learned. This touches on the non-trivial issue about how exactly a learner may recognize the existence of more than one language, which possibly varies depending on the ambient culture and the extent of contact with communities or individuals who speak other languages. At the latest, children certainly become aware of the existence of other languages during early school years.

The  $\mathcal{L}_{any_d}$  inferences are proposed to guide learners' implicit expectations, or predictions, about what novel-language perceptual space might look like. One such expectation might be that the space is carved into categories using similar phonetic dimensions as in their native language. Thus, the  $\mathcal{L}_{any_d}$  inferences inform learners' implicit predictions about which phonetic dimensions are likely to be informative in a new language ( $L2_d$ ), and, consequently, predictions about the distributions over the language's specific sound categories ( $L2_c$ ). These predictions are possibly formed before any actual L2 exposure, thus creating an initial perceptual bias for learners to attend to the dimensions that they have inferred as likely to be informative in the L2.

Let me now turn to the phase of actual exposure to a new language. As discussed in section 2.3.2, I posit that adult learners are able to extract from the L2 speech signal the information about sound and phonetic cue distribution, and use it to help them acquire the L2 phonological system. Therefore, when first exposed to an L2, learners can immediately begin to use two sources of information to learn L2 phonetic categories: (1) L1 perceptual biases, which predispose learners to selectively attend to a subset of phonetic dimensions with

highest inferred informativity, and (2) the statistical distribution of sounds in the L2 input. These two sources of information are predicted to interact as learners try to arrive at more precise representations of  $L2_c$  and  $L2_d$ .

Learning any additional languages straightforwardly follows from the theory, and proceeds analogously to learning an L2. Crucially, the knowledge of each previously acquired language—whether at the early or late stage of learning—is expected to inform the  $\mathcal{L}_{\text{any}}$  level. This means that L3 learning benefits from higher-order inferences made on the basis of two languages, and this number increases with each additional language acquired.

In the case of phonetic category learning, the inferences that learners make about phonetic dimension informativity ( $\mathcal{L}_{\text{any}_d}$ ) will be changing as more knowledge is gained regarding the informativity of dimensions in specific languages (i.e., as the number of inferred  $L_d$  nodes increases).

Note that the amount of information contributed by each language learned to the  $\mathcal{L}_{\text{any}}$  level might vary. For example, learning two closely related languages might not be as informative as learning two languages that are typologically distant. This is because learning an L2 that closely resembles the learner’s native language will not provide the learner with much information on how languages can vary. On the other hand, learning languages that are very different will allow a learner to make more precise  $\mathcal{L}_{\text{any}}$  inferences that better approximate the true distribution of patterns across all existing languages. How exactly this might happen is, however, beyond the scope of this dissertation.

## 2.4 Predictions of the Hierarchical Inductive Model

In this section I describe some empirical consequences of the proposal that perceptual development, both in infancy and adulthood, is guided by the principles of hierarchical inductive inference.

### 2.4.1 General Predictions

As an approach to perceptual development, the inductive theory makes clear predictions about the patterns of non-native speech perception by naive listeners, as well as of distributional learning of sounds – which presumably occurs mostly in an early stage of acquisition.

There are many well-known difficulties in non-native speech perception of sound distinctions that do not exist in a learner’s native language, as discussed in section 2.2.1. The inductive theory can account for these difficulties because distinctions not used in a learner’s native language are predicted to also be inferred as uninformative in a novel language. For example, recall the well-known difficulty Japanese monolinguals experience in learning the distinction between English /r/ and /l/, which in the mapping models is explained by listeners mapping both sounds onto a single Japanese category. In the inductive theory, this difficulty is explained by a strong bias arising during the acquisition of Japanese that the acoustic cues distinguishing /r/ and /l/ are not informative for phonetic categorization.

The hierarchical inductive theory crucially also makes predictions regarding potential *benefits* of previous linguistic knowledge that go beyond direct similarities between L1 and L2 phonetic categories. Specifically, there are three main predictions:

1. *Generalization of phonetic dimension informativity across languages*

If a phonetic dimension is highly informative in L1 (strong weight at  $L1_d$ ), a learner will infer that it is likely to be informative in any language (strong weight at  $\mathcal{L}_{any_d}$ ), which will translate into a bias to attend to that dimension when exposed to a new language. As a consequence, discrimination and learning of novel categories distinguished by this dimension is predicted to be facilitated.<sup>5</sup>

---

<sup>5</sup>Note that this is similar to the proposal within Brown’s (1997, 2000) model of L2 speech perception that transfer of phonological distinctive features from L1 to L2 might facilitate discrimination of sounds that are contrasted by those features.

## 2. *Generalization of phonetic dimension informativity within a novel language*

If informativity of a phonetic dimension is acquired from distributional properties of the  $L_n$  speech signal for some categories (strong weight at  $L_{n_d}$ ), then a learner should infer that this dimension is likely to also be informative for other categories in this language. Consequently, discrimination and distributional learning of any additional  $L_n$  categories distinguished by this dimension is predicted to be facilitated.

## 3. *Combining cross-language and within-language generalizations*

If learners incrementally combine information from previously learned languages with statistical information from new language input, then we can expect that both types of information may interact. That is, initial inferences regarding novel phonetic categories, such as distributions over phonetic dimensions ( $L_{n_c}$ ), are predicted to result from combining generalizations based on previous language knowledge ( $\mathcal{L}_{any_d}$ ) with distributional information extracted from the novel speech signal.

The first prediction crucially implies generalization across segment classes from one language to another: that is, enhancements in discrimination of novel sounds may obtain even among non-native percepts for which a given dimension is not ordinarily informative in the native language. For example, as mention in section 2.3.4, if VOT distinguishes between some categories in a learner's native language (e.g., /p/-/b/, /t/-/d/, /k/-/g/), then discrimination should also be enhanced for analogous non-native categories (e.g., /q/-/G/).

Note that this prediction could be made by the mapping theories only insofar as the two non-native sounds can map onto two distinct L1 categories. In the VOT example, discrimination of /q/ vs. /G/ would be good only if we assume that learners systematically map novel /q/ and /G/ onto two separate L1 categories: for example, /q/<sub>L2</sub> onto /k/<sub>L1</sub> and /G/<sub>L2</sub> onto /g/<sub>L1</sub>. This kind of mapping seems very intuitive in this particular example, but in section 2.4.4 I discuss additional specific predictions of the inductive theory, where sound-to-sound mappings are far less straightforward.

The second prediction is analogous to prediction (1), but it applies within a single language. That is, it implies generalization across segment classes from a subset of  $L_n$  categories that has already been learned to not-yet-learned  $L_n$  categories. There is some experimental evidence suggesting that this kind of generalization might be possible: training infants and adults on a novel contrast along the VOT dimension enhances not just discrimination of the trained sounds (e.g., /p/-/b/), but also other sounds with an analogous contrast (e.g., /k/-/g/) (Maye et al., 2008; McClaskey et al., 1983).

The third prediction suggests that the two sources of information— $\mathcal{L}_{any_d}$  biases and  $L_n$  input statistics—may interact during  $L_n$  acquisition. At the early stages of acquiring an additional language a learner might not have received sufficient language input to fully trust the distributional information extracted thus far from the novel speech signal. At that point, a learner might rely more heavily on previously acquired linguistic knowledge, interpreting the statistical regularities encountered in the new language more in line with  $\mathcal{L}_{any_d}$  generalizations. More specifically, learners are predicted to interpret ambiguous acoustic information in favor of phonetic dimensions that they have previously inferred as informative in known languages.

## 2.4.2 Phonetic Dimensions Studied

In this dissertation, two dimensions were chosen to test the predictions of the inductive theory: short vs. long segmental length and alveolo-palatal vs. retroflex place of articulation.

### Length

Length is a convenient property for testing generalization since it is a relatively salient acoustic-phonetic cue (D. A. Hall et al., 2002) that cross-cuts a wide range of possible segments, both vowels and consonants. In many languages length is phonologically contrastive, as shown in (1).<sup>6</sup>

<sup>6</sup>By convention, long consonants are represented as a sequence of two identical segments (e.g., [tt]), while long vowels are represented as a single vowel followed by a colon (e.g., [a:]).

- (1)
- |    |                        |  |
|----|------------------------|--|
| a. | [taka-] vs. [takka]    | Finnish: <i>back</i> vs. <i>fireplace</i>            |
| b. | [kisasi] vs. [kissaki] | Japanese: <i>empress</i> vs. <i>point of a sword</i> |
| c. | [belo] vs. [bello]     | Italian: <i>I bleat</i> vs. <i>beautiful</i>         |
| d. | [seki] vs. [se:ki]     | Japanese: <i>seat</i> vs. <i>century</i>             |

The acoustic correlates of length are different for different classes of segments. For example, length can be signaled by closure duration (stops), duration of frication noise (fricatives), or duration of voicing. Length is not just raw duration, as length contrasts are also signaled by intensity (sonorant consonants and vowels) or burst strength (for stop consonants).

There have been many different proposals regarding formal means of representing length: using the distinctive feature [ $\pm$ long] (Chomsky & Halle, 1968), timing slots (Levin, 1985; Selkirk, 1991; Tranel, 1991; Hume, Muller, & van Engelenhoven, 1997), moras (Hyman, 1985; Hayes, 1989; Davis, 1999), or a combination of the latter two (Muller, 2001). In this dissertation I make no attempt to differentiate among these different analyses. Instead, of interest is that all of these proposals share one commonality: length is represented as abstracted across different segments, despite different raw acoustic cues that signal segmental length differences. Therefore, I assume that it is justified to treat length as an independent phonetic dimension that learners are sensitive to.

### **Alveolo-Palatal vs. Retroflex Place of Articulation**

Alveolo-palatal and retroflex sibilant consonants are known to exist in relatively few languages, including Polish and Mandarin (Ladefoged & Maddieson, 1996). This distinction is acoustically very subtle, and the main cues to the contrast include spectral properties of the frication noise and formant transition onto the following vowel (Ladefoged & Maddieson, 1996; Nowak, 2006; Li, Edwards, & Beckman, 2007; Li, 2008). The specific distinction chosen for the purposes of this dissertation was the alveolo-palatal vs. retroflex contrast from Polish. Polish has four alveolo-palatals (fricatives: / $\mathfrak{c}$ /, / $\mathfrak{z}$ /; affricates: / $\mathfrak{t}\mathfrak{c}$ /, / $\mathfrak{d}\mathfrak{z}$ /) and four retroflexes (fricatives: / $\mathfrak{s}$ /, / $\mathfrak{z}$ /; affricates: / $\mathfrak{t}\mathfrak{s}$ /, / $\mathfrak{d}\mathfrak{z}$ /). Half of them are voiceless (/ $\mathfrak{c}$ /, / $\mathfrak{t}\mathfrak{c}$ /, / $\mathfrak{s}$ /, / $\mathfrak{t}\mathfrak{s}$ /), and half are voiced (/ $\mathfrak{z}$ /, / $\mathfrak{d}\mathfrak{z}$ /, / $\mathfrak{z}$ /, / $\mathfrak{d}\mathfrak{z}$ /). Some authors describe Polish retroflexes as postalveolars (Jassem, 2003). How-

ever, I follow the arguments that they are better described as slightly retroflex or retracted (P. A. Keating, 1991; Ladefoged & Maddieson, 1996; Hamann, 2004). For the purposes of this dissertation I treat the collection of acoustic cues to the alveolo-palatal vs. retroflex place of articulation contrast as a unified phonetic dimension (henceforth, *place* dimension).

### 2.4.3 Properties of the Languages Studied

Five groups of speakers were chosen for experiments included in this dissertation based on their language background, as shown in (2). In this section I briefly describe the inventory of each language with respect to short and long, as well as alveolo-palatal and retroflex segments.

- (2)
- a. Chapter 3: speakers of Korean, Vietnamese, Cantonese (with some knowledge of Mandarin), and Mandarin (all bilingual in English)
  - b. Chapter 4: speakers of Korean and Mandarin (all bilingual in English)
  - c. Chapter 5: speakers of English (monolingual)
  - d. Chapter 6: speakers of Korean and Mandarin (all bilingual in English)

#### Korean

Korean uses length to distinguish between all vowels (e.g., [pul] ‘fire’ vs. [pu:l] ‘blow’; Lee, 1999). There are only a few lexical items with underlying monomorphemic long consonants (e.g., [p̄alle] ‘worm’; Kim, 2002). More common are morphologically derived long consonants ([ll], [nn], [mm]), which arise from phonological assimilation processes (Sohn, 1999). In addition, Korean tense obstruents ([p̄], [t̄], [k̄], [s̄], [t̄ɕ]) have sometimes been analyzed as long (Choi, 1995).

Korean does not have an alveolo-palatal vs. retroflex distinction. The inventory includes an alveolo-palatal fricative [ç] that is an allophone of /s/, but no retroflex sounds (Hahm, 2007; Sohn, 1999).

## Vietnamese

In Vietnamese, length is phonologically contrastive for two sets of vowels: [a]-[a:] and [ə]-[ə:] (e.g., [bang] ‘state’ vs. [ba:ng] ‘ice’), although the latter have been argued to also differ in vowel quality (Winn et al., 2008). Consonants are always short.

Vietnamese does not have alveolo-palatal or retroflex sounds.

## Cantonese

Length is used in Cantonese to distinguish between vowel categories, but it is generally only one of the cues in addition to distinctions in vowel quality, and all but one short–long vowel pairs are in complementary distribution (Bauer & Benedict, 1997). However, the one pair occurring in the same contexts ([ɐ]-[a:]) is distinguished almost exclusively by length, with only minimal quality differences (Zhang, 2011), and length has been shown to be the primary cue for distinguishing other vowel pairs as well (Kao, 1971; Bauer & Benedict, 1997). Consonants are always short.

Cantonese does not have retroflex sounds, but alveolar sibilants ([ts], [ts<sup>h</sup>], [s]) can be palatalized to the alveolo-palatal place of articulation ([tɕ], [tɕ<sup>h</sup>], [ç]), especially before high front vowels (Bauer & Benedict, 1997).

## Mandarin Chinese

Mandarin does not have segmental length contrasts. Mandarin tones vary in length, and some listeners have been reported to use length to distinguish between tones when the main cue—the F0 pattern—is ambiguous (Tseng, Massaro, & Cohen, 1986; Blicher, Diehl, & Cohen, 1990). At the same time, Mandarin speakers have been found not to rely on length for non-native segmental contrasts on vowels (Bohn, 1995).

Mandarin has voiceless alveolo-palatals (/ç/, /tɕ/) and retroflexes (/ʂ/, /tʂ/) as allophones of the same phonetic category: alveolo-palatals occur before high front vowels and the palatal glide, and retroflexes occur elsewhere (Lin, 2001). In addition, the voiced retroflex fricative ([ʐ]) is a between-speaker vari-

ant of the retroflex approximant ([ɻ]). Other voiced sibilants are assumed to be absent because Mandarin has obstruent distinctions in aspiration, not voicing (Lin, 2001). Note that the alveolo-palatal and retroflex sounds used as stimuli in this dissertation are from Polish, not Mandarin. Polish and Mandarin alveolo-palatals and retroflexes differ, but they share similar spectral cues that distinguish between the two (Ladefoged & Maddieson, 1996).

## English

American English does not use length contrastively. Vowel length varies, but it correlates with the tense-lax distinction (e.g., *beat* vs. *bit*) and the voicing of the following segment (e.g., *cad* vs. *cat*). The differences in vowel length alone never distinguish between words, and length has not been argued to be the primary cue to vowel distinctions. Long consonants are sometimes attested but only at morpheme boundaries (e.g., *dissatisfied*; Benus, Smorodinsky, & Gafos, 2003). Minimal pairs are rare (e.g., *unnamed* vs. *unaimed*), and for most speakers the contrast is neutralized (Kaye, 2005). Furthermore, there is evidence that by 18 months of age English-learning infants process length contrasts differently from infants learning a language that uses length contrastively (e.g., Dutch or Japanese; Dietrich, Swingley, & Werker, 2007; Mugitani, Pons, Fais, Werker, & Amano, 2008). Therefore, even though length plays some role in the American English sound system, for the purposes of this dissertation I treat English as a language where length is not informative, and I leave for future research investigating the extent to which English might provide learners with evidence of length informativity (cf. Mermelstein, 1978; Whalen, 1989; McAllister et al., 2002).

English does not have alveolo-palatal nor retroflex obstruents, although some speakers produce the alveolar approximant /ɻ/ as retroflex (Ladefoged & Maddieson, 1996; Westbury, Hashi, & Lindstrom, 1998).

## Summary

Table 2.1 summarizes the relevant properties of the surface (i.e., not just phonemic) segmental inventories in Korean, Vietnamese, Cantonese, Mandarin, and English.

**Table 2.1:** Summary of the surface segmental inventories in English, Korean, Vietnamese, Cantonese, and Mandarin: length and alveolo-palatal vs. retroflex place of articulation.

SEGMENTS		KOR	VIET	CANT	MAND	ENG
<i>Length</i>	consonants	<i>short</i>	✓	✓	✓	✓
		<i>long</i>	✓			
	vowels	<i>short</i>	✓	✓	✓	✓
		<i>long</i>	✓	✓	✓	
<i>Place</i>	<i>alveolo-palatals</i>				✓	
	<i>retroflexes</i>				✓	

### 2.4.4 Specific Predictions

In this section I describe specific predictions of the inductive theory for perception and learning of the length and place dimensions by speakers of Korean, Vietnamese, Cantonese, Mandarin, and English. Empirical tests of each prediction outlined in this section are provided in subsequent chapters of the dissertation.

#### L1 Perceptual Biases in Non-Native Speech Perception (Chapter 3)

The first general prediction discussed in section 2.4.1 states that if a phonetic dimension is informative in L1, then a learner should infer it as likely to also be informative in L2, and—as a consequence—attend to this dimension when exposed to the new language, even for percepts for which this dimension is not relevant in L1. I will now illustrate this prediction for the dimension of segmental length.

Imagine three languages:  $L_A$  in which length is contrastive for both vowels and consonants,  $L_B$  in which length is contrastive only for vowels, and  $L_C$  in which length is never contrastive, as illustrated in Table 2.2.

**Table 2.2:** Hypothetical language inventories.

SEGMENTS		$L_A$	$L_B$	$L_C$
<i>short</i>	a	✓	✓	✓
	i	✓	✓	✓
	s	✓	✓	✓
<i>long</i>	a:	✓	✓	
	i:	✓	✓	
	ss	✓		

Now imagine that we are testing speakers of all three languages on discrimination of two contrasts in a novel language: [a]–[a:] and [s]–[ss]. Both mapping and hierarchical inductive theories predict that native speakers of  $L_A$  and  $L_B$  will outperform speakers of  $L_C$  in [a]–[a:] contrast discrimination, since that contrast is present in both  $L_A$  and  $L_B$ , and—for the same reason—that speakers of  $L_A$  will outperform those of  $L_C$  in [s]–[ss] discrimination. The inductive theory, however, additionally predicts that speakers of  $L_B$  will outperform those of  $L_C$  in [s–[ss] discrimination, because there is evidence in  $L_B$  that segmental length is a generally informative dimension (Tab. 2.3).

**Table 2.3:** Example predictions for discrimination performance on non-native length contrasts by native speakers of  $L_A$ ,  $L_B$ , and  $L_C$ .

$L_{novel}$ CONTRAST	INDUCTIVE THEORY	MAPPING THEORIES
a–a:	$L_A, L_B > L_C$	$L_A, L_B > L_C$
s–ss	$L_A, L_B > L_C$	$L_A > L_B, L_C$

Following this simple example, we can now outline predictions for speakers of real languages. Recall from section 2.4.3 that Korean has length contrasts for both vowels and consonants, Vietnamese and Cantonese have length contrasts only for vowels, while Mandarin does not have any length contrasts. Therefore, if tested on discrimination of length contrasts in a novel language, the inductive theory predicts that speakers of the first three languages (Korean, Vietnamese, Cantonese) should outperform speakers of Mandarin. Crucially, even if tested only on *consonant* length contrasts, speakers of all three languages should have an advantage over Mandarin speakers because they should have inferred length as a dimension that is likely to be informative in a new language.

Mapping theories, on the other hand, only predict some advantage for Korean speakers who are familiar with some length contrasts on consonants – an advantage that is limited in that it should only obtain for discrimination of those specific length contrasts that are found in Korean.

Similarly, we can make predictions for the alveolo-palatal vs. retroflex place contrast. Here, both theories predict that Mandarin speakers should be better at discriminating the place contrast due to the fact that Mandarin has this distinction, while other languages do not.

All these predictions are illustrated in Table 2.4, and Chapter 3 describes experimental work that tests these claims. The results show that, as predicted by the inductive theory, speakers of Korean, Vietnamese, and Cantonese all outperform Mandarin speakers on discrimination of consonant length contrasts, while Mandarin speakers are better at discrimination of alveolo-palatal vs. retroflex place contrasts. Thus, the results suggest that there is generalization of phonetic dimension informativity (length) across very dissimilar segment classes (from vowels to consonants).

**Table 2.4:** Predictions for discrimination of length and place contrasts by speakers of different native languages.

CONTRAST	INDUCTIVE THEORY	MAPPING THEORIES
<i>consonant length</i>	<i>Kor, Viet, Cant &gt; Mand</i>	<i>Kor &gt; Viet, Cant, Mand</i>
<i>alv. pal. – retroflex place</i>	<i>Kor, Viet, Cant &lt; Mand</i>	<i>Kor, Viet, Cant &lt; Mand</i>

### L1 Perceptual Biases in Word Learning (Chapter 4)

The predictions regarding perceptual advantages on length and place contrasts can be extended to the domain of lexical acquisition. Specifically, if speakers of Korean and Mandarin are differentially sensitive to length and place distinctions (i.e., if Korean speakers are perceptually biased toward length contrasts, and Mandarin speakers are biased toward place contrasts), then similar differences should be observed when learning words in a new language that are minimal pairs distinguished only by length and place. Chapter 4 describes experimental work that tested this prediction. The results reveal

that participants were not able to fully capitalize on their perceptual abilities: only faster learners—as independently assessed by baseline trials—showed enhanced learning involving contrasts in phonetic dimensions that are informative in their native language. This suggests that attention to phonetic detail when learning words might only be possible for adults with better cognitive abilities or higher motivation.

### **Overcoming L1 Perceptual Biases in Phonetic Category Learning (Chapter 5)**

The second prediction of the inductive theory discussed in section 2.4.1 states that if a learner successfully learns a novel L2 contrast for one set of segments, then this should facilitate discrimination and learning of additional L2 categories that vary along the same phonetic dimension. I will illustrate this prediction using the dimension of segmental length.

English does not use length to contrast phonetic categories, and therefore native speakers of English should be perceptually biased against distinctions along the length dimension. Learning a novel language with contrastive length thus requires overcoming initial L1 biases. However, once the distinction is learned for some segments in the new language (e.g., [f]–[ff] and [s]–[ss]), then we should observe generalization to other classes of segments (e.g., [n]–[nn], [j]–[jj]) because learners should be able to infer the informativity of length based on the newly-acquired phonetic categories.

Chapter 5 describes two experiments that tested this prediction. In particular, native speakers of English were exposed to a new language, in which the statistical distribution of sounds suggested a phonetic length distinction. They were subsequently tested on categorization of short and long sounds on both a trained segment class and a novel, dissimilar, segment class. The results revealed that—given successful learning of short vs. long categories for the trained segments—participants are also inclined to categorize novel segments as either short or long. This suggests that there is generalization of phonetic dimension informativity across segment classes within a new language that is being learned, thus providing additional support for the inductive theory.

## **Combining L1 Perceptual Biases and Distributional Cues in Phonetic Category Learning (Chapter 6)**

The third prediction of the inductive theory outlined in section 2.4.1 states that learners' interpretation of novel language statistics may be affected by their L1 perceptual biases. This prediction was tested in a study with Korean and Mandarin speakers, as described in Chapter 6. Given known perceptual sensitivities of these two language groups to length and place contrasts (i.e., Korean speakers are more sensitive to length, and Mandarin speakers are more sensitive to place), it is possible that native speakers of these languages would show differential biases when interpreting ambiguous distributional cues to L2 phonetic categories.

To test this possibility, we constructed stimuli in a novel language with sound distributions that could be interpreted as either a place distinction (alveolo-palatal vs. retroflex) or a length distinction (short vs. long). Korean and Mandarin speakers were first exposed to these distributions, and then tested on phonetic categorization of sounds in this new language. The prediction was that speakers of Korean should be biased toward inferring a length-based category distinction and against inferring a place-based category distinction, and thus interpret this phonetic input as two categories along the length dimension. Speakers of Mandarin, on the other hand, should be biased toward inferring a place-based category distinction and against inferring a length-based category distinction, thus interpreting the input as two categories along the place dimension. The results revealed that Mandarin speakers indeed inferred word distinctions more based on place than length, but speakers of Korean inferred distinctions based on both, suggesting that they might have hypothesized four underlying categories instead of two. This result indicates that the interpretation of distributional information in L2 speech is indeed affected by L1 perceptual biases, providing initial support for the claim that learners incrementally combine two sources of information when learning a new language: L1 biases and L2 statistics.

# Chapter 3

## L1 Perceptual Biases in Non-Native Speech Perception

### 3.1 Introduction

This chapter tests the first prediction outlined in section 2.4.1 stating that if a phonetic dimension is informative in L1, then a learner should infer it as likely to also be informative in L2, and—as a consequence—attend to this dimension when exposed to the new language, even for percepts for which this dimension is not relevant in L1.

This prediction was tested in a perceptual discrimination experiment that involved discriminating consonant length contrasts and alveolo-palatal vs. retroflex place contrasts. We recruited participants of different language backgrounds: 1) Korean, where length is a highly relevant contrastive cue for both vowels and consonants, and where there are no alveolo-palatal vs. retroflex place contrasts; 2) Vietnamese and Cantonese, where the length cue is informative, but more limited (only for vowels, and—for Cantonese—only as an additional cue together with changes in vowel quality), and where there are no alveolo-palatal vs. retroflex place contrasts; and 3) Mandarin Chinese, where length is not very informative, but where there is an allophonic distinction between alveolo-palatals and retroflexes (as described in more detail in section 2.4.3).

As discussed in section 2.4.4, the inductive theory predicts that all length-attuned participants would outperform Mandarin speakers on discriminating between short and long segments, but a reverse performance pattern is expected for the place contrast: Mandarin speakers outperforming other participants on discriminating between alveolo-palatals and retroflexes. Additionally, the theory predicts that Korean speakers might show an advantage over Vietnamese and Cantonese speakers on length contrasts due to stronger bias in favor of length.

## 3.2 Experiment

### 3.2.1 Method

#### Participants

96 undergraduate students at UC San Diego participated in the experiment for course credit. Each participant was from one of four language groups: Korean, Vietnamese, Cantonese (all *length-familiar*, and Mandarin (*place-familiar*). All Cantonese speakers also spoke some Mandarin, which they learned at school. All participants learned the target language (L1) from birth, and were bilingual in English. There were no major differences between language groups on other characteristics (see Tables 3.2–3.5 in the appendix to this chapter), and the tested population included both L1-dominant and English-dominant participants. Participants reported no history of speech or hearing problems.

#### Materials

The materials consisted of nonce words recorded in a soundproof booth by a phonetically-trained native speaker of Polish. The *length* items included short and long consonants, and the *place* items were Polish alveolo-palatal and retroflex consonants (see Table 3.1). Each sound segment was recorded embedded in seven different frames: [pa\_a], [pe\_a], [po\_a], [ta\_a], [te\_a], [ka\_a], [ke\_a], with five repetitions of each word. Then, the stimuli were manipulated through

**Table 3.1:** Sound segments used in stimuli, and the occurrences of corresponding sounds in Korean, Vietnamese, Cantonese, and Mandarin.

SEGMENT		KOR	VIET	CANT	MAND	
<i>Length stimuli</i>	<i>short</i>	m	✓	✓	✓	✓
		n	✓	✓	✓	✓
		l	✓	✓	✓	✓
		s	✓	✓	✓	✓
		j	✓	✓	✓	✓
		w	✓	✓	✓	✓
		f		✓	✓	✓
	<i>long</i>	mm	✓			
		nn	✓			
		ll	✓			
		ss	✓			
		jj				
		ww				
		ff				
	(vowels) <sup>a</sup>	(✓)	(✓)	(✓)		
<i>Place stimuli</i>	<i>alv.-pal.</i>	ɕ	✓			✓
		tɕ				✓
		ʐ				
		dz				
	<i>retroflex</i>	ʂ				✓
		tʂ				✓
		ʐ̥				✓
	dz̥					
<i>Filler stimuli</i>	x		✓	✓	✓	
	ɣ					
	ʙ					
	j					

<sup>a</sup> Long vowels were not included in the stimuli.

splicing to ensure that the minimal-pair words differed only in length or place, with no irrelevant differences present elsewhere in a word. One token of each word type was chosen as a frame (a short-consonant word for length contrasts and a retroflex-word for place contrasts), from which the target segments were removed. Then, the missing segments were spliced out from other recorded tokens and placed in the frame words.<sup>1</sup> For *length* items, only short segments were spliced in, and long segments were created from short ones by either doubling their length (for sonorant consonants: [j], [w], [l], [m], and [n]) or elongating them by half their length (for obstruent consonants: [f] and [s]).<sup>2</sup>

### Procedure

The experiment consisted of a same-different AX discrimination task. In each trial, a pair of words was presented auditorily over headphones. The words were either ‘different’ (e.g., [pama]–[pamma]) or ‘same’ (e.g., [pama]–[pama] or [pamma]–[pamma]). ‘Same’ words in each pair were physically identical and ‘different’ words in each pair always shared a physically identical frame (i.e., the words were identical except for artificial lengthening for length contrasts and a spliced consonant for place contrasts). This was done to ensure that ‘different’ responses resulted only from the manipulation of interest, and not due to irrelevant differences present elsewhere in a word. The words in each pair were separated by an interstimulus interval of 750msec to ensure processing of sounds at a higher, non-sensory level (Werker & Logan, 1985). Each pair was repeated twice throughout the experiment, which yielded a total of 392 pairs (196 pairs with length contrasts and 196 pairs with place and filler contrasts), divided into seven 56-trial blocks that were separated by self-terminated breaks. In each trial, a word pair was played once without a replay option, and the response to one pair triggered presentation of the subsequent pair with a delay of 500msec.

---

<sup>1</sup>The use of splicing meant that one of the cues to the place contrast—vowel transition of alveolo-palatals—was partially removed, possibly making this contrast harder than in natural speech.

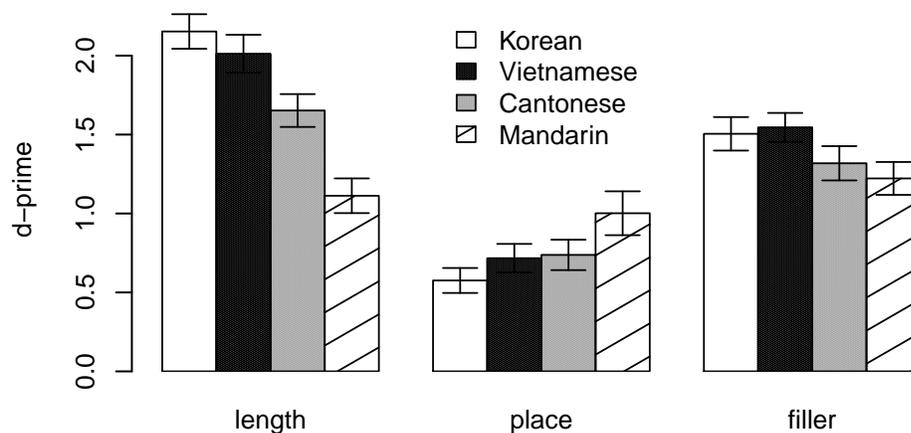
<sup>2</sup>This difference was introduced to mimic natural production, reflecting the fact that inter-vocalic length contrasts are perceptually harder for sonorants than for obstruents due to more blurred segment boundaries (Kawahara, 2007).

Trial order was randomized for every participant. The testing was preceded by a no-feedback practice session with 16 filler trials.

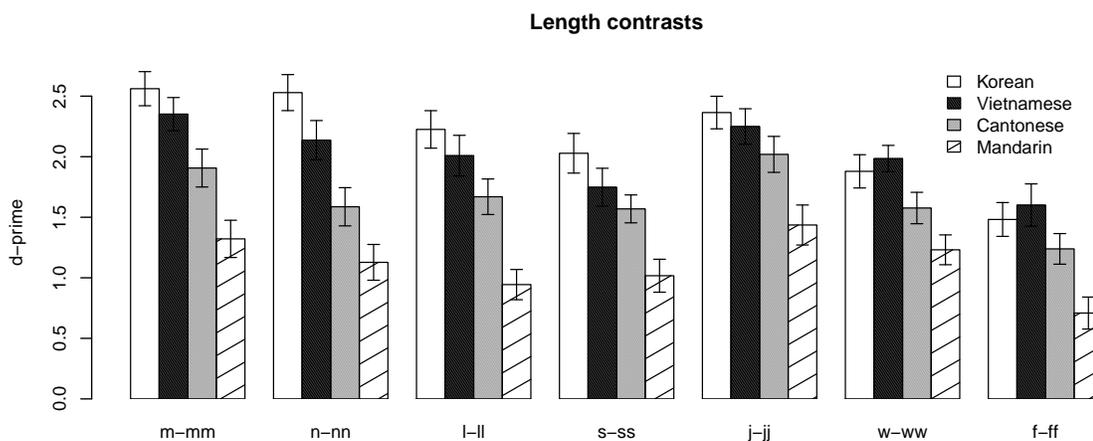
### 3.2.2 Results

The results are plotted in Figures 4.2 and 3.2 (error bars are standard errors). We calculated  $d$ -prime scores (Swets, 1964) for each tested contrast and each participant as a measure of contrast sensitivity, and analyzed the scores using repeated-measures ANOVAs. First, we compared length-familiar participants to place-familiar participants using an ANOVA with the factors LANGUAGE GROUP (*length-familiar, place-familiar*) and CONTRAST (*length, place*). As we predicted, there was a significant interaction between LANGUAGE GROUP and CONTRAST [ $F(1, 94) = 51.7; p < .001$ ]: length-familiar participants were more sensitive to length differences, and place-familiar participants were more sensitive to place contrasts. The results also revealed a main effect of LANGUAGE GROUP [ $F(1, 94) = 6.5; p < .05$ ]: Mandarin speakers performed overall worse than length-familiar participants as a group. The main interaction was not, however, driven by the overall worse performance by Mandarin speakers, since the result was reversed for place contrasts (also supported by a significant interaction between LANGUAGE GROUP and *place vs. filler* CONTRAST [ $F(1, 94) = 24.0; p < .001$ ]), and the differences between the two LANGUAGE GROUPS on length and on filler contrasts were of different magnitudes (as indicated by a significant interaction between LANGUAGE GROUP and *length vs. filler* CONTRAST [ $F(1, 94) = 33.4; p < .001$ ]).

Crucially, the main result was not driven just by Korean performance, but also held for each relevant pairwise LANGUAGE comparison, as indicated by significant interactions between LANGUAGE and CONTRAST (Korean–Mandarin: [ $F(1, 46) = 63.7; p < .001$ ]; Vietnamese–Mandarin: [ $F(1, 46) = 33.1; p < .001$ ], Cantonese–Mandarin: [ $F(1, 46) = 18.7; p < .001$ ]). The results reveal an extremely robust pattern: Korean, Vietnamese, and Cantonese speakers were consistently better at length contrasts than Mandarin speakers for each tested segment (see Figure 3.2).



**Figure 3.1:** Overall results.



**Figure 3.2:** Results from the critical *length* trials split by segment.

As for the comparisons within the length-familiar group, there was no significant difference on length contrasts for the Korean–Vietnamese pair [ $F < 1$ ], but there was a significant difference for both Korean–Cantonese [ $F(1,46) = 11.0; p < .01$ ] and Vietnamese–Cantonese [ $F(1,46) = 5.1; p < .05$ ], with Cantonese speakers performing worse. Given that length is only used in Cantonese as a secondary cue, this result is consistent with the idea that sensitivity to a given phonetic dimension is mediated by the degree of informativity that this dimension has in the learner’s native language.

### 3.3 Discussion

The results reported here are consistent with the predictions of the hierarchical inductive theory: Korean, Vietnamese, and Cantonese speakers were all better at discriminating between short and long consonants than Mandarin speakers, even though Vietnamese and Cantonese only use length distinctions on vowels. This suggests that they generalized a familiar phonetic dimension across different segment classes. This result cannot be attributed to better task performance, as the pattern was reversed for sibilant contrasts, which are more familiar to Mandarin speakers.

The mapping theories of non-native speech perception and learning, as described in section 2.2.3, have not explicitly stated any predictions regarding the dimensions of length and alveolo-palatal vs. retroflex place. However, given their theoretical assumptions, they are likely to predict a different response pattern than the inductive theory. Specifically, mapping theories predict that participants should only have a clear advantage on discrimination of contrasts in cases when the two contrasted sounds would be perceptually mapped onto two different native-language categories. Recall that, on these views, mapping is determined by acoustic or articulatory similarity between sounds, and is also assumed to reflect learners' perceived between-sound similarity. That is, as learners' perceived similarity between a novel L2 sound and an L1 sound category increases, the two sounds become more perceptually confusable, and the probability of the L2 sound being assimilated onto the L1 category rises. Following this reasoning, mapping theories predict enhanced discrimination of length and place contrasts only insofar as they are perceived as similar to two distinct native categories.

For place contrasts, this could yield similar predictions to those of inductive theory: Mandarin speakers should be better than speakers of Korean, Vietnamese, and Cantonese, because only Mandarin has alveolo-palatal and retroflex sounds, which are similar to those included in the stimuli. Therefore, Mandarin speakers could be said to map each novel alveolo-palatal sound onto an alveolo-palatal category in Mandarin, and each novel retroflex sound onto a

retroflex category in Mandarin. Speakers of Korean, Vietnamese, and Cantonese would, on the other hand, be likely to map both alveolo-palatals and retroflexes onto a single sibilant category in their native languages.

For length contrasts, however, mapping theories cannot straightforwardly explain the advantage of Korean, Vietnamese, and Cantonese speakers over Mandarin speakers. The only group that would be predicted to perform somewhat better on consonant length contrasts are speakers of Korean: since Korean is the only language where some long consonants are present—and only if counted generously (see section 2.4.3)—it can be said that Korean speakers mapped short consonants onto similar short categories in Korean, and long consonants onto similar long categories, which would yield enhanced discrimination. However, in order to account for the result that speakers of Vietnamese and Cantonese outperform Mandarin speakers on consonant length contrasts, these theories would have to assume that the Vietnamese/Cantonese vs. Mandarin speaker groups systematically differ in how they map non-native long consonants onto their native-language categories. Yet, there is no obvious reason why this might be the case other than the presence of long vowels in Vietnamese/Cantonese, as already argued in this dissertation. One could posit that, for Vietnamese/Cantonese speakers, native-language long vowel categories pull long consonant percepts toward them, thus facilitating short vs. long consonant discrimination. However, it would then follow that Vietnamese/Cantonese speakers should also perceptually confuse long consonants and long vowels (e.g., [ss] and [a:]) more so than Mandarin speakers, which seems highly unlikely.

Therefore, the results reported here suggest that learners generalize the informativity of phonetic dimensions across languages (from L1 to a novel language), and also across very dissimilar segment classes (vowels and consonants). These results are consistent with predictions of the inductive theory proposed in this dissertation, and are not easily explained by mapping theories.

### 3.4 Chapter Appendix

This appendix includes detailed individual characteristics of the participants from the study reported in this chapter.

**Table 3.2:** Mean and standard deviation of self-reported participant characteristics: Korean speakers.

KOREAN SPEAKERS	L1-dominant (n=12)		Eng-dominant (n=12)	
	M	SD	M	SD
Age	21	3.0	20	1.0
Age of arrival in US <sup>a</sup>	14	5.3	4	5.3
Self rated L1 proficiency <sup>b</sup> (0-none, 10-perfect)	9.1	1.0	7.4	2.1
% time current L1 exposure	44	13.8	34	17.1
L1 use w/family (0-never, 10-always)	9.9	0.3	8.6	1.5
L1 use w/friends (0-10)	6.8	1.7	4.0	2.3
% time preferred L1 use <sup>c</sup>	61	15.6	35	21.5
Age when began regular Eng exposure	10	4.7	4	3.8
Self rated Eng proficiency <sup>b</sup> (0-10)	7.3	1.1	9.4	0.8
% time current Eng exposure	53	13.6	65	16.0
Eng use w/family (0-10)	1.0	1.76	4.3	2.4
Eng use w/friends (0-10)	5.8	2.2	9.0	1.6
% time preferred Eng use <sup>c</sup>	35	16.2	63	20.9

<sup>a</sup> If born in the US, coded as 0.

<sup>b</sup> Mean proficiency speaking & understanding.

<sup>c</sup> "If you could freely choose a language to speak, what percentage of time would you choose to speak each language?"

**Table 3.3:** Mean and standard deviation of self-reported participant characteristics: Vietnamese speakers.

VIETNAMESE SPEAKERS	L1-dominant (n=12)		Eng-dominant (n=12)	
	M	SD	M	SD
Age	20	1.2	20	1.4
Age of arrival in US <sup>a</sup>	5	4.5	0	0.3
Self rated L1 proficiency <sup>b</sup> (0-none, 10-perfect)	8.1	0.9	7.0	0.9
% time current L1 exposure	25	14.7	20	12.2
L1 use w/family (0-never, 10-always)	8.7	1.8	8.1	1.8
L1 use w/friends (0-10)	2.6	2.5	1.7	1.8
% time preferred L1 use <sup>c</sup>	46	9.4	26	14.0
Age when began regular Eng exposure	6	3.5	3	1.8
Self rated Eng proficiency <sup>b</sup> (0-10)	7.7	0.49	9.4	0.6
% time current Eng exposure	74	15.1	77	13.5
Eng use w/family (0-10)	3.6	2.9	6.1	2.1
Eng use w/friends (0-10)	9.1	1.5	9.7	0.6
% time preferred Eng use <sup>c</sup>	55	10.3	69	15.6

<sup>a</sup> If born in the US, coded as 0.

<sup>b</sup> Mean proficiency speaking & understanding.

<sup>c</sup> "If you could freely choose a language to speak, what percentage of time would you choose to speak each language?"

**Table 3.4:** Mean and standard deviation of self-reported participant characteristics: Cantonese speakers.

CANTONESE SPEAKERS	L1-dominant (n=12)		Eng-dominant (n=12)	
	M	SD	M	SD
Age	20	0.9	20	1.5
Age of arrival in US <sup>a</sup>	14	4.4	2	4.4
Self rated L1 proficiency <sup>b</sup> (0-none, 10-perfect)	9.2	1.3	8.3	1.2
% time current L1 exposure	41	24.6	33	8.9
L1 use w/family (0-never, 10-always)	9.1	1.9	9.6	0.8
L1 use w/friends (0-10)	7.1	3.1	4.0	2.1
% time preferred L1 use <sup>c</sup>	58	25.1	33	9.4
Age when began regular Eng exposure	5	2.8	4	1.7
Self rated Eng proficiency <sup>b</sup> (0-10)	7.3	1.0	9.2	0.9
% time current Eng exposure	45	22.9	57	11.4
Eng use w/family (0-10)	1.1	1.2	4.0	2.9
Eng use w/friends (0-10)	7.3	2.1	8.6	2.0
% time preferred Eng use <sup>c</sup>	28	20.3	51	19.6

<sup>a</sup> If born in the US, coded as 0.

<sup>b</sup> Mean proficiency speaking & understanding.

<sup>c</sup> "If you could freely choose a language to speak, what percentage of time would you choose to speak each language?"

**Table 3.5:** Mean and standard deviation of self-reported participant characteristics: Mandarin speakers.

MANDARIN SPEAKERS	L1-dominant (n=12)		Eng-dominant (n=12)	
	M	SD	M	SD
Age	20	0.9	20	1.5
Age of arrival in US <sup>a</sup>	12	4.4	4	4.3
Self rated L1 proficiency <sup>b</sup> (0-none, 10-perfect)	9.4	1.0	8.1	1.0
% time current L1 exposure	46	14.7	16	11.8
L1 use w/family (0-never, 10-always)	9.6	0.9	8.6	1.7
L1 use w/friends (0-10)	5.5	3.2	2.8	2.3
% time preferred L1 use <sup>c</sup>	56	24.7	26	15.1
Age when began regular Eng exposure	8	3.6	5	2.7
Self rated Eng proficiency <sup>b</sup> (0-10)	7.3	0.7	9.0	1.1
% time current Eng exposure	50	14.7	81	18.6
Eng use w/family (0-10)	2.3	1.9	3.3	2.1
Eng use w/friends (0-10)	7.4	1.9	9.6	0.7
% time preferred Eng use <sup>c</sup>	43	24.2	73	15.28

<sup>a</sup> If born in the US, coded as 0.

<sup>b</sup> Mean proficiency speaking & understanding.

<sup>c</sup> "If you could freely choose a language to speak, what percentage of time would you choose to speak each language?"

### 3.5 Acknowledgements

Chapter 3 is a revised version of Pajak (2010) [“Perceptual advantage from generalized linguistic knowledge”. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 369–374). Austin, TX: Cognitive Science Society], and it includes material submitted for publication [Pajak & Levy. “Inductive generalization over phonetic categories in perceptual reorganization”]. The dissertation author was the primary investigator and author of these papers. In addition to being presented to the Cognitive Science Society, this work was also presented at the 84th Annual Meeting of the Linguistic Society of America, and at the 51st Annual Meeting of the Psychonomic Society.

# Chapter 4

## L1 Perceptual Biases in Word Learning

### 4.1 Introduction

The study in Chapter 3 provided evidence that non-native speech perception is guided by perceptual biases that are the result of generalization from previous language knowledge. Specifically, it was found that speakers of Korean, Vietnamese, and Cantonese outperform speakers of Mandarin on length contrast discrimination, but the reverse is true for alveolo-palatal vs. retroflex place contrast discrimination. In this chapter these initial findings are extended to the domain of lexical acquisition: if speakers of Korean and Mandarin are differentially sensitive to length and place distinctions, then similar differences might be observed in a more natural learning situation, for example when learning words in a new language that are minimal pairs distinguished only by length and place.

We know that humans are able to take advantage of many different resources available to them in the course of learning. For example, when learning a new language—whether in infancy or adulthood—humans actively search for regularities by analyzing the input in several alternative ways (e.g., examining either adjacent or non-adjacent dependencies; Gómez, 2002), and are able to simultaneously entertain multiple implicit theories about the input's under-

lying structure (e.g., Gerken, 2010). One of the complex features of learning a language is that it is necessary to perform concurrent analyses of the input at different levels of processing and integrate these multiple pieces of information at once. If, for example, we zoom in to the level of processing single words, one needs to encode phonetic cues and, at the same time, map the phonetic form onto meaning. This means that learning words is a complex task that requires not only remembering a label for a given referent, but also forming a phonetically-rich representation of that label, which in turn relies on proper segmentation of the word into individual sounds and recognizing each sound as an instance of a specific phonetic category.

The word-learning task may be particularly hard for beginning L2 learners who are not yet familiar with the L2 sound system, especially when they are processing words with novel sounds that do not exist in their native language. However, there is evidence that learners capitalize on whatever pieces of information are available to them to achieve this task: they might use lexical cues to make inferences about sound categorization (Feldman, Myers, White, Griffiths, & Morgan, 2011), and—conversely—take advantage of perceptual training on sound categorization to help them make inferences about the lexicon (Perfors & Dunbar, 2010). The question addressed in this chapter is another piece of this puzzle. Specifically, we know from the study reported in Chapter 3 that prior language knowledge can help in acquisition of a new language: L1-based perceptual biases can facilitate perception of novel sound contrasts that differ along phonetic dimensions relevant in L1. How efficiently, then, do adults capitalize on their L1-based phonetic generalizations when learning the lexicon in a new language?

Intuitively, it might seem that whatever perceptual abilities adults have, they should be able to use them when learning novel words. That is, if they hear a distinction between sounds *b* and *p*, they should be able to easily distinguish between words like *ban* and *pan*. However, the picture emerging from prior research is far less clear. In fact, research with young infants suggests that the ability to discriminate perceptually between similar sounds does not in general guarantee immediately successful learning of words that are contrasted by

those sounds. At 14 months, infants can easily discriminate the sounds *b* and *d*. However, when taught that a novel object is called a *bih*, but later on the object is referred to as a *dih*, infants do not notice this mispronunciation (Stager & Werker, 1997). The initial explanation proposed for this result was the limited resource hypothesis (Stager & Werker, 1997; Werker, Fennell, Corcoran, & Stager, 2002): since attending to fine phonetic detail while learning new words is computationally very demanding, young infants—who have limited attentional and cognitive resources—might have difficulty accessing all phonetic detail when focusing their attention on learning meaning. Subsequent research showed that 14-month-old infants succeed only with additional contextual information or under less demanding learning conditions (Fennell & Werker, 2003; Fennell, Waxman, & Weisleder, 2007; Rost & McMurray, 2009; Swingley & Aslin, 2002; Thiessen, 2007; Yoshida, Fennell, Swingley, & Werker, 2009).

There is some evidence suggesting that adults might have similar difficulties when learning words in a new language. In a study by Perfors and Dunbar (2010), native speakers of English were first exposed to a non-native distinction between a prevoiced and a voiceless unaspirated stop ([gipur] vs. [kipur]) through the distributional learning paradigm (Maye & Gerken, 2000; Maye et al., 2002), and then taught word-picture mappings using minimal-pair words distinguished by this non-native contrast. The results showed that while participants performed better than chance at learning similar words with the exact contrast they had been trained on ([gipur]-[kipur]), they were at chance at learning words contrasted by sounds with an analogous contrast ([bipur]-[pipur]). This was despite the fact that, after perceptual training on [g]-[k], participants were able to distinguish [b] and [p] perceptually. Thus, just like 14-month-old infants, adults had difficulty differentiating between similar words in a word-learning task, even though they could tell these words apart in a pure perceptual task.<sup>1</sup>

However, the difficulty in learning similar-sounding words found by

---

<sup>1</sup>Better performance on [gipur] and [kipur] might have been due to the familiarity with these specific lexical items rather than familiarity with the contrast itself from perceptual training, which is consistent with infants also performing better on familiar words (Swingley & Aslin, 2002).

Perfors and Dunbar (2010) might have been due to insufficient familiarity with the novel sounds that differentiated between the words. As Perfors and Dunbar point out, learners' representations of the novel [b] and [p] categories, derived from one session of distributional learning, might have been too fragile to see any advantage in word learning. If this reasoning is correct, then the direct comparison between 14-month-olds and adults in Perfors and Dunbar's (2010) study is perhaps less informative because the 14-month-olds have had extensive exposure to—and easily discriminate between—the sounds *b* and *d*, but still have difficulty learning the words *bih* and *dih*.

In the study reported in this chapter we designed a learning situation that is more comparable to the situation of 14-month-old infants in that we investigated how adults learn similar-sounding words that they can distinguish perceptually due to their L1-based phonetic generalizations. Specifically, we tested native speakers of Korean and of Mandarin, and we used the distinctions that were previously tested in a perceptual study (Chapter 3): *length* (e.g., [taja]–[tajja]) and *place* of articulation between alveolo-palatal and retroflex sounds (e.g., [gotɕa]–[gotʂa]).

## 4.2 Experiment

Participants learned novel word-picture mappings, where each word was in a minimal pair with either a *length* distinction or an alveolo-palatal vs. retroflex *place* distinction. An additional group of participants was recruited for a perceptual discrimination task with the same materials to make sure that previous perceptual results, reported in Chapter 3, are replicated with the newly constructed stimuli. We predicted that if adult L2 learners are able to attend to phonetic detail by using their L1-based resources when learning a new lexicon, then we should observe the same pattern in both the perceptual discrimination and the word-learning tasks: that is, Korean speakers should be more accurate on *length* trials, and Mandarin speakers more accurate on *place* trials.

## 4.2.1 Method

### Participants

90 undergraduate students at UC San Diego participated in the experiment for course credit or payment. 54 were assigned to the word-learning task, and 24 to the perceptual-discrimination task. Half were speakers of Korean, and the other half were speakers of Mandarin. Participants varied in terms of their length of residence in the US: some were born in the US, while others immigrated at some point after birth or were international students who arrived very recently. Consequently, they varied in English proficiency. Importantly, however, they all learned Korean or Mandarin from birth, reported high proficiency in those languages, and still used them regularly, predominantly with family. In most cases they had some high school and/or college exposure to Spanish or French. Some Mandarin speakers were also familiar with Taiwanese, mostly through family exposure. All participants reported no history of speech or hearing problems.

### Materials

The materials consisted of 16 bisyllabic nonce words of the form CVC(C)V, where each was in a minimal pair differing only in the middle consonant. There were 12 *length* words, with either a short or a long middle consonant, and 4 *place* words, with either an alveolo-palatal or a retroflex sibilant (both pronounced as in Polish), as illustrated in Table 4.1. We chose a subset of contrasts tested in the study reported in Chapter 3. Half of the chosen length distinctions exist in Korean ([l]–[ll], [m]–[mm], [n]–[nn]), and the other half of distinctions was novel ([j]–[jj], [w]–[ww], [f]–[ff]). Similarly, half of the chosen place distinctions exist in Mandarin ([tʃ]–[tʃ̚]), and the other half of distinctions was novel ([ʒ]–[ʒ̚]).

The materials were recorded in a soundproof booth by a phonetically-trained native speaker of Polish. There were 10 tokens recorded for each word. For *length* words, two tokens of each word with long consonants were chosen for the experiment. Subsequently, words with short consonants were created

**Table 4.1:** Stimuli.

LENGTH WORDS		PLACE WORDS	
<i>short</i>	<i>long</i>	<i>alveolo-palatal</i>	<i>retroflex</i>
taja	tajja		
tala	talla		
diwa	diwwa	gotɕa	gotʂa
difa	diffa	goʂa	goʂa
kema	kemma		
kena	kenna		

by shortening the tokens with long consonants in a way that, for each word and each recording, the naturally-recorded long consonant was reduced to half its duration<sup>2</sup> so as to maintain a constant 2:1 duration ratio.<sup>3</sup> For *place* words, two tokens each were chosen for the experiment with the goal of maximizing the similarity between the words in minimal pairs with regards to how vowels were pronounced, but at the same time choosing tokens with clearly enunciated sibilants.

The same auditory stimuli were used for both the perceptual-discrimination and the word-learning task. For the word-learning task, each word was paired with a picture of a different kind of mushroom (see two examples in Figure 4.1), which were chosen in order to include objects that were unfamiliar to our participants, but not so unfamiliar that participants would find them bizarre and hard to remember. We selected pictures that varied in shape and color so as to maximize visual differences between them. We created four different one-to-one word-to-picture mappings that were counterbalanced

<sup>2</sup>Note that this procedure was the reverse of the procedure used in Chapter 3, where words with long consonants were created by lengthening the tokens with short consonants. This change in procedure was employed because it yielded more natural-sounding words.

<sup>3</sup>Cross-linguistically, the long-to-short consonant ratio varies between 1.5 to 3 (Ladefoged & Maddieson, 1996), and the exact durations and long-to-short consonant ratios depend on several other factors such as the segmental nature of the consonant, position in a word, or stress (Payne, 2000, 2005; Payne & Eftychiou, 2006; Dmitrieva, 2007; Al-Tamimi, Abu-Abbas, & Tarawnah, 2010; Pajak, to appear). We maintained the natural duration variability between different segments (as recorded in Polish), but we controlled the ratio by keeping it constant across all segments. The chosen 2:1 duration ratio is common cross-linguistically, and it characterizes Korean tense vs. plain consonants (Han, 1992). (Recall from section 2.4.3 that Korean tense consonants are by some authors analyzed as long; Choi, 1995.)

between participants in order to make sure that the results were not driven by any peculiarities in the mappings we chose.

## Procedure

Participants sat in front of a computer, and responded by using a mouse. They were instructed that in this experiment they would be listening to a novel language, and, specifically, either (i) learn this language's sounds (in the perceptual-discrimination task), or (ii) learn the language's words for different types of mushrooms (in the word-learning task). The experiment was completed in a single session, and each participant only took part in one of the tasks. The perceptual-discrimination and the word-learning tasks were made equal in terms of the total auditory exposure to each stimulus in order to keep them as parallel as possible.

### *Perceptual discrimination (PD)*

The experiment consisted of an ABX categorization task. There were 4 blocks, each with 64 trials and lasting about 5 minutes. Blocks were separated by self-terminated breaks. In each trial, three words were presented auditorily through headphones: A <500msec> B <750msec> X (e.g., [taja] [tajja] [taja]). The task was to assess whether X sounded more like A or more like B. There were four types of trials depending on the AB contrast, as illustrated in Table 4.2: (i) *length* (24 trials per block), (ii) *place* (8 trials), (iii) filler-dissimilar (16 trials), and (iv) filler-similar (16 trials). The critical *length* and *place* trials consisted of minimal-pair AB words, and the filler trials consisted of *dissimilar* pairs, which differed in the first CV sequence, and *similar* pairs that shared the initial CV sequence. The X word was always acoustically different from both AB words to make sure that categorization was not based on pure acoustical identity of two tokens. . The AB word order was counterbalanced, and the trial order was randomized for each participant.

**Table 4.2:** Trial types in perceptual discrimination (AB=words presented auditorily) and word learning – testing (AB=labels for visually presented pictures).

CRITICAL		FILLER	
<i>type</i>	<i>example</i>	<i>type</i>	<i>example</i>
	A B		A B
<b>Length</b>	taja–tajja	<b>Dissimilar</b>	tala–goza
<b>Place</b>	gotça–gotşa	<b>Similar</b>	tala–taja



**Figure 4.1:** Example of a screen shot from the word-learning task.

### *Word learning (WL)*

In this experiment participants learned to associate words with pictures of mushrooms. There were 4 training blocks (each with 128 trials, about 10–15 minutes long) and 4 testing blocks (each with 64 trials, about 5 minutes long), interleaved. Blocks were separated by self-terminated breaks. In each trial, two pictures were presented on a computer screen (see Figure 4.1), and a word was played through headphones with a delay of 500msec. Participants were asked to click on the picture that they thought went with the word. In training, feedback was provided following the response in the form of the correct picture staying on the screen. A mouse click triggered the start of the next trial. Presenting feedback after each participant’s response meant that the early responses were necessarily random. Participants were told to guess at first, and that through

feedback they would eventually learn the correct word-to-picture mappings. In testing, no feedback was provided.

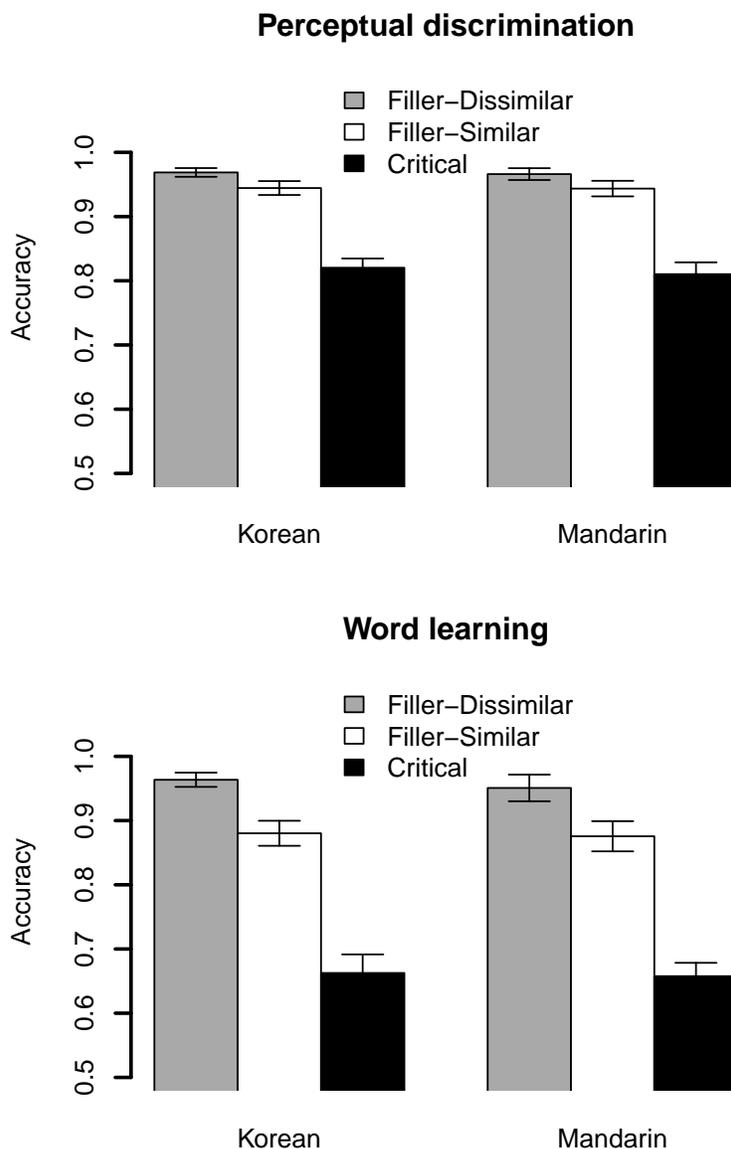
The training trial types consisted of picture pairs that were always associated with dissimilar word pairs (e.g., *taja–diwa*, *gotca–kemma*) so that participants were not directly alerted to the distinctions of interest. The testing trial types were always different from the training trials, and were completely analogous (in form and number) to trials in the perceptual-discrimination task, as illustrated in Table 4.2. Each picture pair in the word-learning task was an analog of an AB word pair in the perceptual-discrimination task, and the auditorily presented word in the word-learning task corresponded to the X word in the perceptual-discrimination task. The picture position was counterbalanced. The trial order was pseudo-randomized: we created four randomized lists, and then altered them manually so that the same word was never repeated in two consecutive trials. Furthermore, the minimal-pair trials were always separated by at least two other trials. Each participant heard each list once, with a different list for each block. The block order was counterbalanced across participants.

#### 4.2.2 Results

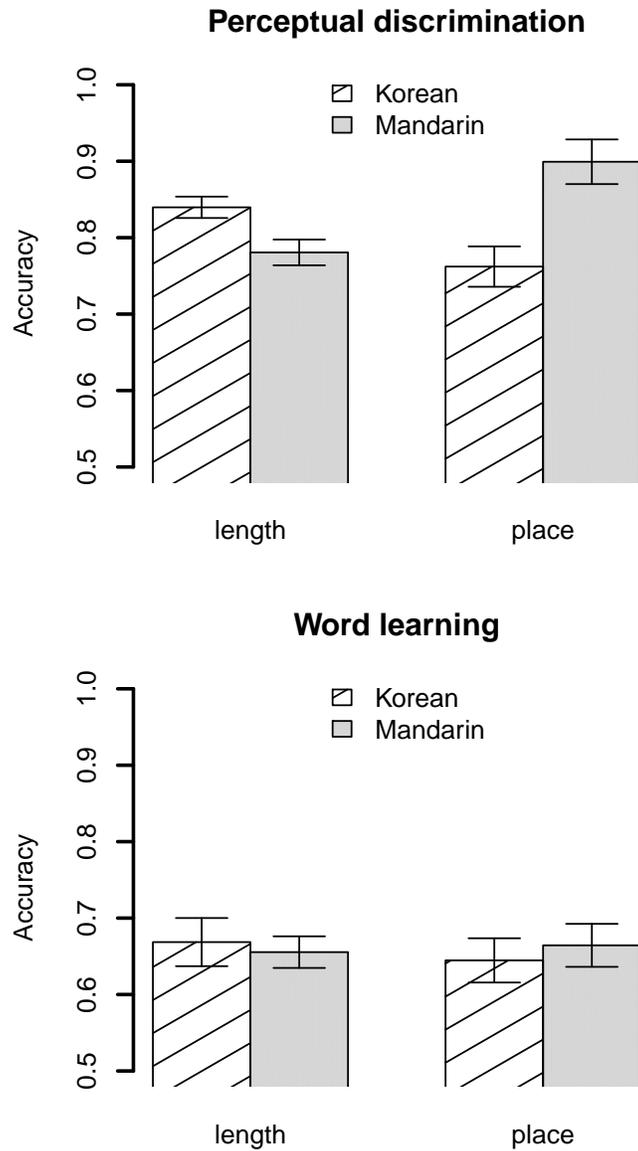
We analyzed accuracy scores from both perceptual discrimination and testing in word learning with mixed-effects logit models (Jaeger, 2008). We included random intercepts for participants and items, and random slopes for participants and items for all effects of interest that were manipulated within participants or within items. We controlled for participants' nonverbal IQ, self-reported L1 proficiency, and current L1 exposure and use by adding them as fixed effects to the models.

We expected that in both perceptual-discrimination and word-learning tasks all participants, regardless of language background, should perform best on *filler-dissimilar* trials, slightly worse on *filler-similar* trials, and worst on *critical* trials. These overall results were borne out, as illustrated in Figure 4.2 (in all figures error bars are standard errors). In models with fixed effects of TRIAL TYPE (*filler-dissimilar*, *filler-similar*, *critical*) and LANGUAGE (*Korean*, *Mandarin*), per-

formed separately for perceptual discrimination and word learning, we found that the responses in the *filler-dissimilar* condition were significantly higher than in the *filler-similar* condition (PD:  $p < .05$ , WL:  $p < .001$ ), which in turn were higher than in the *critical* condition ( $ps < .001$ ). Neither LANGUAGE nor its interactions were significant in the models, suggesting that there were no significant differences between the two language groups in overall response patterns.



**Figure 4.2:** Overall results.



**Figure 4.3:** Results from the critical *length* and *place* trials.

Next, we compared Korean and Mandarin speakers on critical trials in models with fixed effects of CRITICAL TRIAL TYPE (*length*, *place*) and LANGUAGE (*Korean*, *Mandarin*). If perceptual discrimination results reported in Chapter 3 are replicated, then we should observe a difference in performance between the two language groups in the perceptual-discrimination task: Korean speakers should be more accurate on *length* trials, and Mandarin speakers more accurate

on *place* trials. This was indeed the case, as illustrated in Figure 4.3 (top), and as evidenced by a significant interaction between CRITICAL TRIAL TYPE and LANGUAGE ( $p = .001$ ).

Now, if learners are able to capitalize on their L1-based perceptual resources when beginning to learn new words, we should observe the same response pattern in the word-learning task. However, there was no significant interaction between CRITICAL TRIAL TYPE and LANGUAGE ( $p = .21$ ), indicating that Korean and Mandarin speakers did not differ in their accuracy when learning similar-sounding words that differed in either length or place, as illustrated in Figure 4.3 (bottom). Comparing the two tasks in one model revealed a significant three-way TASK \* LANGUAGE \* CRITICAL TRIAL TYPE interaction ( $p < .001$ ), indicating that performance was indeed different across the two tasks.<sup>4</sup> Overall, these results suggest that learners were not able to take full advantage of their perceptual abilities in a word-learning task.

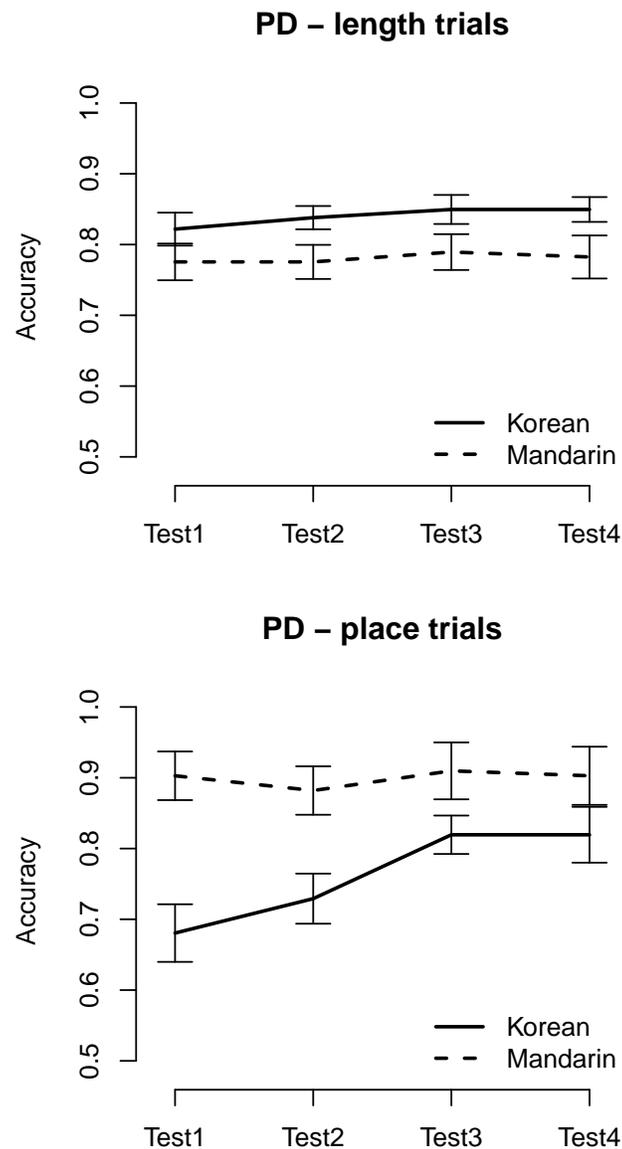
To improve our understanding of the discrepancy between the perceptual-discrimination and word-learning task results, we examined the *length* and *place* trial data from each task split by testing block (Figures 4.4–4.5). We analyzed the results by adding to the models block number as a continuous covariate. For both tasks we found significant main effects of BLOCK (PD:  $p < .05$ ; WL:  $p < .001$ ), indicating that participants improved throughout the experiment. However, there were also significant interactions between CRITICAL TRIAL TYPE and BLOCK (PD:  $p < .05$ ; WL:  $p < .01$ ): in word learning, the improvement was more prominent for the *length* trials than for the *place* trials; the opposite seemed to be the case for perceptual discrimination – more improvement on *place* than on *length* trials.

Examining the perceptual discrimination data more closely reveals that Korean speakers were catching up with Mandarin speakers on *place* trials, as indicated by a significant interaction between LANGUAGE and BLOCK ( $t = -2.22$ ) in a linear model (the interaction was not, however, significant in a mixed logit model;  $p = .49$ ), and a significant main effect of BLOCK in a model with only

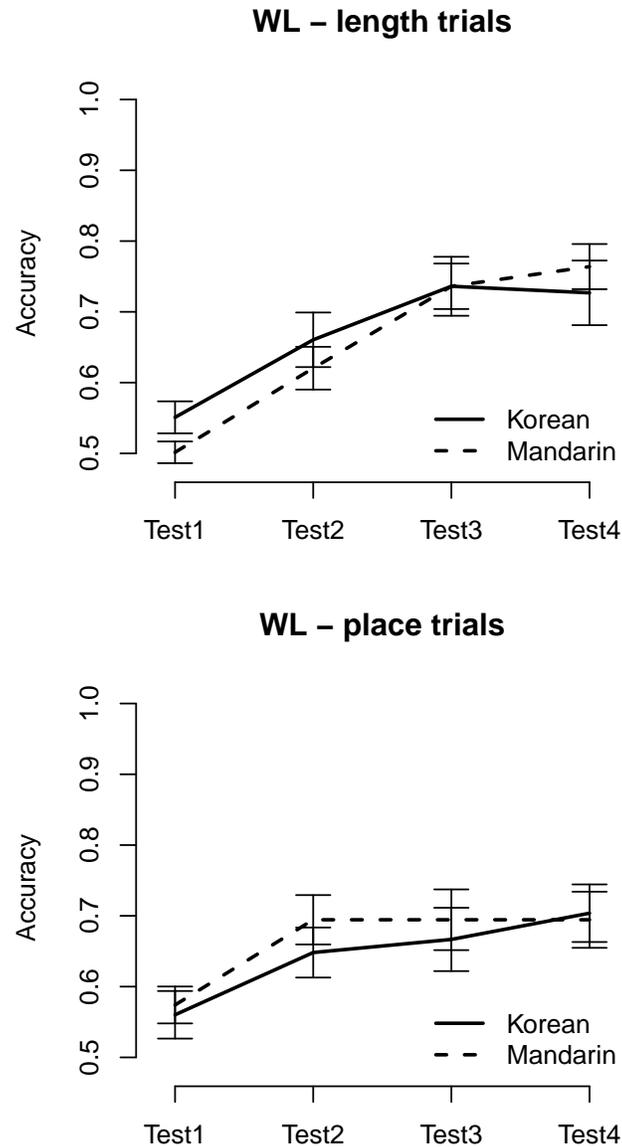
---

<sup>4</sup>In both perceptual-discrimination and word-learning tasks, accuracy did not differ on trials with familiar segments vs. trials with novel segments varying along familiar dimensions.

Korean-speaker data (both linear,  $t = 3.29$ , and logit,  $p < .001$ ). No such improvement was observed on the *length* trials for Mandarin speakers. This suggests that, for length, whatever benefit there was of perceptual learning in the perceptual-discrimination task, it happened in the first testing block, and there was no more improvement shown by the last block of testing. For place, on the other hand, Korean speakers were still benefiting from exposure until the end of the experiment.



**Figure 4.4:** Critical results from the perceptual-discrimination task by block.



**Figure 4.5:** Critical results from the word-learning task by block.

Therefore, the results reported here revealed a different pattern of responses in the perceptual-discrimination and the word-learning tasks: differential sensitivities to length and place trials by Korean and Mandarin speakers in the perceptual-discrimination task, but not in the word-learning task. However, we know that learners vary in their attention, motivation and learning skills. Furthermore, previous research suggests that performing a word-learning task may create a strain on perceptual abilities (Perfors & Dunbar, 2010). Thus, we

asked whether only faster learners are able to use their L1 resources and attend to fine phonetic detail in word learning.

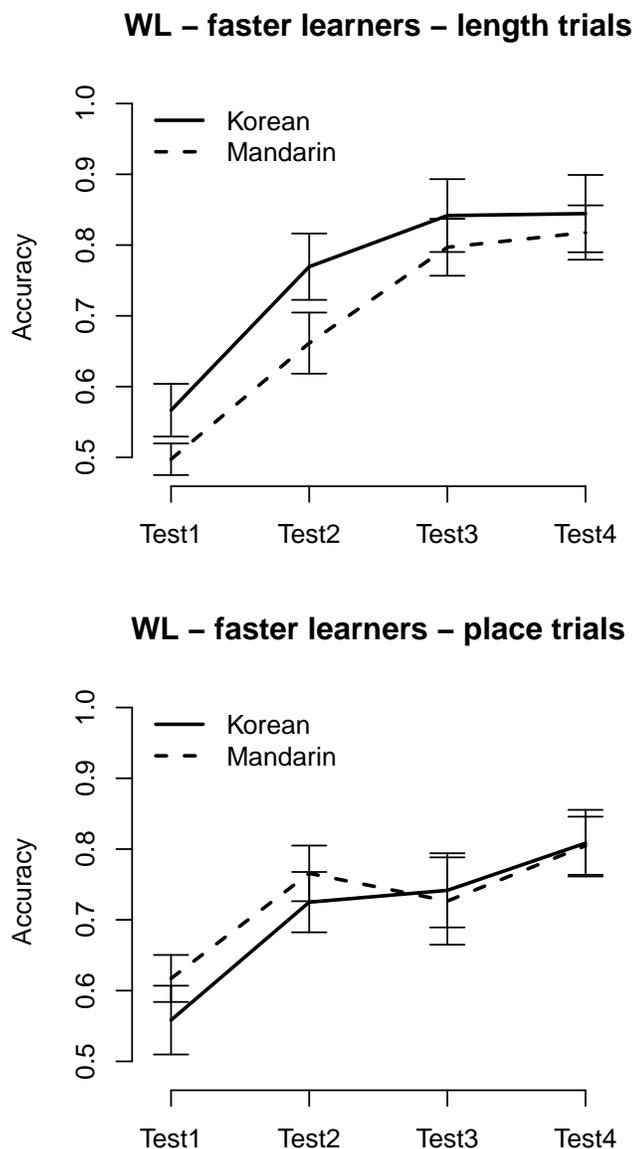
To answer this question we split the word-learning participants into faster and slower learners based on their performance on filler trials, which was a dimension independent from the variables of interest (Davidson, Shaw, & Adams, 2007, showed that this kind of split can be useful in analyzing results from a word-learning task in a novel language). The median score on all fillers combined was 94.5% accuracy. All participants scoring above median were included in the faster-learners group, while all participants scoring below median were included in the slower-learners group. There were 7 participants who scored right at 94.5%. We performed two separate analyses of the data, where the 7 participants were either all included in the faster-learners group or in the slower-learners group, later referred to as split-1 and split-2, respectively. The results were equivalent in both cases, as discussed further below. Thus, for simplicity reasons, we only illustrate the split-1 results.

The distribution of participants in terms of their language background was fairly equal in both groups. For split-1: Korean = 16 and Mandarin = 15 in the faster-learner group, and Korean = 11 and Mandarin = 12 in the slower-learner group. For split-2: Korean = 11 and Mandarin = 13 in the faster-learner group, and Korean = 16 and Mandarin = 14 in the slower-learner group. The filler scores for both faster and slower learners are provided in Table 4.3 (for split-1 only). Both groups were highly accurate on filler pairs (at least 80% accuracy), but there was much more variability in the slower-learner group.

**Table 4.3:** Results from the word-learning filler trials.

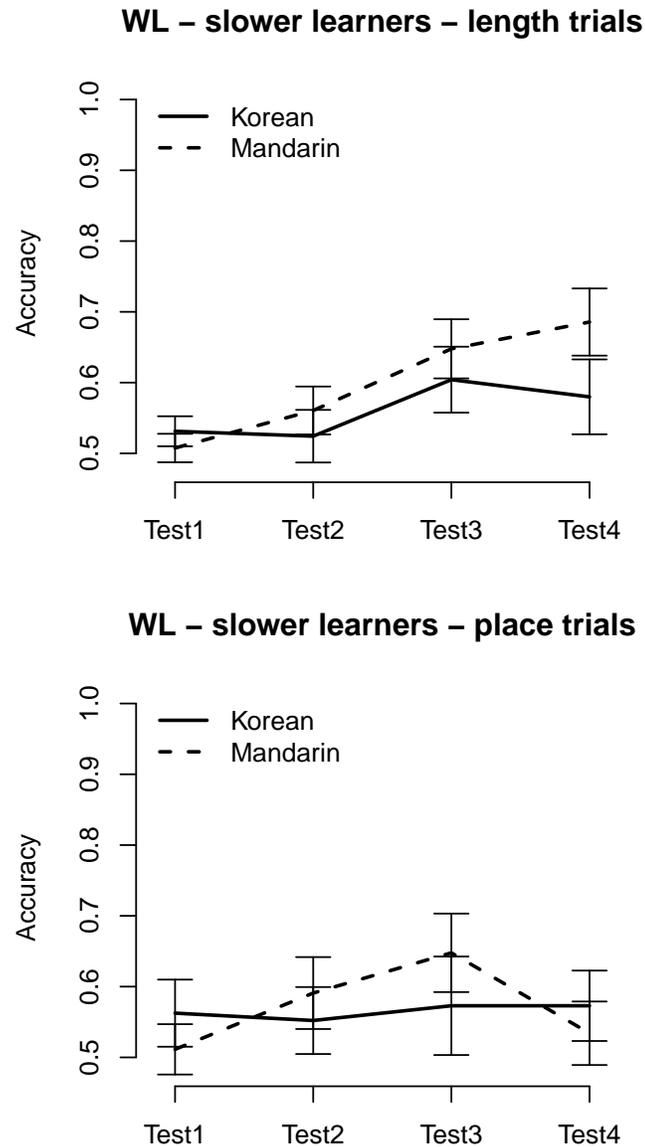
BLOCK	TOP HALF		BOTTOM HALF	
	<i>Korean</i>	<i>Mandarin</i>	<i>Korean</i>	<i>Mandarin</i>
Test 1	.95 (.01) <sup>a</sup>	.91 (.01)	.78 (.04)	.84 (.02)
Test 2	.98 (.01)	.98 (.00)	.88 (.02)	.85 (.06)
Test 3	.98 (.00)	.97 (.01)	.89 (.01)	.84 (.06)
Test 4	.99 (.01)	.98 (.01)	.88 (.02)	.84 (.05)

<sup>a</sup> Numbers in parentheses indicate standard errors.



**Figure 4.6:** Results from the word-learning task: faster learners.

The results split by faster and slower learners are illustrated in Figures 4.6 and 4.7. Even by visual inspection alone, the results look strikingly different in the faster vs. slower-learner group: in the faster-learner group, participants were clearly learning the minimal-pair words, as indicated by their much higher levels of accuracy. In the slower-learner group, on the other hand, participants' responses were close to chance, with only minimal signs of improvement through-



**Figure 4.7:** Results from the word-learning task: slower learners.

out the experiment. We analyzed these results with a model with fixed effects of CRITICAL TRIAL TYPE (*length, place*) and LANGUAGE (*Korean, Mandarin*), and an additional fixed effect of FILLER PERFORMANCE (*top, bottom*), separately for split-1 and split-2. In both cases we found significant three-way interactions between CRITICAL TRIAL TYPE, LANGUAGE, and FILLER PERFORMANCE ( $ps < .05$ ), indicating distinct response patterns for Korean vs. Mandarin speakers on *length*

and *place* trials depending on their overall success rate in learning, as measured by their accuracy on filler trials.

Separate LANGUAGE \* CRITICAL TRIAL TYPE analyses indicated that the three-way interaction was driven by the faster-learner group. Specifically, for the faster-learner group, we found the pattern more in line with what we predicted if participants were taking advantage of their perceptual biases: Korean speakers were more accurate on *length* trials than Mandarin speakers, but not on *place* trials, as indicated by significant interactions between CRITICAL TRIAL TYPE and LANGUAGE ( $ps < .05$ ). These interactions were already marginal (split-1,  $p = .08$ ) or close to marginal (split-2,  $p = .11$ ) in the first block of testing, suggesting that the effect was there even before the participants were alerted to the presence of the length and place minimal pairs, since the pictures corresponding to these minimal pairs were never shown together in training. Furthermore, models examining *length* trials only revealed marginal main effects of LANGUAGE ( $ps < .08$ ). On place trials, Korean and Mandarin speakers were not significantly different, but the numerical tendency was the opposite of that seen in the length trials: Mandarin speakers were slightly more accurate than Korean speakers. The lack of difference between the two language groups on place trials may be a consequence of Korean speakers improving on perception of the place distinction throughout the experiment, as observed in the perceptual-discrimination task (Figure 4.4, bottom).

For slower-learner group, on the other hand, we found no significant interaction between CRITICAL TRIAL TYPE and LANGUAGE. There were also no significant differences between Korean and Mandarin speakers when only *length* ( $ps > .34$ ) or only *place* ( $ps > .73$ ) trials were examined. Mandarin speakers did seem to improve on *length*—but not *place*—trials toward the end of the experiment. This apparent trend is, however, most likely to be noise, since the four-way LANGUAGE \* CRITICAL TRIAL TYPE \* FILLER PERFORMANCE \* BLOCK interactions did not approach significance for either split-1 ( $p = .52$ ) or split-2 ( $p = .27$ ).

### 4.3 Discussion

The study reported in Chapter 3 has shown beneficial effects of L1 properties on L2 discrimination, but do discrimination benefits extend to vocabulary learning? We asked participants of different language backgrounds to either discriminate words containing particular sound contrasts, or to map those words to novel referents. Some sound contrasts were similar to contrasts in each listener's L1, and others were not. For the word-learning task participants as a whole there was no clear effect indicating that participants made effective use of their L1 resources, but there is evidence that faster learners (as independently assessed by filler performance) were able to do so. This is in contrast to the perceptual discrimination results, where L1-based perceptual advantages were observed for all participants. This result thus reveals an intermediate effect between failure to learn similar-sounding words (as observed for 14-month-old infants) and a full ability to use existing perceptual abilities in learning (which should mimic the perceptual discrimination data).

Overall, the results reported in this chapter suggests that there is something inherently hard about the early stage of word learning that precludes attention to fine phonetic detail that is otherwise available during phonetic processing. But what is the source this difficulty? One answer is that learning novel words is simply a highly complex task, and only individuals with better attentional or general cognitive abilities can effectively manage simultaneous information at multiple levels of processing. Crucially, what might make this task harder is that learners must rely heavily on their working memory to remember label-referent mappings. This means that even in cases when perception of subtle sounds is good, learners might still have trouble remembering which label corresponds to which picture when the labels are highly similar to each other. In other words, word learning is most likely not just about discrimination because the noise level associated with the label-referent mapping might be much higher than the noise level associated with perception. This idea is corroborated by the filler results in our word learning study, where performance on similar fillers was much worse relative to performance on dissimilar fillers, despite the

fact that—in the perceptual discrimination task—perception of both types of contrasts was near ceiling.

## 4.4 Acknowledgements

Chapter 4 is a revised version of Pajak, Creel, & Levy (2012) [“Can native-language perceptual bias facilitate learning words in a new language?” In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 2174–2179). Austin, TX: Cognitive Science Society], and it includes material currently being prepared for submission for publication [Pajak, Creel, & Levy. “Can native-language perceptual bias facilitate learning similar words in a new language?”]. The dissertation author was the primary investigator and author of these papers. In addition to planned presentation to the Cognitive Science Society in August 2012, this work was also presented at the 86th Annual Meeting of the Linguistic Society of America.

# Chapter 5

## Overcoming L1 Perceptual Biases in Phonetic Category Learning

### 5.1 Introduction

This chapter tests the second prediction of the inductive theory outlined in section 2.4.1 stating that if a learner successfully learns a novel L2 contrast for one set of segments, then this should facilitate discrimination and learning of additional L2 categories that vary along the same phonetic dimension.

This prediction was tested using the distributional learning paradigm, in which listeners (here, monolingual English speakers) are exposed to a new language through listening to stimuli sampled from a continuum of sounds that vary along some phonetic dimension (here, segmental length). The stimuli are sampled from either a bimodal frequency distribution, suggesting that there are two categories along the continuum (here, short and long segments), or a unimodal distribution, suggesting only one category (and, thus, no contrast between short and long segments). Crucially, all participants are exposed to the same inventory of stimuli, differing only in relative frequency of occurrence among stimuli within the inventory. Thus, any differences between bimodal and unimodal conditions in subsequent testing must be due to participants' interpretation of the novel sounds as influenced by training and not just to auditory sensitization.

Beyond its relevance for testing the predictions made by the inductive theory, this work contributes to the perceptual learning literature by investigating distributional learning on a previously unstudied phonetic dimension—*segmental length*—and generalization across segment classes (sonorants and voiceless fricatives). Unlike previously studied voicing dimension, length cross-cuts a wide range of possible segments, and is not in any form contrastive in the participants' native language (English).

## 5.2 Experiment 1

We exposed monolingual English speakers to evidence suggesting a novel phonetic contrast along the length dimension. We used the distributional learning paradigm, as applied by Maye and Gerken (2001) in a study with adult participants. Subsequently, we tested their categorization of short and long segments for trained and untrained segment classes (sonorants and voiceless fricatives). We predicted that participants would generalize the relevance of length in sound categorization from a trained class to an untrained class.

### 5.2.1 Method

#### Participants

48 undergraduate students at UC San Diego participated in the experiment for course credit. They were all monolingual speakers of English, in most cases with some limited high school and/or college exposure to Spanish or French. Crucially, none of them had any exposure to any language that uses length contrastively. All participants reported no history of speech or hearing problems.

#### Materials

The materials consisted of nonce words recorded in a soundproof booth by a phonetically-trained native speaker of Polish. The critical length items in-

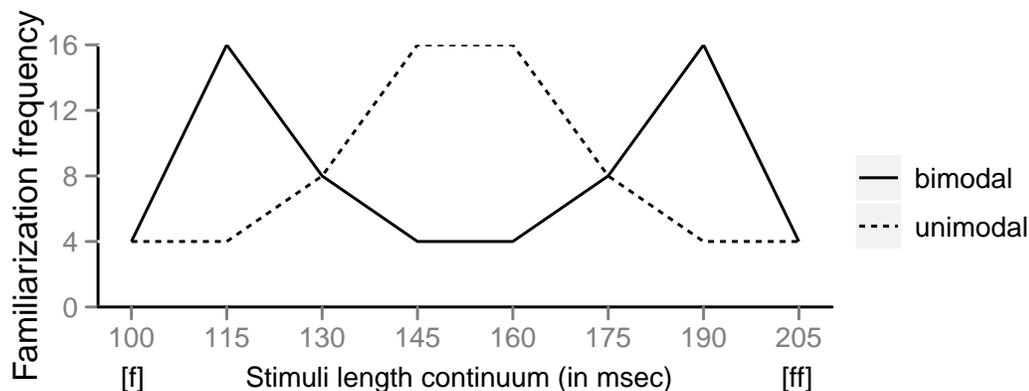
cluded segments from two classes: sonorants ([j], [l], [m], [n]), and voiceless fricatives ([s], [f], [θ], [ʃ]). They were recorded as words with long consonants: [ajja], [illa], [amma], [inna], [assa], [iffa], [aθθa], [iʃʃa]. Subsequently, the consonant length in each word was manipulated to create length continua, each with eight tokens. There are several ways in which such continua could be created. One way would be to maintain natural between-segment duration differences (e.g., sonorant consonants are generally shorter than fricatives<sup>1</sup>), but manipulate relative durations so that for each continuum the endpoints are always in the same duration ratio. Another way, which we adopted, is to use the same distribution on absolute durations for all segments (see the discussion section for more on the consequences of this choice). In the continua we created, durations of all consonants ranged from 100msec (short) to 205msec (long), and each adjacent token differed by 15msec. The fillers resembled the critical items, but different consonants were used: [ira], [iʔa], [aɟa], [atʃa], [idza], [itsa], [aba], [apa], [ida], [ita], [aga], [aka], [ixa], [iχa], [aβa], [aʕa].

## Procedure

The experiment adhered as closely as possible to the procedure used by Maye and Gerken (2001), and consisted of two main parts: training and testing.

In training, participants listened to single words presented over headphones that were of one of two STIMULUS TYPES: *critical* or *filler*. Each participant was trained on critical items from one TRAINED SEGMENT CLASS (either *sonorants* or *fricatives*), and in one of two CONDITIONS: (1) *bimodal*, imitating a language with phonemic contrasts between short and long consonants, and (2) *unimodal*, imitating a language with no phonemic length contrasts (see Figure 5.1). All participants were trained on the same filler items: the words [ira], [iʔa], [aɟa], [atʃa]. To maintain participants' attention on the experimental items, they were instructed to push a button after they heard each word. The response

<sup>1</sup>The ranges of duration for English consonants that are equivalent to those used in the experiment are roughly the following (in msec): [j] 39–100, [l] 42–85, [m] 50–89, [n] 38–83, [s] 61–126, [f] 88–138, [θ] 46–90, [ʃ] 88–138 (based on the phonetically annotated portion of the Switchboard corpus, as described in 'The Switchboard Transcription Project' report by Steven Greenberg, 1996.)



**Figure 5.1:** Critical training stimuli in Experiment 1.

to a given stimulus triggered the presentation of the following stimulus with a delay of 1 second. Training consisted of a total of 384 words and lasted for about 10 minutes. This included four repetitions of a training block, where each block had 64 critical items (16 tokens from each of the four length continua) and 32 filler items (8 different recordings of each of the four filler items). Stimulus order was randomized for each participant, and there was a self-terminated break after each block.

The testing was identical for all participants, and consisted of an AX discrimination task. Participants listened to pairs of words, and were asked to judge whether these were two different words or two repetitions of the same word. For critical pairs, these were endpoints of each continuum, either ‘different’ (100msec – 205msec, 205msec – 100msec) or ‘same’ (100msec – 100msec, 205msec – 205msec). For filler ‘different’ pairs, these were two words that differed by one segment: the contrasts were either in voicing ([ɖ]–[tʃ], [ɗ]–[ts], [b]–[p], [d]–[t], [g]–[k]), in place of articulation ([x]–[χ], [ʁ]–[ʀ]), or in both ([r]–[ʀ]). The ‘same’ pairs were always physically identical. The TESTED SEGMENTS were of one of two types: *trained* (i.e., heard in training) or *untrained* (i.e., heard for the first time in testing). There was a total of 384 word pairs, which included 6 repetitions of a testing block. One block consisted of 32 critical pairs (16 ‘same’ and 16 ‘different’) and 32 filler pairs (16 ‘same’ and 16 ‘different’). The words in each pair were separated by an interstimulus interval of 750msec. As with

training, stimulus order was randomized for each participant, and there was a self-terminated break after each block. Participants responded by pushing a button on a gamepad. They were instructed to respond according to their intuition based on what they learned during the training period, and were assured that there were no strictly right or wrong answers. The instructions included a short practice with English words, where ‘different’ words were minimal pairs (e.g., *mass* – *miss*), and ‘same’ words were repetitions of the same word pronounced with different intonations. Testing lasted about 20 minutes.

### 5.2.2 Results

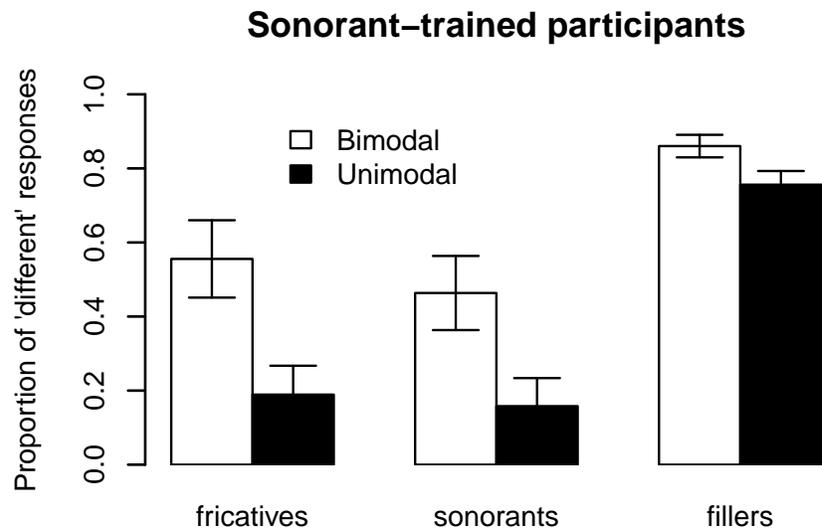
We predicted that successful distributional training should lead to a difference between the bimodal and the unimodal conditions on critical length trials: bimodal training should result in more ‘different’ responses (since the training should suggest that short and long consonants are contrastive in this language), while unimodal training should lead to fewer ‘different’ responses (because the training provided no evidence that short and long consonants belong to different categories). Furthermore, we predicted that participants would generalize the relevance of length from trained to untrained words (reflected in no difference in performance on trained and untrained items), and that this generalization would be bidirectional (i.e., from sonorants to fricatives, and vice versa).

Since performance was at ceiling on ‘same’ trials (> 95% correct for each CONDITION, TRAINED SEGMENT CLASS, and TESTED SEGMENT type), we only analyzed the responses from ‘different’ trials,<sup>2</sup> using mixed-effects logit models with random slopes and intercepts for participant and item.<sup>3</sup>

---

<sup>2</sup>Note that in the distributional learning experiments we did not calculate d-prime scores, but instead we analyzed the raw accuracy on ‘different’ trials. D-prime is used to measure sensitivity to a distinction between two stimuli, but in the distributional learning paradigm we are not asking participants to detect a difference between the stimuli, but rather to make a judgment about phonetic categorization. This means that we expect ‘same’ responses even in cases when a difference between two words has been detected. Thus, d-prime is not an adequate measure in this case.

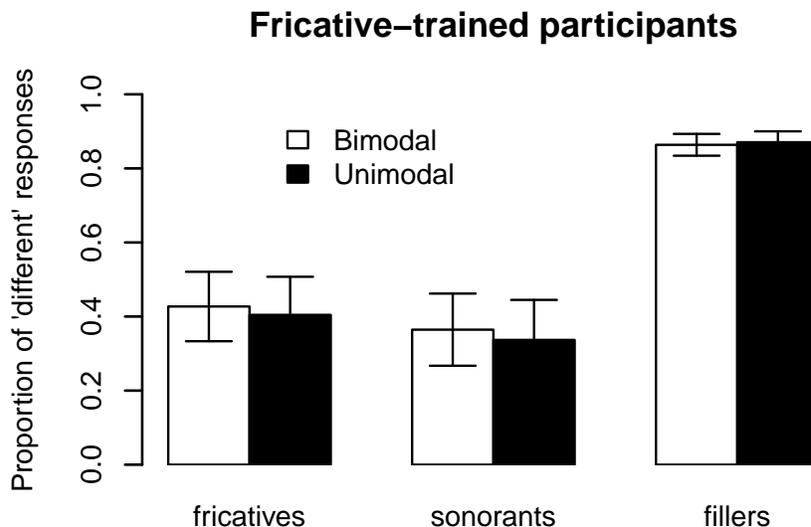
<sup>3</sup>We also performed ANOVA analyses and found no major differences in results. Minor discrepancies are reported in footnotes.



**Figure 5.2:** Experiment 1: proportion of ‘different’ responses on ‘different’ trials by participants trained on the sonorant segment class.

First, we examined the critical trials for the fixed effects of *CONDITION* (*bimodal*, *unimodal*), *TESTED SEGMENT* (*trained*, *untrained*), and *TRAINED SEGMENT CLASS* (*sonorant*, *fricatives*). There was a main effect of *CONDITION* ( $p < .05$ ): as predicted, participants in the bimodal condition responded ‘different’ more often than in the unimodal condition. However, there was also a significant interaction between *CONDITION* and *TRAINED SEGMENT CLASS* ( $p < .05$ ): the difference between the bimodal and the unimodal conditions was driven by the participants trained on the sonorant class.<sup>4</sup> That is, as can be seen in the left part of Figure 5.2, participants trained on sonorants responded ‘different’ more often in the bimodal than in the unimodal condition. However, as illustrated in the left part of Figure 5.3, all participants trained on fricatives performed similarly regardless of the condition, even on the trained items (in all figures error bars are standard errors). These results suggest that the distributional training was successful when it was done on sonorant length continua, but not when the training continua involved fricatives, in which case there was no difference between the bimodal and the unimodal conditions on any tested words: whether critical or filler, or trained and untrained.

<sup>4</sup>Both of these effects were only marginal in ANOVAs with  $p = .06$  and  $p = .08$ , respectively.



**Figure 5.3:** Experiment 1: proportion of ‘different’ responses on ‘different’ trials by participants trained on the fricative segment class.

Since the training was only successful for the sonorant-trained participants, we examined the critical trials for the effect of generalization for this group alone. We used a mixed model with fixed effects of CONDITION (*bimodal*, *unimodal*) and TESTED SEGMENT (*trained*, *untrained*). As expected by previous main effect, there was a significant main effect of CONDITION ( $p < .01$ ) with participants in the bimodal condition responding ‘different’ more often than in the unimodal condition. Furthermore, as predicted by the inductive theory, there was no significant main effect of TESTED WORD, meaning that participants in both bimodal and unimodal conditions performed similarly on trained and untrained items. Separate pairwise comparisons revealed that the difference between bimodal and unimodal conditions was significant for both trained and untrained critical items ( $ps < .01$ ). These results suggest that participants generalized length to the novel segment class.

This effect was not due to a simple bias of bimodally-trained participants to respond ‘different’ on any trial, as reflected by a significant interaction between CONDITION and STIMULUS TYPE (*critical*, *filler*) ( $p < .05$ ), as well as the same interaction for only untrained items ( $p < .05$ )<sup>5</sup>: the difference between the

<sup>5</sup>For these cases the models with the full random effects structure failed to converge.

bimodal and the unimodal conditions was significantly larger for the critical than for the filler trials, even when just the untrained items were considered.

The fact that testing was identical for all participants, but the distributional training was only successful for the sonorant-trained group and not for the fricative-trained one, allows us to make a direct comparison between the two groups. By treating the performance of the fricative-trained group as a baseline (38% ‘different’ responses), we can see the net effect of bimodal vs. unimodal training by comparing the performance of sonorant-trained participants to the baseline. This comparison reveals that successful bimodal training increased ‘different’ responses by 13%, whereas successful unimodal training decreased ‘different’ responses by 21%.

### 5.3 Interim Discussion

This study yielded two key results. First, monolingual speakers of English can be trained through distributional learning to recognize a phonetic category distinction along a dimension (segmental length) which is never contrastive in their native language. After only one ten-minute training session of 256 critical items, participants exposed to sonorants sampled from a bimodally distributed length continuum categorized words differing only in sonorant length as being distinctive more often than did participants exposed to sonorants of unimodally distributed length. Second—and even more crucially to predictions of the inductive theory—speakers generalized the relevance of length for sound categorization to a different set of consonants, voiceless fricatives. This generalization was quite aggressive, with the effect on fricative categorization during testing just as strong as the effect on sonorant categorization. This result seems not to be reducible to greater general sensitization to any phonetic distinctions for the bimodally trained group, since the effect on performance for fillers—even those to which participants received no exposure during training—was smaller (though this comparison must be taken with caution

---

Thus, we iteratively removed random effects with the smallest variance until convergence was successful.

since performance for fillers was higher across the board than for critical trials).

This result contrasts with Maye and Gerken's (2001) study of distributional learning of a novel voicing distinction, where no evidence of generalization was found.<sup>6</sup> Since Maye and Gerken only used one segment continuum for training, our results suggest that training on a wider range of segments might yield stronger cross-segment generalization. This is in line with previous research showing that exposure to greater acoustic variability, such as different phonetic contexts or multiple talkers, improves acquisition of both native and non-native contrasts (Lively, Logan, & Pisoni, 1993; Bradlow, Pisoni, Akahane-Yamada, & Tokhura, 1997; Wang, Spence, Jongman, & Sereno, 1999; Rost & McMurray, 2009). Interestingly, the benefit of high-variability training is mostly observed in cases when listeners have to generalize to new talkers or new words produced by old talkers used in training (Lively et al., 1993). Thus, cross-segment generalization of phonetic dimensions is likely to also be more robust when training is enriched with acoustic variability that learners encounter in natural languages.

For participants trained on fricatives, in contrast, the choice of bimodal versus unimodal distribution of segment length had no discernible effect on word categorization. The most likely reason for this might be related to the differences in duration between these two classes of consonants in naturally spoken English: voiceless fricatives are generally longer than sonorants. Since we created uniform length continua for both segment classes, this meant that all the tokens from the sonorant continua were longer than their usual duration range in English, while for fricatives these ranges partially overlapped. This might have been the reason why the fricative-trained participants did not pick up on the distributional information: they may have heard the fricatives of around 200msec as unusually long, but still interpreted them as within reasonable English-like duration range, which consequently was not sufficient for bimodally-trained participants to infer contrastiveness of the length dimension. If this is correct, then modifying the fricative continua (by including longer du-

---

<sup>6</sup>Perfors and Dunbar (2010) did obtain both learning and generalization for a voicing distinction similar to Maye and Gerken's, but they used much more training and no fillers.

rations) should be more effective in guiding participants' inferences. This is confirmed by a follow-up experiment reported in the next section: when the fricative continua range from 140msec to 280msec, the results for fricative-trained participants look similar to those for sonorant-trained participants in Experiment 1.

In the face of the learning failure observed in the present experiment for fricative-trained participants, the generalization by sonorant-trained participants to fricatives is all the more impressive: distributional evidence as to whether length is contrastive for sonorants informs participants' perception of fricative length contrastiveness even within a range of the continuum which would not itself drive learning through exposure to fricatives.

## 5.4 Experiment 2

This experiment is a follow-up to Experiment 1. Given the failure of distributional learning for fricative-trained participants in Experiment 1, this experiment used modified fricative stimuli—sampled from a longer length continuum—with the goal to replicate the result found for sonorant-trained participants.

As in Experiment 1, monolingual English speakers were exposed to evidence suggesting a novel contrast, segmental length, using the distributional learning paradigm (Maye & Gerken, 2000, 2001; Maye et al., 2002). In this study, however, participants were only trained on voiceless fricatives. Subsequently, participants' categorization of short and long segments was tested for trained (voiceless fricatives) and untrained segment classes (sonorants). The prediction was that longer fricative durations should yield more effective learning and generalization compared to Experiment 1.

## 5.4.1 Method

### Participants

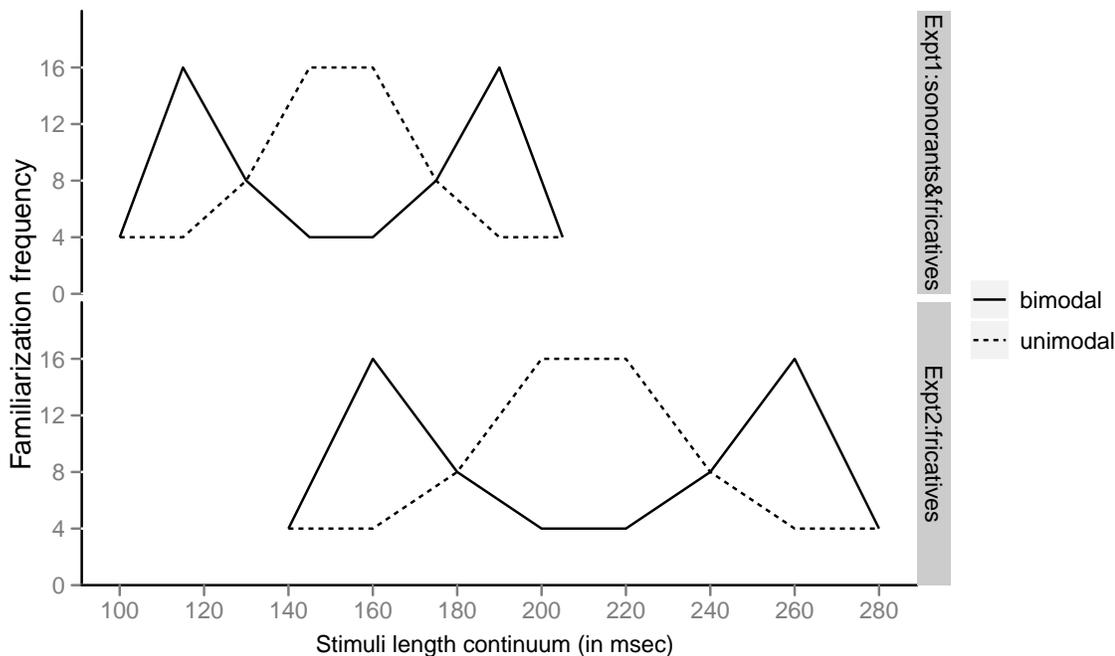
54 undergraduate students at UC San Diego participated in the experiment for course credit. They were all monolingual speakers of English, in most cases with some limited high school and/or college exposure to Spanish or French. Crucially, none of them had any exposure to any language that uses length contrastively. All participants reported no history of speech or hearing problems.

### Materials

We used the same materials as in Experiment 1 with the exception of the critical fricative segments. In the new fricative continua we created, durations of the consonants ranged from 140msec (short) to 280msec (long), and each adjacent token differed by 20msec. (Recall that in Experiment 1 the continua ranged from 100msec to 205msec, and that each adjacent token differed by 15msec.)

### Procedure

The procedure was exactly the same as in Experiment 1, except that we only trained on one segment class: voiceless fricatives. As before, there were two STIMULUS TYPES (*critical* or *filler*), and two CONDITIONS: (1) *bimodal*, imitating a language with phonemic contrasts between short and long consonants, and (2) *unimodal*, imitating a language with no phonemic length contrasts (see Figure 5.4). The testing consisted of an AX discrimination task, where participants listened to pairs of words, and were asked to judge whether these were two different words or two repetitions of the same word. For critical pairs, these were endpoints of each continuum, either 'different' (for sonorants: 100msec – 205msec, 205msec – 100msec; for fricatives: 140msec – 280msec, 280msec – 140msec) or 'same' (for sonorants: 100msec – 100msec, 205msec – 205msec; for fricatives: 140msec – 140msec, 280msec – 280msec). There were two types of TESTED SEGMENTS: *trained* (fricatives) and *untrained* (sonorants).



**Figure 5.4:** Critical training stimuli in Experiments 1 and 2.

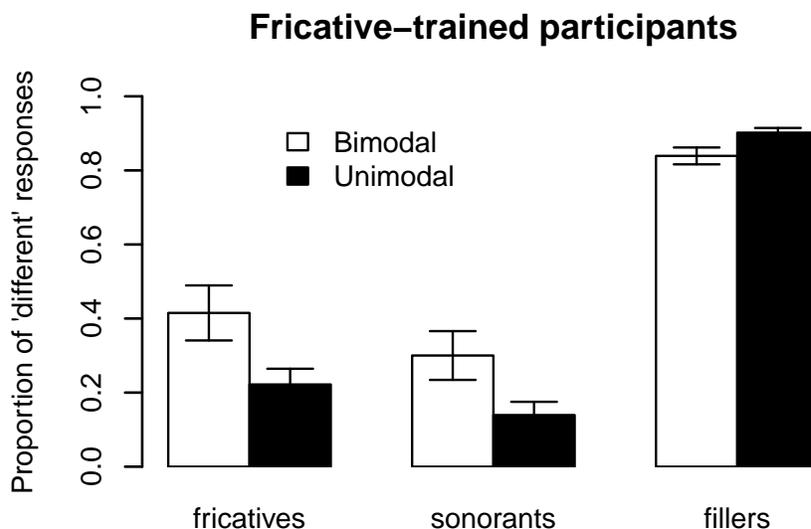
## 5.4.2 Results

We predicted that successful distributional training should lead to a difference between the bimodal and the unimodal conditions on trained fricative-length trials. Furthermore, we predicted that participants would generalize the relevance of length from trained fricatives to untrained sonorants.

Performance was at ceiling on ‘same’ trials (> 96% correct for each STIMULUS TYPE, CONDITION, and TESTED SEGMENT type), and so we only analyzed the responses from ‘different’ trials, using mixed-effects logit models with random slopes and intercepts for participant and item.<sup>7</sup>

We examined the critical trials for the fixed effects of CONDITION (*bimodal*, *unimodal*) and TESTED SEGMENT (*trained*, *untrained*). There was a main effect of CONDITION ( $p < .05$ ) and no interaction between CONDITION and TESTED SEGMENT ( $p = .10$ ): as predicted, participants in the bimodal condition responded ‘different’ more often than in the unimodal condition, and this

<sup>7</sup>We also performed ANOVA analyses and found no major differences in results.



**Figure 5.5:** Experiment 2: proportion of ‘different’ responses on ‘different’ trials.

was the case for both trained and untrained segments (see Figure 5.5). There was also a significant main effect of TESTED SEGMENT ( $p < .001$ ): overall, participants responded ‘different’ more often for fricatives than for sonorants. This result indicates that distributional training was successful, and—as was the case for sonorant-trained participants in Experiment 1—there was generalization to untrained segments.

## 5.5 Discussion

The results of Experiment 2 confirmed our hypothesis that the failure of learning in Experiment 1 was at least in part due to the way the length continua were constructed. With longer fricative continua the results for fricative-trained participants looked similar to those for sonorant-trained participants in Experiment 1.

Interestingly, when comparing the rate of ‘different’ responses to fricative stimuli across the two experiments, one notices that the responses after the bimodal training were basically at the same level in both experiments, whereas the responses after the unimodal training went down in Experiment 2 relative to Experiment 1. If Experiment 1 is treated as baseline in this case, then distribu-

tional training in Experiment 2 seems to only have affected unimodally-trained participants. This suggests that the evidence for two categories along the length continuum was still relatively weak in the bimodal training. This raises the question of whether learners *can* actually acquire phonetic categories from this type of distributional evidence. In Experiment 2 we tried to create stimuli that would be more conducive to learning, but in many cases of L2 acquisition there *will* be some overlap between native and non-native phonetic categories, similar to the overlap between English fricatives and fricative stimuli in Experiment 1. How is then distributional learning a viable explanation of actual learning? The answer might lie in the overall variability of stimuli in the exposure phase. As already discussed in section 5.3, greater acoustic variability aids learning of novel sound contrasts. Therefore, variability in talkers, phonetic contexts, segments, etc.—found in natural language data—might be indispensable for more robust learning from distributional evidence.

Overall, the results of the two experiments reported in this chapter revealed that—given successful learning of short vs. long categories for the trained segments—participants are also inclined to categorize novel segments as either short or long. This suggests that there is generalization of phonetic dimension informativity across segment classes within a new language that is being learned.

These results are problematic for the mapping approaches to L2 speech perception and learning because these approaches have no straightforward explanation of distributional learning, much less of generalization. If we assume that phonological categorization of novel sounds proceeds through mapping of these sounds onto the most similar L1 categories, then frequency of exposure to sounds from a given phonetic continuum (as in distributional learning) should not have any effect on how the endpoints of that continuum are mapped. The results reported here show, however, a clear difference in responses between bimodally- and unimodally-trained participants. Furthermore, under mapping approaches there is no reason why exposure to novel stimuli from one segment class should affect perception and categorization of stimuli from another class. Yet the results discussed here show this exact kind of dependency.

The findings reported in this chapter suggest that learners make immediate inferences about the informativity of phonetic dimensions based on a very limited exposure to a new language, and are able to use these inferences to make predictions about the informativity of these dimensions in a language as a whole. These results are consistent with the second prediction outlined in section 2.4.1, and support the proposed approach to understanding L2 learning as a process of inductive inference.

## 5.6 Acknowledgements

Chapter 5 is a revised version of Pajak & Levy (2011) [“Phonological generalization from distributional evidence”. In L. Carlson, C. Hölscher & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 2673–2678). Austin, TX: Cognitive Science Society] and Pajak & Levy (to appear) [“How abstract are phonological representations? Evidence from distributional perceptual learning”. In *Proceedings of the 47th Annual Meeting of the Chicago Linguistic Society*. Chicago, IL: University of Chicago]. The dissertation author was the primary investigator and author of these papers. This work was presented to the Cognitive Science Society and the Chicago Linguistic Society.

## Chapter 6

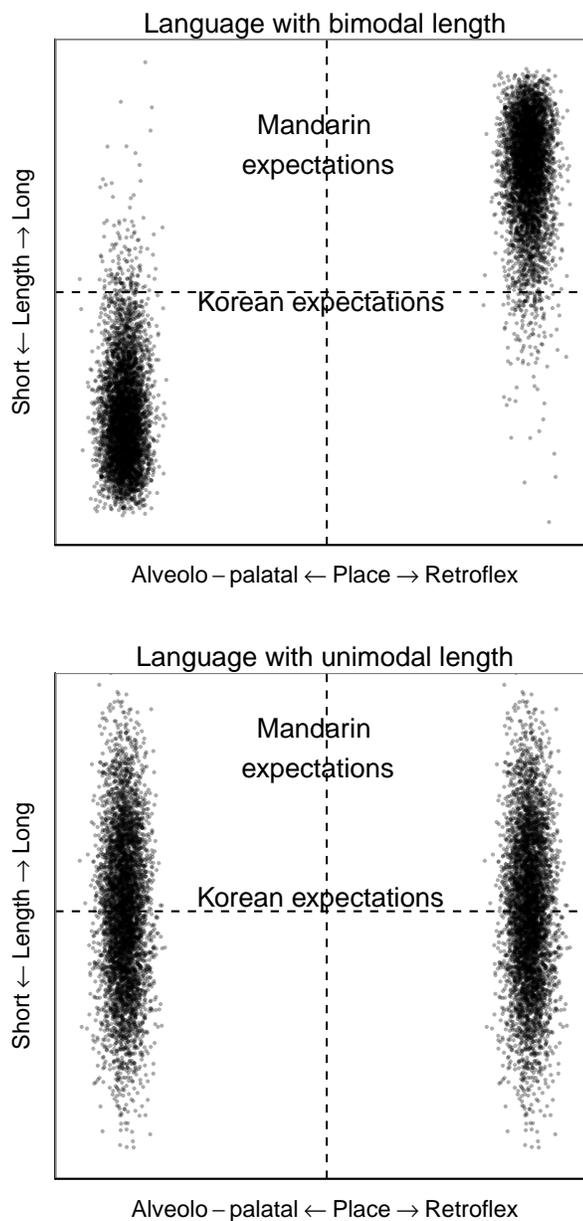
# Combining L1 Perceptual Biases and Distributional Cues in Phonetic Category Learning

### 6.1 Introduction

This chapter tests the third prediction of the inductive theory outlined in section 2.4.1 stating that learners' interpretation of novel language statistics may be affected by their L1 perceptual biases. This prediction was tested in a study with Korean and Mandarin speakers. Given known perceptual sensitivities of these two language groups to length and place contrasts (i.e., Korean speakers are more sensitive to length, and Mandarin speakers are more sensitive to place, as discussed in Chapter 3), it is possible that native speakers of these languages would show differential biases when interpreting ambiguous distributional cues to L2 phonetic categories.

### 6.2 Experiment

Speakers of Korean and Mandarin were recruited for this experiment, and the stimuli were constructed using the length and the alveolo-palatal vs. retroflex place distinctions. We designed two languages with sound dis-



**Figure 6.1:** Schematic representation of segment statistics in two languages constructed for the purpose of this study. **Top:** strongly bimodal *place* and weakly bimodal *length*. **Bottom:** strongly bimodal *place* and unimodal *length*

tributions illustrated in Figure 6.1. In both languages sounds are distributed along both dimensions of interest: length and place. Length is distributed continuously from short to long. For place, on the other hand, there is a clearer

acoustic separation between alveolo-palatals and retroflexes. However, we expected that—following language exposure—the sounds located at the extreme points of the length dimension would be equally confusable as the extreme points of the place dimension. This reasoning is motivated by the fact that length distinctions are discriminated relatively more easily than the alveolo-palatal vs. retroflex place distinctions (perhaps since temporal cues are more salient than spectral ones; D. A. Hall et al., 2002; see also Chapter 3). Thus, the gradient distribution and overlap along the length dimension might reduce the relative salience of the length cue and increase confusability between tokens along that dimension. Despite the lack of overlap along the place dimension, the between-cluster confusability is expected to be high due to the fact that this distinction is acoustically very subtle (Nowak, 2006; see also Lisker, 2001; Żygis & Padgett, 2010) and, as confirmed by the study reported in Chapter 3, poorly discriminable by both Mandarin and Korean speakers (but better by the former).

Given the confusability expectations, the data from the bimodal-length language could be interpreted as either a place distinction (alveolo-palatal vs. retroflex) or a length distinction (short vs. long), whereas the data from the unimodal-length language should be interpreted as an unambiguous place distinction with irrelevant variability along the length dimension. These interpretations might, however, vary depending on the listener's language background. Following the predictions of the inductive theory, speakers of Korean should be generally biased toward inferring length-based category distinctions and against inferring place-based category distinctions. Speakers of Mandarin, on the other hand, should be biased toward inferring place-based category distinctions and against inferring length-based category distinctions. Thus, with length distributed bimodally—as in the top graph—Korean speakers might be inclined to interpret the input as two categories along the length dimension, while Mandarin speakers might interpret it as two categories along the place dimension. With input unimodally distributed in length—as in the bottom graph—neither group of speakers should infer that length is contrastive, but Korean speakers are expected to be less inclined to infer a place distinction than Mandarin speakers. The study reported here tested these predictions.

## 6.2.1 Method

### Participants

144 undergraduate students at UC San Diego participated in the experiment for course credit or payment. Half were Korean-English bilinguals, and the other half were Mandarin-English bilinguals. All learned Korean or Mandarin from birth, and reported to be at least competent speakers of those languages. In most cases they had some limited high school and/or college exposure to Spanish or French. Some Mandarin-English bilinguals were also familiar with Taiwanese, mostly through family exposure. All participants reported no history of speech or hearing problems.

### Materials

The materials consisted of nonce words recorded in a soundproof booth by a phonetically-trained native speaker of Polish. The critical items included segments from two classes: alveolo-palatals ([ç], [tç]) and retroflexes ([ʂ], [tʂ]). They were recorded as words with long intervocalic consonants: [açça], [attça], [aʂʂa], [attʂa]. Subsequently, two recordings of each word were chosen, and the consonant length in each word was manipulated to create length continua, each with eight tokens, where durations of consonants ranged from short (140msec) to long (280msec) in a 2:1 long-to-short duration ratio. Each token adjacent on the continuum differed in duration by 20msec. For affricates, the frication portion was held constant throughout the continuum (90msec), and only the closure duration was manipulated (ranging from 50 to 190msec).<sup>1</sup> The fillers resembled the critical items, but included different consonants: [afa], [ava], [axa], [aχa], [aba], [aβa], [asa], [aθa], [ada], [aða], [aβa], [aʃa], [atsa], [adza], [aka], [aqa]. Eight different recordings of each filler word were used in the experiment. There were no length manipulations on fillers.<sup>2</sup>

<sup>1</sup>This is how long affricates are often naturally produced in different languages (Maddieson, 1980; Tarnóczy, 1988; Thurgood & Demenko, 2003; Pycha, 2009).

<sup>2</sup>This meant that most segments in the language were short, which could have provided learners with an additional bias against length distinctions. While this might have affected the overall number of participants' 'different' responses, it should not have had any bearing on

## Procedure

We followed the general procedure of the distributional learning paradigm (Maye & Gerken, 2000; Maye et al., 2002), as applied by Maye and Gerken (2001) in a study with adult participants, where the main idea is that by manipulating the frequency of exposure to sounds that vary along a given dimension, participants can recover the underlying structure along that dimension and, for example, infer two categories when the input is bimodally-distributed, but only one category when the input is unimodal (e.g., as shown in Chapter 5 in a study with English monolinguals).

The general overview of the critical part of the experiment is the following: In training, participants were exposed to a novel language by listening to tokens that varied along the length and place dimensions, as was depicted in Figure 6.1. The place contrast was indicated by including naturally recorded tokens of both alveolo-palatals and retroflexes. The evidence for the length contrast was provided by varying frequency of exposure to different tokens along the length continuum. In testing, participants heard pairs of words that had clear place or length contrasts, and were asked to judge whether these were two different words or two repetitions of the same word. A detailed description of the study is provided below.

Each participant was randomly assigned to one of four conditions: (1) *discrimination* (13 Korean, 13 Mandarin), (2) *filler training* (13 Korean, 13 Mandarin), (3) *bimodal-length training* (23 Korean, 23 Mandarin), (4) *unimodal-length training* (23 Korean, 23 Mandarin). The first two conditions were introduced in order to assess baseline performance. In each condition participants were presented with the same exact testing. The conditions differed only in instructions and/or training provided prior to and in the middle of testing.

The instructions included a short practice. In the *discrimination* condition, the practice consisted of acoustically identical ('same') pairs and acoustically distinct ('different') pairs of words from the new language that were not included in subsequent training nor in testing. In the *training* conditions,

---

the predicted main effect of bimodal vs. unimodal training since both training regimes were identical in this respect.

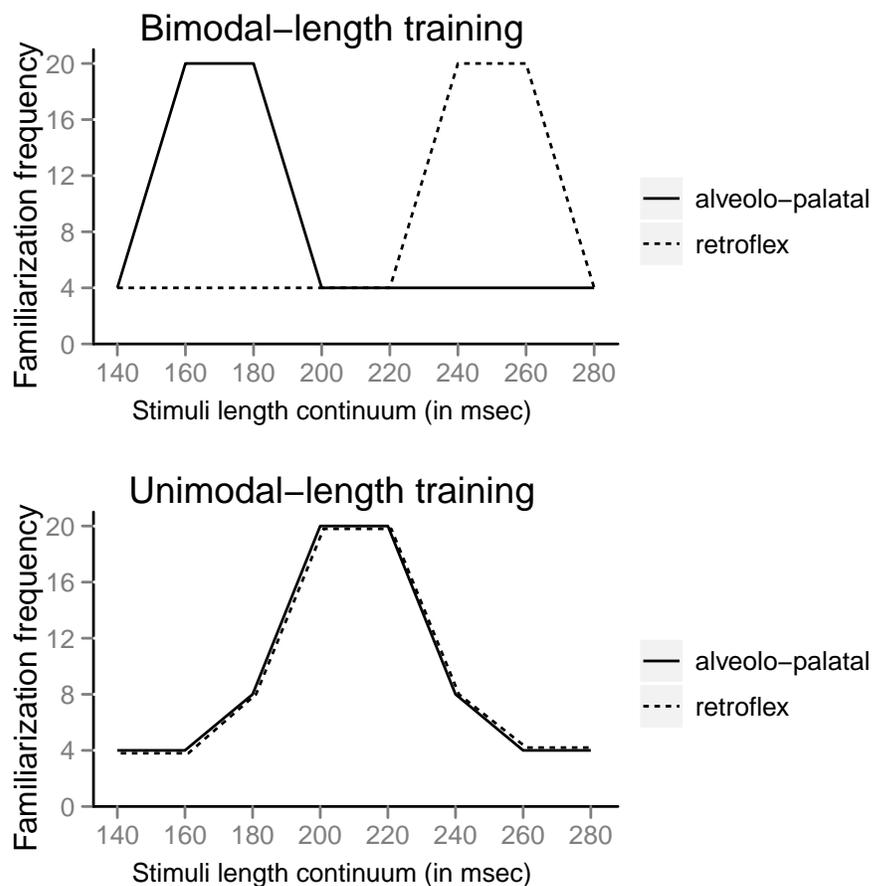
the practice consisted of English words, where ‘different’ words were minimal pairs (e.g., *mass* – *miss*), and ‘same’ words were repetitions of the same word pronounced with different intonations.

In the *discrimination* condition, participants were told that the goal of the experiment was to assess how well they can hear differences between sounds in a new language. There was no exposure to the language besides the testing trials.

In the *training* conditions, participants were told that they would first listen to words in a new language (training) and then would be asked to use what they learned in testing. In training, participants listened to single words presented over headphones and were asked to push a button after hearing each word. The response to a given stimulus triggered the presentation of the following stimulus with a delay of 1 second. There were two training sessions: one prior to testing, and another after the first half of testing. The first training session consisted of a total of 384 words (four repetitions of one 96-trial training block) and lasted about 10 minutes. The second training session consisted of a total of 192 words (two repetitions of one 96-trial training block) and lasted about 5 minutes. Stimulus order was randomized for each participant, and there was a self-terminated break after each block.

In the *filler training* condition, participants were exposed to 12 filler words ([afa], [ava], [axa], [aχa], [aba], [aβa], [asa], [aθa], [ada], [aða], [ava], [aʕa]) with no variability along the length dimension (i.e., all segments were short). One training block included 8 repetitions of each word (a total of 96 trials), where each repetition was a different recording of the word.

In the *bimodal-length training* condition, participants were exposed to words that were either critical or filler items. One training block consisted of 64 critical items (8 tokens from each length continuum type: [ç]–[çç], [tç]–[ttç], [ʂ]–[ʂʂ], [tʂ]–[ttʂ]) and 32 fillers (8 repetitions each of the words [afa], [ava], [axa], [aχa], where each repetition was a different recording of the word). The critical items from the length continua were presented with different frequencies, as illustrated in Figure 6.2 (top): alveolo-palatals were most frequently short, and retroflexes were most frequently long, suggesting a bimodal distribution along

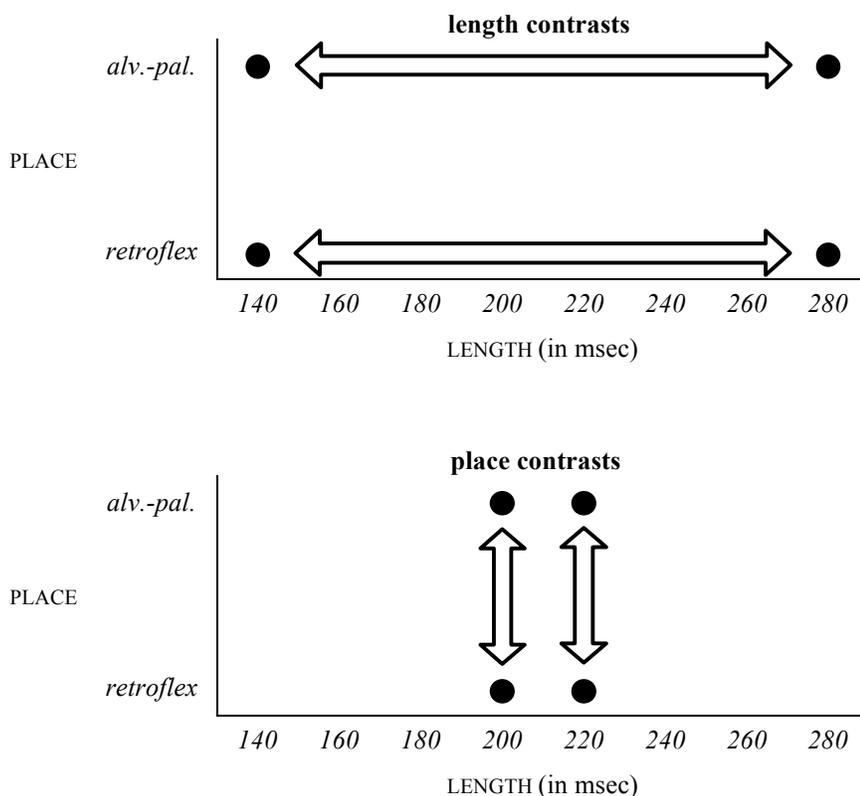


**Figure 6.2:** Critical training stimuli.

the length dimension.

The *unimodal-length training* condition differed from the *bimodal-length training* only in the frequencies of critical items, as in Figure 6.2 (bottom): both alveolo-palatals and retroflexes were most frequently of medium length, indicating a unimodal distribution along the length dimension.

The testing was identical for all participants, and consisted of a same-different AX discrimination task. Participants listened to pairs of words, and were asked to answer whether these were ‘same’ or ‘different’ by pushing one of two buttons. In the *discrimination* condition, participants were instructed to answer ‘different’ whenever they heard any kind of difference between the two words. In the *training* conditions, on the other hand, participants were asked to



**Figure 6.3:** Critical ‘different’ word pairs in testing.

make an intuitive judgment, based on what they learned during training, about what differences counted as ‘different’ in this language and whether the words in a pair were two different words or two repetitions of the same word.

There were two critical contrasts: *length* and *place*. The ‘different’ critical pairs are illustrated in Figure 6.3. For *length*, these were endpoints of each length continuum differing only in length (e.g., [aɕa]–[aɕɕa]), but each word in a pair originated from a different recording of the word. For *place*, these were items of medium length that differed only in place (e.g., [aɕa]–[aʂa]). The ‘same’ pairs were always two different recordings of a word from the same point along the length and place dimensions (e.g., [aɕa]<sub>rec1</sub>–[aɕa]<sub>rec2</sub>). Just like for ‘different’ pairs, only items from the endpoints and the middle of the length continuum were used. For filler ‘different’ pairs, these were two words that differed by

one segment: the contrasts were either in voicing ([afa]–[ava], [atsa]–[adza]), place of articulation ([axa]–[ax̣a], [asa]–[aθa], [aβa]–[aʃa], [aka]–[aqa]), or place and/or manner ([aba]–[aβa], [ada]–[aða]). The ‘same’ pairs were again always two different recordings of the same word.

There was a total of 384 word pairs in testing, which included 6 repetitions of the testing block. The block consisted of 32 critical pairs (16 ‘same’ and 16 ‘different’) and 32 filler pairs (16 ‘same’ and 16 ‘different’). The content of the block was balanced with each pair occurring twice. The words in each pair were separated by an interstimulus interval of 750msec. As with training, stimulus order was randomized for each participant, and there was a self-terminated break after each block. Testing lasted about 20 minutes.

## 6.2.2 Results

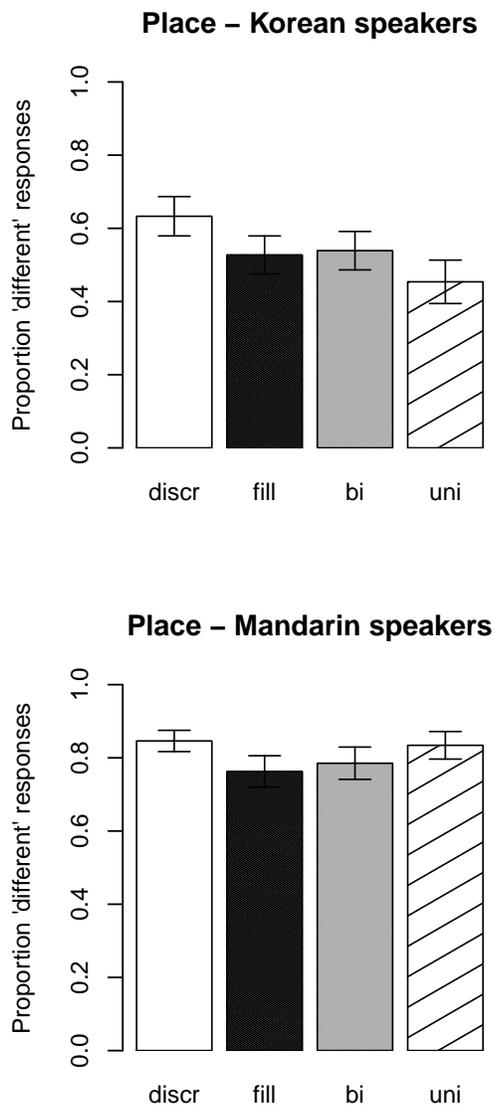
The results from ‘same’ trials are provided in Table 6.1. Participants rarely responded ‘different’ on ‘same’ trials, and there were no significant differences between CONDITIONS.

**Table 6.1:** Proportion of ‘different’ responses on ‘same’ trials.

CONDITION		KOREAN	MANDARIN
<i>Discrimination</i>	<i>length</i>	.11 (.02) <sup>a</sup>	.13 (.03)
	<i>place</i>	.15 (.03)	.14 (.04)
<i>Filler training</i>	<i>length</i>	.10 (.02)	.09 (.03)
	<i>place</i>	.12 (.02)	.10 (.03)
<i>Bimodal-length training</i>	<i>length</i>	.12 (.02)	.12 (.02)
	<i>place</i>	.14 (.02)	.13 (.02)
<i>Unimodal-length training</i>	<i>length</i>	.12 (.02)	.13 (.02)
	<i>place</i>	.15 (.03)	.14 (.02)

<sup>a</sup> Numbers in parentheses indicate standard errors.

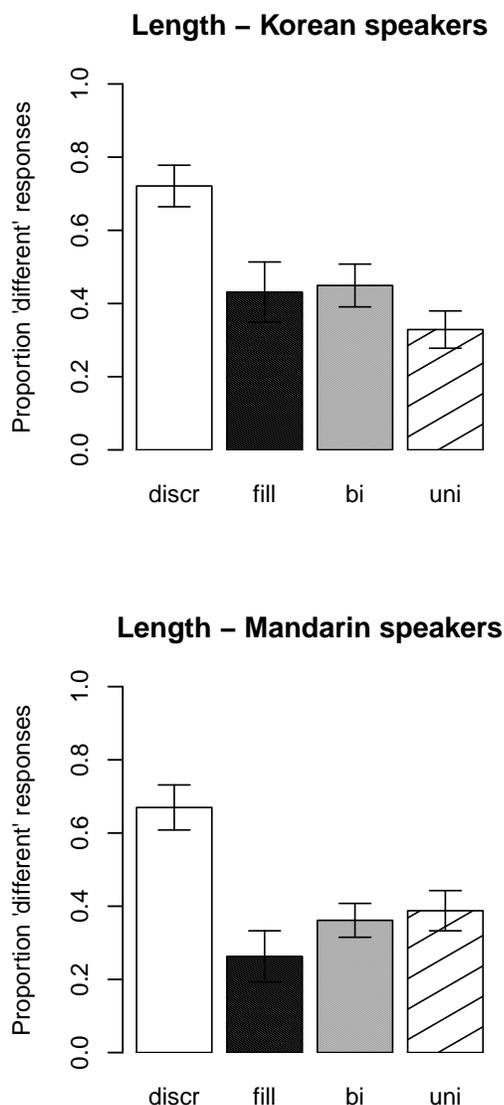
Therefore, we only analyzed responses from ‘different’ trials, using mixed-effects logit models (Jaeger 2008). We included random intercepts for participants and items, and random slopes for participants and items for all effects of interest that were manipulated within participants or within items.



**Figure 6.4:** Proportion of ‘different’ responses on ‘different’ *place* trials.

We controlled for main effects of participants’ dominant language, length of residence in the US, and—for bimodal- vs. unimodal-length comparisons—performance on filler items by adding them as fixed effects to the models.

The results from ‘different’ trials are illustrated in Figures 6.4–6.5 (error bars are standard errors). First, we predicted that—in agreement with their L1 biases—Mandarin speakers should overall give more ‘different’ responses than Korean speakers on the place trials, while the reverse should be true for the



**Figure 6.5:** Proportion of 'different' responses on 'different' *length* trials.

length trials. We examined this in a model with fixed effects of LANGUAGE (Korean, Mandarin) and CONTRAST (*place*, *length*), and found a significant interaction between the two effects ( $p < .001$ ) in the predicted direction: Mandarin speakers responding more 'different' on *place*, and Korean speakers responding more 'different' on *length*. In addition, there was a significant main effect of CONTRAST ( $p < .001$ ), with more 'different' responses for *place* than for *length*, suggesting that the place contrast was perhaps relatively more salient

than the length contrast. Finally, there was a significant main effect of LANGUAGE ( $p < .01$ ): Mandarin speakers gave overall more ‘different’ responses than Korean speakers.

As the next step, we looked at the data from the two baseline conditions, *discrimination* and *filler-training*, to assess how perceptual sensitivity compared to phonetic category judgments with no prior training on *place* or *length* items. We examined this in a model with fixed effects of CONDITION (*discrimination*, *filler-training*), LANGUAGE (*Korean*, *Mandarin*), and CONTRAST (*place*, *length*). We expected at least as many ‘different’ responses in the *discrimination* as in the *filler-training* condition, since perceptual sensitivity should constitute a ceiling for category judgments. We found a significant main effect of CONDITION ( $p < .001$ ) with more ‘different’ responses in *discrimination* than in *filler-training*, which was consistent with our prediction. Furthermore, we expected an interaction between LANGUAGE and CONTRAST, as already found in the overall model, which was indeed significant ( $p < .001$ ). In addition, as in the overall model, we found a significant main effect of CONTRAST ( $p < .001$ ) with more ‘different’ responses on *place* than on *length*. Finally, there was an unexpected significant interaction between CONDITION and CONTRAST ( $p < .001$ ): for *place*, ‘different’ responses were only slightly less frequent in the *filler-training* than in the *discrimination* condition; for *length*, on the other hand, the ‘different’ responses were considerably lower in the *filler-training* than in the *discrimination* condition. This result suggests that after exposure to only short segments in training, the expectations for a length contrast decreased significantly with respect to participants’ perceptual sensitivity. The expectations for a place contrast, on the other hand, did not seem considerably affected by the filler training, and were maintained at nearly the same level as perceptual sensitivity.

Next, we examined the data from all training conditions for fixed effects of CONDITION (*filler*, *bimodal-length*, *unimodal-length*), LANGUAGE (*Korean*, *Mandarin*), and CONTRAST (*place*, *length*).<sup>3</sup> We predicted a three-way interaction, because speakers of Korean — but not Mandarin — should be highly sensitive to

---

<sup>3</sup>The model with the full random effects structure failed to converge. Thus, we removed the interaction between CONDITION and LANGUAGE from random effects for items.

the distribution of length. For *length*, we expected more ‘different’ responses for *bimodal-length training* than for *unimodal-length training*, and we expected the reverse for *place*. However, the analysis revealed no significant three-way interaction. Instead, there was a two-way interaction of LANGUAGE and CONTRAST ( $p < .001$ ), but it did not interact with CONDITION. The key reason why the three-way interaction did not come out as expected is the fact that the Korean/*place*/*unimodal* response was—unlike what we predicted—lower than the Korean/*place*/*bimodal* response.

We followed up on this result by running pairwise CONDITION by CONTRAST comparisons within each language for all training conditions. As found in previous models, in all tests there was a significant main effect of CONDITION ( $ps < .05$ ). For Korean speakers, the only other nearly significant difference was the marginal main effect of CONDITION for *bimodal-length* vs. *unimodal-length* ( $p = .052$ ): more ‘different’ responses after training on *bimodal-length* than on *unimodal-length*. When the data were examined separately for *length* and for *place* trials, there was a significant effect of CONDITION for *length* ( $p < .05$ ), but not for *place*. For *length* trials, the responses in the *filler-training* condition did not differ significantly from either *bimodal-length* or *unimodal-length* (which was perhaps due to the smaller number of participants in the *filler-training* condition). However, if we interpret the results numerically, responses for *bimodal-length* were slightly higher than for *filler-training*, and responses for *unimodal-length* were considerably lower. Overall, this suggests that Korean speakers were sensitive to subtle distributional cues present on *length* in training. For Mandarin speakers, on the other hand, the response pattern was quite different. There were marginal ( $p = .076$ ) and close to marginal ( $p = .13$ ) effects of CONDITION when comparing *filler* vs. *unimodal-length*, and *filler* vs. *bimodal-length* conditions, respectively. Furthermore, contrary to what we found for speakers of Korean, Mandarin speakers’ responses on *length* trials were numerically higher after both *bimodal-length training* and *unimodal-length training* compared to the *filler-training* condition, and this difference was close to significant ( $p = .11$ ) when the *length*-trials data from *bimodal-length* and *unimodal-length* conditions were pooled together. This result suggests that Mandarin speakers were not at-

tending to the distributional length cues in training, but rather increased their ‘different’ responses after any exposure to variability in length. (Note that the smaller number of participants in the *filler-training* condition may place some limits on statistical power in this last analysis.)

### 6.3 Discussion

Taken together, the type of training did not have a clear effect on Mandarin speakers, but—if anything—training on our length/place materials sensitized them overall to subtle differences and increased their proclivity to infer that tokens differing either in place or length are different words. Korean speakers, on the other hand, started out overall fairly sensitive to both distinctions (insofar as they could perceive them), and the main effect of training on the length/place stimuli was to *desensitize* them to differences when the length distribution was unimodal.

The question then remains why Korean speakers had greater proclivity to answer ‘different’ for both contrasts in bimodal-length than in unimodal-length training. At this point, we can speculate that Korean speakers may have tended to infer four categories in the bimodal-length training, not just two. The reasons for this are unclear, but it may be that—despite our initial assumptions—the evidence for place contrasts in training was overall more salient than the evidence for length. This interpretation would be consistent with the discrepancy between our perceptual-discrimination results in the present study, where Korean speakers were nearly as accurate in identifying place distinctions as they were in identifying length distinctions, and those from Chapter 3, where Korean speakers were far more accurate on length contrasts than on place contrasts.

Another possibility is that the correlation between the length and place cues in bimodal-length training (i.e., alveolo-palatals were mostly short and retroflexes were mostly long) actually helped Korean speakers become more sensitized to the place distinction. That is, Korean speakers might have first picked up on the length differences and categorized the stimuli based on length. But once short and long tokens were teased apart, listeners’ attention might

have been drawn to the correlated place cues, which then led to their enhanced discriminability and separate categorization. This is in line with the domain-general learning mechanism of *acquired distinctiveness* (Lawrence, 1949) that has already been proposed to play a role in native-language phonetic category learning (Yeung & Werker, 2009). This mechanism crucially relies on the fact that different perceptual cues often occur in unique contexts. In particular, the occurrences of value A in context X and value B in context Y are believed to highlight cues that differentiate value A from value B, leading to facilitation in discriminating between A and B.

Overall, the results presented in this chapter replicate the finding presented in Chapter 3 that non-native speech perception is guided by L1-derived perceptual biases. The results also provide new evidence that perceptual abilities do not fully predetermine how novel sounds are categorized in a new language, but instead—as predicted by the inductive theory—that categorization of L2 sounds is a result of combining L1-shaped perceptual biases with the distributional information from L2 input. Crucially, the interpretation of L2 statistics varies depending on the expectations that learners have about L2 given their previous language background. The results reported here are, however, far from conclusive. There are many unanswered questions regarding how exactly distributional information and prior L1 biases interact, and what factors might play a role in how each piece of information is weighed. Nevertheless, the study presented here can be considered a first step in understanding the complex nature of how previous linguistic experience affects the use of distributional evidence in the acquisition of phonetic categories in a new language.

## 6.4 Acknowledgements

Chapter 6 is a revised version of Pajak & Levy (to appear) [“Distributional learning of L2 phonological categories by listeners with different language backgrounds”. In A. K. Biller, E. Y. Chung, & A. E. Kimball (Eds.), *Proceedings of the 36th Boston University Conference on Language Development* (pp. 400–413). Somerville, MA: Cascadilla Press]. The dissertation author was

the primary investigator and author of these papers. This work was presented at the 36th Boston University Conference on Language Development.

# Chapter 7

## Conclusion

In this dissertation I proposed a general framework to study native and non-native language acquisition throughout the life span, based on the principles of hierarchical inductive inference in learning. I used this general framework to construct a more detailed inductive theory of phonological acquisition in the specific areas of speech sound perception and phonetic category learning. The theory was intended to bear on three related aspects of language learning in both infancy and adulthood: (1) perceptual reorganization in infancy, (2) non-native speech perception by naive listeners, and (3)  $L_n$  phonetic category learning.

I proposed that phonological learning in both native and non-native languages is guided by the principles of hierarchical inductive inference, where learners—in addition to acquiring the specific distributional properties of the language being learned—make higher-order inferences on at least two levels of abstraction: inferences about the structure of the language's sound system as a whole, as well as inferences about the likely properties of sound systems in other languages. I argued that the observed patterns of non-native speech perception by naive listeners are a by-product of these higher-level generalizations. Furthermore, I argued that learning phonetic categories in additional languages relies on combining the distributional information extracted from the  $L_n$  speech signal with previously formed generalizations about likely properties of other languages – the latter acting as inductive biases in learning.

## 7.1 Summary of Main Findings

I focused on one type of generalization that learners might make about a language's sound system: inferences regarding the underlying set of informative phonetic dimensions from which the system as a whole is constructed. I proposed that inferring high informativity of a given phonetic dimension to distinguish between phonetic categories leads to enhanced perceptual acuity for any contrasts along that dimension. I suggested that this implicit learning strategy may benefit acquisition because languages often use the same phonetic dimension to distinguish between multiple sets of categories (e.g., voiced vs. voiceless consonants of different place of articulation). If sensitivity to phonetic dimensions is not tied to any particular segments, then learning a contrast for one set of categories should lead to facilitation in learning analogous contrasts for other sets of categories within that language.

This general dimension-based acuity obtained through native-language perceptual tuning was then predicted to constrain the processing of non-native sounds by naive listeners. In particular, discrimination of non-native contrasts along dimensions previously inferred as informative should be enhanced relative to non-native contrasts along other dimensions. This was confirmed in the study reported in Chapter 3, where speakers of length-informative languages (Korean, Vietnamese, and Cantonese) all outperformed speakers of a length-uninformative language (Mandarin) on discrimination of consonant length contrasts. Most strikingly, Vietnamese and Cantonese only have length contrasts on vowels, and not consonants, suggesting that they generalized phonetic dimension informativity (length) across very dissimilar segment classes (vowels and consonants). Interestingly, these perceptual advantages do not easily extend to vocabulary learning, as demonstrated by the word learning study with Korean and Mandarin speakers (Chapter 4): only faster learners within the Korean-speaker group showed enhanced learning of minimal-pair words differentiated by consonant length, suggesting that the ability to take advantage of fine phonetic detail during vocabulary learning might only be possible for adults with better cognitive abilities or higher motivation.

Furthermore, I argued that learning of non-native languages is based on similar underlying mechanisms as native-language acquisition. If that is indeed the case, learning a non-native distinction along a phonetic dimension that is uninformative in the native language should lead to enhanced perceptual acuity to any distinctions along that dimension. This, in turn, should produce facilitation in picking up on the distributional properties of any sounds along that dimension, and thus to enhanced learning of additional phonetic categories that make use of that dimension. The studies reported in Chapter 5 confirmed this prediction: monolingual English speakers who were exposed to distributional evidence that suggested a novel length distinction for one set of segments, not only learned the distinction for the trained segments, but also generalized it to a novel set of segments. This result indicated generalization of phonetic dimension informativity across dissimilar segment classes (sonorants and voiceless fricatives) within a new language being learned.

Finally, interpreting distributional properties of the speech signal in a novel language was predicted to be affected by previous inferences about the informativity of phonetic dimensions. In particular, learners should be biased to expect phonetic category distinctions along dimensions that are informative in their native language. In Chapter 6 I showed that this is indeed the case: Korean and Mandarin speakers drew different inferences about novel length and place distinctions based on exposure to the same distributional cues, each group in line with their predicted native-language biases. This result thus suggests that L2 learners incrementally combine their L1 biases with L2 input statistics.

Overall, the studies reported in this dissertation demonstrate that language learning goes beyond the acquisition of specific phonetic categories, and includes higher-order generalizations regarding the relative importance of phonetic dimensions in the language as a whole. As such, these results provide support for my proposal to view language acquisition as a process of hierarchical inductive inference.

## 7.2 Implications

The hierarchical inductive theory proposed in this dissertation has strong implications for our understanding of acquisition of both native and non-native languages. In addition, however, the proposed approach allows us to go beyond learning, and also probe learners' mental representations of languages they already speak. Below I discuss the implications of the theory in both of these areas.

### 7.2.1 Language Learning

The inductive theory unifies native and non-native language acquisition. Traditionally, these two areas of research have been regarded as very distinct, and have generally been studied separately from each other. Second language researchers did try to compare many aspects of L1 and L2 acquisition, but the theories of L2 learning were generally not intended to explain any underlying links between the two. While native-language acquisition in infancy and additional language learning later in life are clearly different in many crucial respects (cognitive maturation, motivation, learning environment, etc.), I believe that there are similar cognitive and learning mechanisms underlying the acquisition of any language at any point in an individual's life. This dissertation constitutes one of the first steps in uncovering these potential mechanisms, which are likely derived from the domain-general learning mechanisms that humans recruit in many other learning tasks (Tenenbaum et al., 2011).

My theoretical proposal reconceptualizes language learning—and phonological acquisition in particular—as hierarchical inductive inference. First of all, this proposal has implications for our understanding of what happens during perceptual reorganization that tunes infants to the sounds of their native language. Instead of simply viewing the reorganization process as a rearrangement of perceptual sensitivities—enhanced at native-language phonetic category boundaries and diminished within categories—this proposal adds an additional level of implicit inferences over the sound system as a whole. The role of these inferences is to guide this rearrangement process in a way that will yield optimal allocation of cognitive and neural resources to auditory process-

ing of native-language sounds.

Second, the inductive theory is an alternative to classic approaches to non-native speech perception and learning, in which processing and learning of novel sounds proceeds through their *mapping* onto L1 phonetic categories that are acoustically or articulatorily most similar: if two novel sounds are mapped onto a single L1 category, then their discrimination and learning is diminished with respect to sounds that are mapped onto distinct L1 categories. These approaches provide an intuitive explanation of perceptual difficulties in non-native sound perception and learning, but they cannot explain the patterns of generalization reported in this dissertation. Furthermore, they do not offer any satisfactory answer to why learning would even work this way: given that with sufficient exposure the adult perceptual system is able to adapt to new sound systems, why would L2 learners implicitly link all L2 sounds with existing L1 categories? It seems natural to expect that the L1 sound system will affect the way L2 sounds are processed, but creating L2 representations based entirely on the L1-attuned perceptual space does not straightforwardly lead to any clear learning advantages. If anything, establishing L2 phonetic categories in a relationship with L1 categories is likely to hinder the formation of accurate L2 categories – which is exactly the point of the mapping approaches.

However, from the point of view of rational approaches to learning, the underlying mechanisms of learning must at least in some respects be optimal under given conditions and for the specific task at hand. Therefore, it is natural to expect that  $L_n$  learners—while probably cognitively and neurally more constrained than infants—make implicit use of their available resources in a way that maximizes benefits in learning. The hierarchical inductive theory provides an account of learning that follows this exact logic.

My proposal also has clear implications for the area of third and additional language acquisition. Until fairly recently, it had been generally assumed that there is no significant difference between second vs. third or additional (L3+) language acquisition. However, in the past decade it has become widely accepted that there is indeed a fundamental distinction between L2 and L3+ acquisition due to more possibilities for between-language interactions in the

latter process. There has been a growing interest in this recently established subfield, investigating, among other things, the complex cross-language influences in a multilingual mind (see e.g., Edwards, 1994; Cenoz & Genesee, 1998; Cenoz & Jessner, 2000; Cenoz et al., 2001; De Angelis, 2007). The research in L3+ phonological acquisition has been so far very limited in scope and methodology, with many studies being based on introspection or observational data from a very small number ( $\leq 5$ ) of participants. However, some work has shown that L3+ speech production can be influenced not only by the native language, but also by previously learned non-native languages (Rivers, 1979; Singh & Carroll, 1979; Hammarberg & Hammarberg, 1993; Hammarberg, 2001; Pyun, 2005; Wrembel, 2010).

The hierarchical inductive theory provides an account of these kinds of cross-language interactions, and it offers a principled way of predicting what sorts of patterns might be expected in L3+ speech perception and production based on the exact properties of previously learned languages. In particular, learners' higher-order inferences regarding the likely underlying structure of any language are predicted to be informed by the properties of each language learned: language characteristics that are stable across several languages (e.g., the informativity of the second formant to distinguish between front and back vowels) will increase learners' confidence about their invariance in any language encountered in the future; language properties that vary from language to language (e.g., the informativity of segmental length) will, on the other hand, lower learners' confidence about their role in any other language. Learners' implicit expectations are likely to also be affected by multiple other factors, such as the perceived relatedness between the languages, which have previously been found to affect L3+ speech production (S. Williams & Hammarberg, 1998; Cenoz, 2001; Hammarberg, 2001). The inductive theory accounts for these results because—if two languages are perceived to be closely related—several of their properties should be inferred as likely to be shared. The necessity of incorporating these additional factors into the theory indicates, however, that the hierarchical structure of learners' generalizations is even more complex than what I have proposed in this dissertation, and more work is required to identify

all the relevant variables and understand the relationships between them.

Finally, although the results reported here only bear on one type of possible inference about learners' native-language sound system, the hierarchical inductive theory predicts that learners make other types of higher-order generalizations about linguistic structures that should affect non-native language processing, both in the sound domain and in other aspects of language. Thus, this work contributes to the broader literature investigating generalization of linguistic knowledge by both children and adults in different language domains (e.g., Xu & Tenenbaum, 2007; Wonnacott, Newport, & Tanenhaus, 2008; Gerken, 2010), and it provides theoretical unification with many other domains in which inductive approaches to learning have proven fruitful.

## 7.2.2 Phonological Representations

In addition to uncovering the underlying mechanisms of language learning, the proposed approach can also help us probe the degree of abstractness of phonological knowledge in the native language, which has direct implications for any phonological theory.

Phonological representations are typically thought of as abstract: for example, phonemes are argued to be comprised of abstract subphonemic units of some sort, whether distinctive features (e.g., Chomsky & Halle, 1968), articulatory gestures (e.g., Browman & Goldstein, 1989), or acoustic-phonetic dimensions (e.g., Pierrehumbert, 2000). There is direct evidence that people are sensitive to these phonemic and subphonemic units. A good example comes from studying speech errors: the phrase "big and fat" can sometimes be mispronounced by transposing whole segments (**big and fat** → **fig and bat**, where two initial stops are switched), but also just parts of segments (**big and fat** → **pig and vat**, where only the voicing appears to be switched between the initial stops; Fromkin, 1973). This kind of evidence suggests that in speakers' mind not only segments are represented as single units, but that subphonemic properties such as voicing also have some psychological reality in that they can be abstracted away from individual segments, even in running speech.

The line of research proposed in this dissertation can provide further evidence regarding the abstractness of phonological representations: by investigating patterns of cross-segment generalization, it can inform our understanding of what subsegmental properties people are sensitive to and what exact properties are represented in the mind as shared among classes of segments. For example, the results showing that length generalizes across segments suggest that this particular subsegmental property is represented psychologically as independent from individual segments, abstracting away from the segments' acoustic-phonetic (or gestural) properties. Thus, these results provide some additional insight into how length should be formally represented in any phonological theory.

### 7.3 Future Directions

This dissertation provides a foundation for a research program dedicated to studying the role of hierarchical inductive inference in language learning across the life span. Future work will first and foremost consist of elaborating the theory: reducing the number of simplifying assumptions and spelling out precise predictions.

One way of formalizing the theory is to implement it computationally. The basic assumptions of the inductive theory can be naturally interpreted in terms of Bayesian inference, where previous phonological knowledge gives the prior distribution over the range of potentially relevant phonetic features in new languages. This is then combined with distributional information from a novel language to form a new posterior distribution, which can be used as a prior for future language learning. I plan to implement these assumptions in a computational model with the goal of predicting what exact inferences a rational learner would make on the basis of known cross-linguistic phonological patterns in order to optimize future language learning.

One particular aspect of the theory that I am interested in elaborating further concerns the issue of phonetic dimension informativity and generalizability. As discussed briefly in section 2.3, there are many variables that can po-

tentially affect learners' inferences about the degree of informativity of different phonetic dimensions. To take one example, one might imagine that the incremental process of inferring phonetic dimension informativity as a language is being acquired might be affected by the number of segments and the segment class variability along that dimension. For instance, if a learner encounters two sets of segmental categories distinguished by a given dimension where all segments belong to the same segment class (e.g., stops), then the inferred informativity of that dimension might at that point remain fairly low and not immediately generalize to other segment classes. On the other hand, exposure to two sets of categories that vary along the same dimension and belong to different segment classes (e.g., stops and fricatives) might trigger higher generalizability of the dimension to other segment classes (e.g., affricates). Additionally, as more segments are added to the mental representations of a language, this effect will be amplified. Thus, more generally, as the number and range of segments varying along a given dimension increases, the inferred informativity of that dimension is predicted to rise.

Another part of the theory that has not been fully developed in this dissertation concerns third and additional language acquisition. One specific aspect that I am particularly interested in investigating is the prediction that bi-/multilingual learners might be able to make generalizations over properties of two or more languages. A good example concerns the phonetic dimension of VOT, which distinguishes between voiced and voiceless stop consonants (e.g., [b] and [p]). This dimension has been of great interest to language researchers because there is cross-linguistic variation in where exactly the boundary between voiced and voiceless stops is along this dimension. For example, Spanish speakers tend to place this boundary at around 0msec, while the corresponding boundaries in English and Mandarin are both around 35msec (Lisker & Abramson, 1970; L. Williams, 1977; P. Keating, 1984; Chen, Chao, & Peng, 2007). The novel prediction of the inductive theory is that speakers of two languages where the category boundary between voiced and voiceless stops differs (such as English-Spanish bilinguals), should have an advantage in learning additional novel contrasts along the VOT dimension compared to speakers of two

languages where the category boundary is roughly at the same place (such as English-Mandarin bilinguals). This is because only the former group has linguistic experience supporting the inference that languages may differ in the position of the category boundary.

Finally, the proposed theory is very general in scope, and can thus be applied to learning in many different language domains. A natural extension of this research is then to use the proposed framework to investigate learners' inferences in domains such as syntax or morphology (e.g., regarding morpho-phonological alternations or word and clitic ordering), both in comprehension and production. For example, analogously to VOT, speakers of two languages that have different word orders (e.g., SOV and SVO) might make inferences about word order in any language that go beyond these two specific orders learned. That is, learners might infer that languages can vary in this property, which would in turn facilitate learning of an additional language that has a yet different word order (e.g., VSO).

I have outlined only a few specific directions emerging from the proposed inductive theory that I would like to pursue in the future. In the long term, however, this line of research is intended as a comprehensive investigation of the process of language acquisition, including learning in any linguistic domain and at any point in human development.

## References

- Abramson, A. S., & Lisker, L. (1970). Discriminability along the voicing continuum: cross-language tests. In *Proceedings of the 6th international congress of phonetic sciences* (pp. 569–573). Academia.
- Al-Tamimi, F., Abu-Abbas, K. H., & Tarawnah, R. (2010). Jordanian arabic final geminates: An experimental clinical phonetic study. *Poznań Studies in Contemporary Linguistics*, 46(2), 111–125.
- Bauer, R. S., & Benedict, P. K. (1997). *Modern Cantonese phonology*. Berlin/New York: Mouton de Gruyter.
- Benus, S., Smorodinsky, I., & Gafos, A. (2003). Gestural coordination and the distribution of English ‘geminates’. In *Proceedings of the 27th Annual Penn Linguistics Colloquium* (pp. 33–46). Philadelphia, PA: University of Pennsylvania.
- Best, C. T. (1993). Emergence of language-specific constraints in perception of non-native speech: a window on early phonological development. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. MacNeilage, & J. Morton (Eds.), *Developmental neurocognition: speech and face processing in the first year of life*. Dordrecht, The Netherlands: Kluwer.
- Best, C. T. (1994). The emergence of native-language phonological influence in infants: A perceptual assimilation model. In J. Goodman & H. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 167–224). Cambridge: MIT Press.
- Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: issues in cross-language research* (pp. 171–204). Timonium, MD: York Press.
- Best, C. T., & Hallé, P. A. (2010). Perception of initial obstruent voicing is influenced by gestural organization. *Journal of Phonetics*, 38, 109–126.
- 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, 775–794.

- Best, C. T., McRoberts, G. W., & Sithole, N. M. (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., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics*, 20, 305–330.
- Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: in honor of James Emil Flege* (pp. 13–34). Amsterdam/Philadelphia: John Benjamins.
- Blicher, D. L., Diehl, R. L., & Cohen, L. B. (1990). Effects of syllable duration on the perception of the Mandarin Tone 2/Tone 3 distinction: evidence of auditory enhancement. *Journal of Phonetics*, 18, 37–49.
- Bohn, O.-S. (1995). Cross-language speech perception in adults: first language transfer doesn't tell it all. In W. Strange (Ed.), *Speech perception and linguistic experience: issues in cross-language research* (pp. 275–300). Timonium, MD: York Press.
- Bohn, O.-S., & Munro, M. J. (Eds.). (2007). *Language experience in second language speech learning: in honor of James Emil Flege*. Amsterdam/Philadelphia: John Benjamins.
- Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tokhura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/. IV: Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America*, 101, 2299–2310.
- Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology*, 6, 201–251.
- Brown, C. (1997). *Acquisition of segmental structure: Consequences for speech perception and second language acquisition*. Unpublished doctoral dissertation, McGill University.
- Brown, C. (2000). The interrelation between speech perception and phonological acquisition from infant to adult. In *Speech perception and phonological acquisition* (pp. 4–63).
- Burnham, D. K. (1986). Developmental loss of speech perception: exposure to and experience with a first language. *Applied Psycholinguistics*, 7, 207–240.
- Cenoz, J. (2001). The effects of linguistic distance, L2 status and age on cross-linguistic influence in third language acquisition. In J. Cenoz, B. Hufeisen,

- & U. Jessner (Eds.), *Cross-linguistic influence in Third Language Acquisition: psycholinguistic perspectives* (pp. 8–20). Clevedon: Multilingual Matters.
- Cenoz, J., & Genesee, F. (Eds.). (1998). *Beyond bilingualism: multilingualism and multilingual education*. Clevedon: Multilingual Matters.
- Cenoz, J., Hufeisen, B., & Jessner, U. (Eds.). (2001). *Cross-linguistic influence in Third Language Acquisition: psycholinguistic perspectives*. Clevedon: Multilingual Matters.
- Cenoz, J., & Jessner, U. (Eds.). (2000). *English in Europe: the acquisition of a third language*. Clevedon: Multilingual Matters.
- Cenoz, J., & Valencia, J. F. (1994). Additive trilingualism: evidence from the Basque Country. *Applied Psycholinguistics*, 15, 195–207.
- Chater, N., & Manning, C. D. (2006). Probabilistic models of language processing and acquisition. *Trends in Cognitive Science*, 10(7), 335–344.
- Chen, L.-m., Chao, K.-Y., & Peng, J.-F. (2007). VOT productions of word-initial stops in Mandarin and English. In *Proceedings of the 19th Conference on Computational Linguistics and Speech Processing* (pp. 303–317). Taipei, Taiwan.
- Choi, D.-I. (1995). Korean “tense” consonants as geminates. *Kansas Working Papers in Linguistics*, 20, 25–38.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge.
- Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper and Row.
- Davidson, L., Shaw, J., & Adams, T. (2007). The effect of word learning on the perception of non-native consonant sequences. *Journal of the Acoustical Society of America*, 122(6), 3697–3709.
- Davis, S. (1999). On the representation of initial geminates. *Phonology*, 16, 93–104.
- De Angelis, G. (2007). *Third or additional language acquisition*. Clevedon: Multilingual Matters.
- de Boer, B., & Kuhl, P. K. (2003). Investigating the role of infant-directed speech with a computer model. *Acoustic Research Letters Online*, 4(4), 129–134.
- Diamond, A., Werker, J. F., & Lalonde, C. E. (1994). Toward understanding commonalities in the development of object search, detour navigation, categorization, and speech perception. In C. Dawson & K. W. Fischer (Eds.), *Human behavior and the developing brain* (pp. 380–426). New York, NY: Guilford Press.

Dietrich, C., Swingley, D., & Werker, J. F. (2007). Native language governs interpretation of salient speech sound differences at 18 months. *Proceedings of the National Academy of Sciences*, 104, 16027-16031.

Dmitrieva, O. (2007). *Consonant length in Russian: factors affecting variability in production*. Poster at the Workshop on Variation, Gradience and Frequency in Phonology.

Edwards, J. (1994). *Multilingualism*. London: Routledge.

Eimas, P. D. (1978). Developmental aspects of speech perception. In R. Held, H. W. Leibowitz, & H. L. Teuber (Eds.), *Handbook of sensory physiology* (Vol. 8). Berlin: Springer-Verlag.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (Eds.). (1996). *Rethinking innateness: A connectionist perspective on development*. MIT Press.

Feldman, N. H., Griffiths, T. L., & Morgan, J. L. (2009). Learning phonetic categories by learning a lexicon. In *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 2208–2213). Austin, TX: Cognitive Science Society.

Feldman, N. H., Myers, E., White, K., Griffiths, T. L., & Morgan, J. L. (2011). Learners use word-level statistics in phonetic category acquisition. In *Proceedings of the 35th Boston University Conference on Language Development* (pp. 197–209). Somerville, MA: Cascadilla Press.

Fennell, C. T., Waxman, S. R., & Weisleder, A. (2007). With referential cues, infants successfully use phonetic detail in word learning. In *Proceedings of the 31st Boston University Conference on Language Development*. Cascadilla Press.

Fennell, C. T., & Werker, J. F. (2003). Early word learners' ability to access phonetic detail in well-known words. *Language and Speech*, 46(2-3), 245–264.

Flege, J. E. (1988). The production and perception of foreign language speech sounds. In H. Winitz (Ed.), *Human communication and its disorders: a review – 1988* (pp. 224–401). Norwood, NJ: Ablex Publishing.

Flege, J. E. (1992). The intelligibility of English vowels spoken by British and Dutch talkers. In R. D. Kent (Ed.), *Intelligibility in speech disorders: theory, measurement, and management* (pp. 157–232). Amsterdam, The Netherlands: John Benjamins.

Flege, J. E. (1995). Second-language speech learning: theory, findings and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: issues in cross-language research* (pp. 229–273). Timonium, MD: York Press.

- Flege, J. E., & Eefting, W. (1986). Linguistic and developmental effects on the production and perception of stop consonants. *Phonetica*, *43*, 155–171.
- Flege, J. E., & Eefting, W. (1987). Production and perception of English stops by native Spanish speakers. *Journal of Phonetics*, *15*, 67–83.
- Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second language acquisition. *Journal of Memory and Language*, *41*(1), 78–104.
- Fromkin, V. A. (1973). The non-anomalous nature of anomalous utterances. In V. A. Fromkin (Ed.), *Speech errors as linguistic evidence* (pp. 215–242). The Hague/Paris: Mouton.
- Gerken, L. (2010). Infants use rational decision criteria for choosing among models of their input. *Cognition*, *115*, 362–366.
- Gibson, E. J. (1969). *Principles of perceptual learning and development*. Englewood Cliffs, NJ: Prentice-Hall, Inc.
- Gibson, E. J. (1991). *An Odyssey in learning and perception*. Cambridge, MA: Bradford Books (MIT Press).
- Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston, MA: Houghton-Mifflin.
- Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, MA: Houghton-Mifflin.
- Goad, H., & White, L. (2006). Ultimate attainment in interlanguage grammars: a prosodic approach. *Second Language Research*, *22*(3), 243–268.
- Gómez, R. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*(5), 431–436.
- Goodman, N. (1955). *Fact, fiction, and forecast*. Cambridge, MA: Harvard University Press.
- Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds “L” and “R”. *Neuropsychologia*, *9*, 317–323.
- Goudbeek, M., Cutler, A., & Smits, R. (2008). Supervised and unsupervised learning of multidimensionally varying non-native speech categories. *Speech Communication*, *50*, 109–125.
- Hahm, H.-J. (2007). The effects of following vowel on Korean fricatives. *Linguistic Research*, *24*(1), 57–82.

- Hakuta, K., Bialystok, E., & Wiley, E. (2003). Critical evidence: a test of the critical-period hypothesis for second-language acquisition. *Psychological Science*, *14*(1), 31–38.
- Hall, D. A., Johnsrude, I. S., Haggard, M. P., Palmer, A. R., Akeroyd, M. A., & Summerfield, A. Q. (2002). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex*, *12*, 140–149.
- Hall, K. C. (2009). *A probabilistic model of phonological relationships from contrast to allophony*. Unpublished doctoral dissertation, Ohio State University.
- Hallé, P. A., Best, C. T., & Levitt, A. (1999). Phonetic vs. phonological influences on French listeners' perception of American English approximants. *Journal of Phonetics*, *27*, 281–306.
- Hamann, S. (2004). Retroflex fricatives in Slavic languages. *Journal of the International Phonetic Organization*, *34*(1), 53–67.
- Hammarberg, B. (2001). Roles of L1 and L2 in L3 production and acquisition. In J. Cenoz, B. Hufeisen, & U. Jessner (Eds.), *Cross-linguistic influence in Third Language Acquisition: psycholinguistic perspectives* (pp. 21–41). Clevedon: Multilingual Matters.
- Hammarberg, B., & Hammarberg, B. (1993). Articulatory re-setting in the acquisition of new languages. *Reports from the Department of Phonetics, University of Umea, Phonum*, *2*, 61–67.
- Han, J.-I. (1992). On the Korean tensed consonants and tensification. In *Proceedings of the 28th Annual Meeting of the Chicago Linguistic Society* (pp. 206–223).
- Hancin-Bhatt, B. J. (1994). Segment transfer: A consequence of a dynamic system. *Second Language Research*, *10*, 241–269.
- Hansen Edwards, J. G., & Zampini, M. L. (Eds.). (2008). *Phonology and second language acquisition*. Amsterdam/Philadelphia: John Benjamins.
- Hawkins, R., & Chan, C. Y.-H. (1997). The partial availability of universal grammar in second language acquisition: the 'failed functional feature hypothesis'. *Second Language Research*, *13*, 187–226.
- Hawkins, R., & Liszka, S. (2003). Locating the source of defective past tense marking in advanced L2 English speakers. In R. van Hout, A. Hulk, F. Kuiken, & R. Towell (Eds.), *The lexicon-syntax interface in second language acquisition* (pp. 21–44). Amsterdam/Philadelphia: John Benjamins.
- Hayes, B. (1989). Compensatory lengthening in moraic phonology. *Linguistic Inquiry*, *20*, 253–306.

- Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged 6–12. *Journal of Phonetics*, 28, 377–396.
- Hume, E., Muller, J., & van Engelenhoven, A. (1997). Non-moraic geminates in Leti. *Phonology*, 14, 371–402.
- Hyman, L. (1985). *A theory of phonological weight*. Dordrecht: Foris.
- Jaeger, T. F. (2008). Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed effects models. *Journal of Memory and Language*, 59, 434–446.
- Jamieson, D., & Morosan, D. (1986). Training nonnative speech contrasts in adults: acquisition of the /ð/ - /θ/ contrast by francophones. *Perception and Psychophysics*, 40, 205–215.
- Jassem, W. (2003). Polish. *Journal of the International Phonetic Association*, 33(1), 103–108.
- Jusczyk, P. W. (1985). On characterizing the development of speech perception. In J. Mehler & R. Fox (Eds.), *Neonate cognition: Beyond the blooming buzzing confusion*. Hillsdale, NJ: Erlbaum.
- Jusczyk, P. W. (1992). Developing phonological categories from the speech signal. In C. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: models, research, implications* (pp. 17–64). Timonium, MD: York Press.
- Kao, D. L. (1971). *Structure of the syllable in Cantonese*. The Hague: Mouton.
- Kaushanskaya, M. (2012). Cognitive mechanisms of word learning in bilingual and monolingual adults: The role of phonological memory. *Bilingualism: Language and Cognition*, 15(3), 470–489.
- Kaushanskaya, M., & Marian, V. (2009). The bilingual advantage in novel word learning. *Psychonomic Bulletin & Review*, 16(4), 705–710.
- Kawahara, S. (2007). Sonorancy and geminacy. In *University of Massachusetts Occasional Papers in Linguistics 32: Papers in Optimality III* (pp. 145–186). Amherst: GLSA.
- Kaye, A. (2005). Geminata in English. *English Today*, 21, 43–55.
- Keating, P. (1984). Phonetic and phonological representation of stop consonant voicing. *Language*, 60(2), 286–319.
- Keating, P. A. (1991). Coronal places of articulation. In C. Paradis & J.-F. Prunet (Eds.), *Phonetics and phonology. The special status of coronals: internal and external evidence* (pp. 29–48). San Diego: Academic Press.

- Kemp, C., Perfors, A., & Tenenbaum, J. B. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science*, 10(3), 307–321.
- Kim, Y.-S. (2002). On non-moraic geminates. *Studies in Phonetics*, 8(2), 187–200.
- Klatt, D. H. (1989). Review of selected models of speech perception. In W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 169–226). Cambridge, MA: MIT Press.
- Kruschke, J. K. (1992). ALCOVE: an exemplar-based connectionist model of category learning. *Psychological Review*, 99(1), 22–44.
- Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics*, 50(2), 93–107.
- Kuhl, P. K. (1992). Psychoacoustics and speech perception: internal standards, perceptual anchors, and prototypes. In L. A. Werner & E. W. Rubel (Eds.), *Developmental psychoacoustics* (pp. 293–332). Washington, DC: American Psychological Association.
- Kuhl, P. K. (1994). Learning and representation in speech and language. *Current Opinion in Neurobiology*, 4(6), 812–822.
- Kuhl, P. K. (2000a). Language, mind, and brain: Experience alters perception. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences (2nd ed.)* (pp. 99–115). Cambridge, MA: MIT Press.
- Kuhl, P. K. (2000b). A new view of language acquisition. *Proceedings of the National Academy of Sciences*, 97(22), 11850–11857.
- Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews Neuroscience*, 5, 831–843.
- Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society*, 363, 979–1000.
- Kuhl, P. K., & Iverson, P. (1995). Linguistic experience and the "perceptual magnet effect". In W. Strange (Ed.), *Speech perception and linguistic experience: issues in cross-language research* (pp. 121–154). Timonium, MD: York Press.
- 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.

- Lacerda, F. (1995). The perceptual magnet-effect: An emergent consequence of exemplar-based phonetic memory. In K. Ellenius & P. Branderud (Eds.), *Proceedings of the 13th International Congress of Phonetic Sciences* (pp. 140–147). Stockholm: KTH and Stockholm University.
- Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Oxford, UK; Cambridge, MA: Blackwell.
- Lalonde, C. E., & Werker, J. F. (1995). Cognitive influences on cross-language speech perception in infancy. *Infant Behavior and Development*, 18, 459–475.
- Lawrence, D. H. (1949). Acquired distinctiveness of cues: I. Transfer between discriminations on the basis of familiarity with the stimulus. *Journal of Experimental Psychology*, 39, 770–784.
- Lee, H. B. (1999). Korean. In *Handbook of the International Phonetic Association* (pp. 120–123). Cambridge: Cambridge University Press.
- Levin, J. (1985). *A metrical theory of syllabicity*. Ph.D. dissertation, Massachusetts Institute of Technology.
- Li, F. (2008). *The phonetic development of voiceless sibilant fricatives in English, Japanese and Mandarin Chinese*. Unpublished doctoral dissertation, Ohio State University.
- Li, F., Edwards, J., & Beckman, M. E. (2007). Spectral measures for sibilant fricatives of English, Japanese, and Mandarin Chinese. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the XVIIth International Congress of Phonetic Sciences* (pp. 917–920). Dudweiler, Germany: Pirrot GmbH.
- Lim, S.-j., & Holt, L. (2011). Learning foreign sounds in an alien world: Videogame training improves non-native speech categorization. *Cognitive Science*, 35(7), 1390–1405.
- Lin, H. (2001). *A grammar of Mandarin Chinese*. München: Lincom Europa.
- Lisker, L. (2001). Hearing the Polish sibilants [s ś š]: phonetic and auditory judgments. In N. Grønnum & J. Rischel (Eds.), *Travaux du Cercle Linguistique de Copenhague. To honour Eli Fischer-Jørgensen* (Vol. XXXI, pp. 226–238). Copenhagen: Reitzel.
- Lisker, L., & Abramson, A. S. (1970). The voicing dimensions: Some experiments in comparative phonetics. In *Proceedings of the Sixth International Congress of Phonetic Sciences*. Prague: Academia.
- Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America*, 94, 1242–1255.

- Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: a first report. *Journal of the Acoustical Society of America*, 89(2), 874–886.
- MacWhinney, B. (1987). The Competition Model. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 249–308). Hillsdale, NJ: Lawrence Erlbaum.
- MacWhinney, B. (2008). A unified model. In P. Robinson & N. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 341–371). Mahwah, NJ: Lawrence Erlbaum Associates.
- Maddieson, I. (1980). Palato-alveolar affricates in several languages. *UCLA Working Papers in Phonetics*, 51, 120–126.
- Maye, J., & Gerken, L. (2000). Learning phonemes without minimal pairs. In S. C. Howell, S. A. Fish, & T. Keith-Lucas (Eds.), *Proceedings of the 24th Annual Boston University Conference on Language Development* (pp. 522–533). Somerville, MA: Cascadilla Press.
- Maye, J., & Gerken, L. (2001). Learning phonemes: how far can the input take us? In A. H.-J. Do, L. Domínguez, & A. Johansen (Eds.), *Proceedings of the 25th Annual Boston University Conference on Language Development* (pp. 480–490). Somerville, MA: Cascadilla Press.
- Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: facilitation and feature generalization. *Developmental Science*, 11(1), 122–134.
- Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82, B101–B111.
- McAllister, R., Flege, J. E., & Piske, T. (2002). The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English, and Estonian. *Journal of Phonetics*, 30, 229–258.
- McClaskey, C. L., Pisoni, D. B., & Carrell, T. D. (1983). Transfer of training of a new linguistic contrast in voicing. *Perception and Psychophysics*, 34(4), 323–330.
- McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second-language word learning: minimal instruction produces rapid change. *Nature Neuroscience*, 7(7), 703–704.
- McMurray, B., Aslin, R. N., & Toscano, J. C. (2009). Statistical learning of phonetic categories: insights from a computational approach. *Developmental Science*, 12(3), 369–378.

- Mermelstein, P. (1978). On the relationship between vowel and consonant identification when cued by the same acoustic information. *Perception and Psychophysics*, 23, 331–336.
- Mitchell, R., & Myles, F. (1998). *Second language learning theories*. London: Arnold.
- Miyawaki, K., Strange, W., Verbrugge, R. R., Liberman, A. M., Jenkins, J. J., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception and Psychophysics*, 18, 331–340.
- Morgan-Short, K., Finger, I., Grey, S., & Ullman, M. T. (2012). Second language processing shows increased native-like neural responses after months of no exposure. *PLOS ONE*, 7(3), 1–18.
- Mugitani, R., Pons, F., Fais, L., Werker, J. F., & Amano, S. (2008). Perception of vowel length by Japanese- and English-learning infants. *Developmental Psychology*, 45(1), 236–247.
- Muller, J. (2001). *The phonology and phonetics of word-initial geminates*. Ph.D. dissertation, Ohio State University.
- Narayan, C. R., Werker, J. F., & Beddor, P. S. (2010). The interaction between acoustic salience and language experience in developmental speech perception: evidence from nasal place discrimination. *Developmental Science*, 13(3), 407–420.
- Nation, R., & McLaughlin, B. (1986). Novices and experts: and information processing approach to the good language learner problem. *Applied Psycholinguistics*, 7, 41–56.
- Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, 47(2), 204–238.
- Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, 115(1), 39–57.
- Nowak, P. M. (2006). The role of vowel transitions and frication noise in the perception of Polish sibilants. *Journal of Phonetics*, 34, 139–152.
- Odlin, T. (1989). *Language transfer: cross-linguistic influence in language learning*. Cambridge, UK: Cambridge University Press.
- Pajak, B. (to appear). Non-intervocalic geminates: typology, acoustics, perceptibility. *San Diego Linguistics Papers*.

- Payne, E. (2000). *Consonant gemination in Italian: phonetic evidence for a fortition continuum*. Ph.D. dissertation, University of Cambridge.
- Payne, E. (2005). Phonetic variation in Italian consonant germination. *Journal of the International Phonetic Association*, 35, 153–189.
- Payne, E., & Eftychiou, E. (2006). Prosodic shaping of consonant gemination in Cypriot Greek. *Phonetica*, 63, 175–198.
- Perfors, A., & Dunbar, D. (2010). Phonetic training makes word learning easier. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 1613–1618). Austin, TX: Cognitive Science Society.
- Perfors, A., Tenenbaum, J. B., & Wonnacott, E. (2010). Variability, negative evidence, and the acquisition of verb argument constructions. *Journal of Child Language*, 37, 607–642.
- Pierrehumbert, J. (2000). The phonetic grounding of phonology. *Bulletin de la Communication Parlee*, 5, 7–23.
- 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 stop consonants. *Journal of Experimental Psychology: Human Perception and Performance*, 8, 297–314.
- Pisoni, D. B., & Remez, R. E. (Eds.). (2005). *The Handbook of Speech Perception*. Malden, MA: Blackwell Publishing.
- Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *Journal of the Acoustical Society of America*, 89(6), 2961–2977.
- Polka, L. (1992). Characterizing the influence of native language experience on adult speech perception. *Perception and Psychophysics*, 52(1), 37–52.
- Polka, L., Colantonio, C., & Sundara, M. (2001). A cross-language comparison of /d/ - /ð/ perception: evidence for a new developmental pattern. *Journal of the Acoustical Society of America*, 109(5), 2190–2201.
- Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, 77(3), 353–363.
- Pycha, A. (2009). Lengthened affricates as a test case for the phonetics–phonology interface. *Journal of the International Phonetic Association*, 39(1), 1–31.
- Pyun, K.-S. (2005). A model of interlanguage analysis – the case of Swedish by Korean speakers. In B. Hufeisen & R. J. Fouser (Eds.), *Introductory readings in L3* (pp. 55–70). Tübingen, Germany: Stauffenberg Verlag.

- Ritchie, W. C., & Bhatia, T. K. (Eds.). (2009). *The new handbook of second language acquisition*. Bingley, UK: Emerald Group Publishing Limited.
- Rivers, W. M. (1979). Learning a sixth language: an adult learner's daily diary. *The Canadian Modern Language Review*, 36(1), 67–82.
- Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science*, 12(2), 339–349.
- Sanz, C. (2000). Bilingual education enhances third language acquisition. *Applied Psycholinguistics*, 21, 23–44.
- Schwartz, B., & Sprouse, R. (1994). Word order and nominative case in non-native language acquisition: a longitudinal study of (L1 Turkish) German interlanguage. In T. Hoekstra & B. Schwartz (Eds.), *Language acquisition studies in generative grammar: papers in honor of Kenneth Wexler for the GLOW 1991 workshops* (pp. 317–368). Amsterdam: John Benjamins.
- Selinker, L. (1969). Language transfer. *General Linguistics*, 9, 67–92.
- Selkirk, E. (1991). On the inalterability of geminates. In P. M. Bertinetto, M. Kenstowicz, & M. Loporcaro (Eds.), *Certamen Phonologicum ii: Papers from the Second Annual Cortona Phonology Meeting* (pp. 187–210). Turin: Rosenberg & Sellier.
- Singh, R., & Carroll, S. (1979). L1, L2 and L3. *Indian Journal of Applied Linguistics*, 5, 51–63.
- Sohn, H.-M. (1999). *The Korean language*. Cambridge, UK: Cambridge University Press.
- Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, 388, 381–382.
- Stevens, G. (1999). Age at immigration and second language proficiency among foreignborn adults. *Language in Society*, 28, 555–578.
- Strange, W. (1987). Speech input and the development of speech perception. In J. F. Kavanagh (Ed.), *Otitis media and child development* (pp. 12–26). Parkton, MD: York Press.
- Strange, W. (Ed.). (1995). *Speech perception and linguistic experience: issues in cross-language research*. Timonium, MD: York Press.
- Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception and Psychophysics*, 36(2), 131–145.

- Strange, W., & Shafer, V. L. (2008). Speech perception in second language learners: the re-education of selective perception. In J. G. Hansen Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 153–191). Amsterdam/Philadelphia: John Benjamins.
- Swets, J. A. (1964). *Signal detection and recognition by human observers*. New York: Wiley.
- Swingle, D., & Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science*, 13(5), 480–484.
- Tarnóczy, T. (1988). The formation, analysis and perception of Hungarian affricates. In R. Channon & L. Shockey (Eds.), *In honor of Ilse Lehiste* (pp. 255–270). Dordrecht: Foris.
- Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and bayesian inference. *Behavioral and Brain Sciences*, 24, 629–640.
- Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. (2011). How to grow a mind: statistics, structure, and abstraction. *Science*, 331, 1279–1285.
- Thiessen, E. D. (2007). The effect of distributional information on children's use of phonemic contrasts. *Journal of Memory and Language*, 56, 16–34.
- Thomas, J. (1988). The role played by metalinguistic awareness in second and third language learning. *Journal of Multilingual and Multicultural Development*, 9(3), 235–246.
- Thurgood, E., & Demenko, G. (2003). Phonetic realizations of Polish geminate affricates. In M. J. Sol / 'e, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 1895–1898). Barcelona: Causal Production.
- Towell, R., & Hawkins, R. (1994). *Approaches to second language acquisition*. Clevedon/Philadelphia/Adelaide: Multilingual Matters.
- Tranel, B. (1991). CVC light syllables, geminates and Moraic Theory. *Phonology*, 8, 113–136.
- Tremblay, K., Kraus, N., Carrell, T. D., & McGee, T. (1997). Central auditory system plasticity: generalization to novel stimuli following listening training. *Journal of the Acoustical Society of America*, 102(6), 3762–3773.
- Tréville, M.-C. (1996). Lexical learning and reading in L2 at the beginner level: the advantage of cognates. *Canadian Modern Language Review*, 53, 173–190.
- Trubetzkoy, N. (1939/1969). *Principles of phonology [Grundzüge der Phonologie]*. Berkeley, CA: University of California Press.

- Tseng, C.-Y., Massaro, D. W., & Cohen, M. M. (1986). Lexical tone perception in Mandarin Chinese: Evaluation and integration of acoustic features. In H. S. R. Kao & R. Hoosain (Eds.), *Linguistics, psychology, and the Chinese language* (pp. 91–104). Hong Kong, China: Centre of Asian Studies, University of Hong Kong.
- Tsimpli, I. M. (2003). Clitics and determiners in L2 Greek. In J. M. Liceras, H. Zobl, & H. Goodluck (Eds.), *Proceedings of the 6th Generative Approaches to Second Language Acquisition Conference (GASLA 2002)* (pp. 331–339). Somerville, MA: Cascadilla Press.
- Valencia, J. F., & Cenoz, J. (1992). The role of bilingualism in foreign language acquisition: learning English in the Basque Country. *Journal of Multilingual and Multicultural Development*, 13(5), 433–449.
- Vallabha, G. K., McClelland, J. L., Pons, F., Werker, J. F., & Amano, S. (2007). Unsupervised learning of vowel categories from infant-directed speech. *Proceedings of the National Academy of Sciences*, 104(33), 13273–13278.
- Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America*, 106(6), 3649–3658.
- Werker, J. F. (1989). Becoming a native listener. *Scientific American*, 77(1), 54–59.
- Werker, J. F., Fennell, C. T., Christopher T., Corcoran, K. M., & Stager, C. L. (2002). Infants' ability to learn phonetically similar words: effects of age and vocabulary size. *Infancy*, 3(1), 1–30.
- Werker, J. F., Gilbert, J. H. V., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Development*, 52, 349–355.
- Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception and Psychophysics*, 37, 35–44.
- 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. *Journal of the Acoustical Society of America*, 75, 1866–1878.
- Westbury, J. R., Hashi, M., & Lindstrom, M. J. (1998). Differences among speakers in lingual articulation for American English /ɹ/. *Speech Communication*, 26(3), 203–226.

- Whalen, D. (1989). Vowel and consonant judgements are not independent when cued by the same information. *Perception and Psychophysics*, *46*, 284–292.
- Williams, L. (1977). The perception of stop consonant voicing by Spanish-English bilinguals. *Perception and Psychophysics*, *21*, 289–297.
- Williams, L. (1979). The modification of speech perception and production in second-language learning. *Perception and Psychophysics*, *26*(2), 95–104.
- Williams, S., & Hammarberg, B. (1998). Language switches in L3 production: implications for a polyglot speaking model. *Applied Psycholinguistics*, *7*, 141–156.
- Winn, M., Blodgett, A., Bauman, J., Bowles, A., Charters, L., Rytting, A., & Shamoo, J. (2008). Vietnamese monophthong vowel production by native speakers and American adult learners. In *Proceedings of Acoustics '08* (pp. 6125–6130).
- Wonnacott, E., Newport, E. L., & Tanenhaus, M. K. (2008). Acquiring and processing verb argument structure: distributional learning in a miniature language. *Cognitive Psychology*, *56*, 165–209.
- Wrembel, M. (2010). L2-accented speech in L3 production. *International Journal of Multilingualism*, *7*(1), 75–90.
- Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review*, *114*(2), 245–272.
- Yeung, H. H., & Werker, J. F. (2009). Learning words' sounds before learning how words sound: 9-month-olds use distinct objects as cues to categorize speech information. *Cognition*, *113*, 234–243.
- Yoshida, K. A., Fennell, C. T., Swingley, D., & Werker, J. F. (2009). Fourteen-month-old infants learn similar-sounding words. *Developmental Science*, *12*(3), 412–418.
- Zhang, L. (2011). Vowel length perception in Cantonese. In *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 2292–2295). Hong Kong, China.
- Zhang, L., Kuhl, P. K., Imada, T., Kotani, M., & Tohkura, Y. (2005). Effects of language experience: neural commitment to language-specific auditory patterns. *NeuroImage*, *26*(3), 703–720.
- Żygis, M., & Padgett, J. (2010). A perceptual study of Polish fricatives, and its implications for historical sound change. *Journal of Phonetics*, *38*, 207–226.