

From Features to Words: A Study of Developmental Continuity

by
Jie Ren

B.A. in Linguistics and Literature, Hebei Normal University, 2006
M.A. in Linguistics and Applied Linguistics, Beijing Language and Culture University, 2009
M.Sc. in Cognitive Science, Brown University, 2011
A.M. in Biostatistics, Brown University, 2015

A Dissertation submitted in partial fulfillment of the
requirements for the Degree of Doctor of Philosophy
in the Department of Cognitive, Linguistic and Psychological Sciences at Brown University

Providence, Rhode Island
May, 2016

© Copyright 2016 by Jie Ren

This dissertation by Jie Ren is accepted in its present form
by the Department of Cognitive, Linguistic and Psychological Sciences as satisfying the
dissertation requirement for the degree of Doctor of Philosophy

Date_____

Sheila E. Blumstein, Ph.D., Advisor

Recommended to the Graduate Council

Date_____

James L. Morgan, Ph.D., Reader

Date_____

Linda Polka, Ph.D., Reader

Date_____

Dima Amso, Ph.D., Reader

Approved by the Graduate Council

Date_____

Peter M. Weber, Ph.D., Dean of the Graduate School

Vitae

Jie Ren was born on 3 August 1984 in Guyuan, Hebei Providence, China.

Education

- *Ph.D.* in Cognitive Science, Brown University, Providence, Rhode Island, USA, May 2016.
- *A.M.* in Biostatistics, Brown University, Providence, Rhode Island, USA, May 2015.
- *M.Sc.* in Cognitive Science, Brown University, Providence, Rhode Island, USA, May 2011.
- *M.A.* in Psycholinguistics, Beijing Language and Culture University, Beijing, China, 2009.
- *B.A.* in Linguistics & Literature, Hebei Normal University, Shijiazhuang, China, 2006.

Published Conference Proceedings and Book Chapters

- Jie Ren and James Morgan. “The Devil in the Details: Underspecification in Infants’ and Adults’ Lexical Representations”. 36th Proceedings of Boston University Conference on Language Development, (pp. 500-511), Somerville, MA: Cascadilla Press, 2012.
- Jie Ren and James Morgan. “Developmental Continuity in Early Lexical Representation”. Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS XVII) (pp. 1686-1689), August, Hong Kong, 2011.
- Jie Ren and James Morgan. “Do Infants have MORE detailed Lexical Representation than Adults?” Proceedings of the 47th Chicago Linguistic Society Annual Meeting, Volume 47, number 2, (pp. 185-192). Chicago, Illinois: MetaPress, 2011.

Conference Presentations

- Jie Ren and James Morgan. “Segmental and Suprasegmental Details in Early Lexical Representations”. Paper presented at Boston University Conference on Language Development, 40, November 2015, Boston, MA.
- Jie Ren and James Morgan. “Developmental Continuity in Lexical representations”. Poster presented at the 39th Boston University Conference on Language Development, November 2014, Boston, MA.

- Jie Ren and James Morgan. “Developmental (Dis)Continuity in the Nature of Human Lexical Representations”. Poster presented at XIX Biennial International Conference on Infant Studies, July 2014, Berlin, Germany.
- Jie Ren and James Morgan. “Interpreting Language Universal Constraints with Bayesian Inference”. Paper presented at the proceedings of 84th Annual Meeting of the Linguistic Society of America, January 2014, Minneapolis, MN.
- Jie Ren and James Morgan. “Resolving Asymmetric Findings on Asymmetries in Infant Speech Perception”. The 2013 Society for Research in Child Development Biennial Meeting, April 2013, Seattle, WA.
- Jie Ren and James Morgan. “Segmental and Supra-segmental Details in Early Lexical Representations”. Paper presented at the proceedings of 83rd Annual Meeting of the Linguistic Society of America, January 2013, Boston, MA.
- James Morgan and Jie Ren. “Speech Sound Frequency and Speech Sound Discrimination”. XVIII Biennial International Conference on Infant Studies, June 2012, Minneapolis, MN.
- Jie Ren and James Morgan. “Developmental Continuity in Early Lexical Representation”. Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS XVII), August, 2011, Hong Kong.
- Jie Ren and James Morgan. “Do Infants have MORE detailed Lexical Representation as Adults?” Paper presented at the 47th Chicago Linguistic Society Annual Meeting, April 2011, Chicago, IL.
- Jie Ren and James Morgan. “Infants' Representation of Word-Final Consonants”. The 2011 Society for Research in Child Development Biennial Meeting, March 2011, Montreal, QC, Canada.
- Jie Ren, Liquan Gao and James Morgan. “Mandarin Speakers' Knowledge of the Sonority Sequencing Principle”. The XX Colloquium of Generative Grammar, March 2010, Barcelona, Spain.
- Jie Ren and Liquan Gao. “On Stress Assignment Algorithms of Mandarin Disyllabic Compound Words”. The 7th Pan-Asiatic International Symposium on Languages and Linguistics, December 2008, Guangzhou, China.
- Jie Ren and Liquan Gao. “Syntax is Free from Phonology--The Prosodic-driven PF Movement Hypothesis”. The 12th Symposium on Contemporary Linguistics, October 2008, Wuhan, China.
- Jie Ren and Liquan Gao. “Phonological Grammar is Categorical and Gradient”. The International Symposium of Phonetic Frontiers, April 2008, Beijing, China.

ACKNOWLEDGEMENTS

I would never have been able to finish my dissertation without the guidance of my committee members, help from friends, and support from my family and friends.

First and foremost, I would like to extend my deepest gratitude to my advisor, Dr. Sheila Blumstein, for her excellent guidance. I would never have been able to finish my dissertation without her patient support and care. I would like to thank Dr. James Morgan, who taught me to explore the wonderland of infant language research. I would also like to thank Dr. Dima Amso for all her insightful suggestions on my thesis and consistent encouragement to me on my career development. Special thanks to Dr. Linda Polka, for guiding my research in infant speech perception for the past several years, for traveling internationally to attend my dissertation defense, and for providing excellent feedback on my academic works.

I thank my collaborator Dr. Joseph Austerweil. All my computational modeling work was done in collaboration with him. It was such a great pleasure to work with him on these computational projects and learn from his expertise. I would also like to thank Dr. Uriel Cohen-Priva for all his insightful discussions with me on phonology over the years. I also thank Dr. William Heindel, Dr. Bertram Malle, and Dr. Jack Wright for their encouragement, support, and help on my way to becoming a statistician.

Special thanks to the Open Graduate Education program at Brown University for providing me with the opportunity to receive formal training in statistics. Many thanks to Dr. Christopher Schmid, my director of Graduate Studies in the Biostatistics program at Brown. Thank you for helping me improve my statistics skills.

Special thanks to everyone in the Brown Speech Lab and the Metcalf Infant Research Lab for all their great help over the years.

I have been lucky to be surrounded by many kind and generous friends. I would like to thank Alyssa Feldman, Evon Okidi, Anja Zgodic, Xinmiao Tan, and Matthew Borgia for all the good times we have spent together. Special thanks to my officemate Joanna Korman for her company over the past several years. Last but not least, many thanks to my parents for their nurture and confidence in me. I am so lucky to be born in such a happy family. My special thanks go to my husband, Dr. Deqing Sun, for his lasting love.

TABLE OF CONTENTS

GENERAL INTRODUCTION.....	1
PART I: MAPPING DISTINCTICE FEATURES ONTO WORDS	
Chapter 1: Acquisition of Native Phonetic Categories.....	11
1.1 Acquisition of Native Segmental Categories.....	11
1.2 Acquisition of Native Suprasegmental Categories.....	15
1.2.1 Acquisition of Native Lexical Stress.....	16
1.2.2 Acquisition of Native Pitch Accent.....	20
1.2.3 Acquisition of Native Phoneme Duration.....	21
1.2.4 Acquisition of Native Lexical Tone.....	24
1.3 Chapter Conclusion.....	27
Chapter 2: Segmental and Supra-segmental Detail in Early Lexicon.....	28
2.1 Segmental Detail in Early Lexical Representations.....	28
2.1.1 The Holistic Hypothesis.....	28
2.1.2 The Detailed Hypothesis.....	30
2.1.3 Coda Representations in Early Lexicon.....	32
EXPERIMENT 1.....	33
Methods.....	33
Participants.....	33
Stimuli.....	34
Procedure.....	36
Results and Discussion.....	37
2.2 Suprasegmental Detail in Early Lexical Representations.....	41
2.2.1 Early Lexical Representations of Lexical Stress.....	43
2.2.2 Early Lexical Representations of Lexical Tones.....	44
EXPERIMENT 2.....	45

Methods.....	45
Participants.....	45
Stimuli.....	46
Procedure.....	51
Results and Discussion.....	52
2.3 Discussion of Chapter 2.....	56
 Part II: FEATURE (UNDER)SPECIFICATION IN INFANTS' AND ADULTS'	
LEXICON	
Chapter 3: Underspecification of Distinctive Features.....	60
3.1 The Special Status of Coronal Place of Articulation.....	60
3.2 Underspecification of Distinctive Features in The Lexicon.....	62
3.3 Psychological Predictions of Underspecification.....	63
3.4 Underspecification and Developmental Continuity.....	64
3.5 Underspecification Re-Evaluation	70
3.6 Stop versus Fricative Asymmetry.....	73
3.7 Perceptual Asymmetry in Vowels.....	75
3.8 Chapter Conclusion.....	80
Chapter 4: Underspecification in Toddler's Lexical Representations.....	84
EXPERIMENT 3.....	86
Methods.....	86
Participants.....	86
Stimuli.....	86
Procedure.....	89
Results	90
Discussion.....	100
Chapter 5: Underspecification in Adults' Lexical Representations.....	103

EXPERIMENT 4.....	104
Methods.....	104
Participants.....	104
Stimuli.....	105
Procedure.....	106
Results	107
Discussion.....	114
EXPERIMENT 5.....	116
Methods.....	116
Participants.....	116
Stimuli.....	116
Procedure.....	116
Results	117
Discussion.....	121
Chapter 6: Interpreting Underspecification Using Bayesian Inference.....	128
6.1 The Special Status of Coronal Sounds.....	128
6.2 Interpreting Perceptual Asymmetry Using Bayesian Inference.....	130
6.3 Discussion.....	139
GENERAL CONCLUSION.....	144
REFERENCE.....	154
APPENDICES.....	174
Appendix A. Assignment of stimulus pairs to conditions in Experiment 1.....	174
Appendix B. Results of vocabulary questionnaire for tonal stimulus selection...	175
Appendix C. Assignment of stimulus pairs to conditions in Experiment 2.....	176
Appendix D. Items eligible for Experiment 2 in the 16-month-old CDI of Hao.	177
Appendix E. Assignment of stimulus pairs to conditions in Experiment 3a.....	178

Appendix F. Assignment of stimulus pairs to conditions in Experiment 3b.....	181
Appendix G. JAGS code for Bayesian model comparison.....	184
Appendix H. Assignment of stimulus pairs to conditions in Experiment 4a.....	185
Appendix I. Assignment of stimulus pairs to conditions in Experiment 4b.....	188
Appendix J. JAGS code for Bayesian parameter estimation.....	191
SUPPLIMENTARY MATERIALS.....	192
I. Visual stimulus pairings to in Experiment 1.....	192
II. Visual stimulus pairings to in Experiment 2.....	195
III. Visual stimulus pairings to in Experiment 4a & 5.....	198
IV. Visual stimulus pairings to in Experiment 4b.....	234

LIST OF TABLES

Table 1. Sample stimuli for coda mispronunciations in Experiment 1.....	35
Table 2. Expected feature-based perceptual distance matrix.....	48
Table 3. Sample stimuli for tone mispronunciations in Experiment 2.....	50
Table 4. Sample stimuli for onset mispronunciations.	88
Table 5. Sample stimuli for coda mispronunciations.....	88
Table 6. Priors for the standardized coefficients.	99
Table 7. Priors for the standardized coefficients used in Bayesian model comparisons.....	112
Table 8. Frequencies of segments in CALLHOME transcripts by POA.....	128

LIST OF FIGURES

Figure 1. Sample visual stimulus pair in Experiment 1.....	34
Figure 2. Looking proportion and standard errors, Experiment 1.....	40
Figure 3. Sample visual stimulus pair in Experiment 2.....	51
Figure 4. Looking proportion and standard errors, Experiment 2.....	54
Figure 5. Sample visual stimulus pair in Experiments 3a & 3b.....	87
Figure 6. Experiment 3a: Toddlers' sensitivities to different directions of onset mispronunciation.....	92
Figure 7. Experiment 3b: Toddlers' sensitivities to different directions of coda mispronunciation.....	92
Figure 8. Planned contrasts for toddlers' proportional looking time in Experiment 3a.....	96
Figure 9. Planned contrasts for toddlers' proportional looking time in Experiment 3b.....	97
Figure 10. Sample visual stimulus pair in Experiments 4a & 4b.....	106
Figure 11. Experiment 4a: Adults' sensitivities to different directions of onset mispronunciation.....	107
Figure 12. Experiment 4b: Adults' sensitivities to different directions of coda mispronunciations	108
Figure 13. Planned contrasts for adults' proportional looking time in Experiment 4a.....	111
Figure 14. Planned contrasts for adults' proportional looking time in Experiment 4b.....	112
Figure 15. Experiment 5: Adults' sensitivities to different directions of onset mispronunciations when the baseline phase is removed from the task.....	117
Figure 16. Planned contrasts for adults' proportional looking time in Experiment 5.....	120
Figure 17: Coronal category boundary shift caused by unequal category frequency.....	132
Figure 18: Coronal category boundary stretch caused by unequal category variance.....	136
Figure 19: Coronal category boundary stretch caused by unequal category variance.....	138

ABSTRACT

The goal of this thesis is to investigate the question of developmental continuity in language acquisition. In particular, it aims to address whether representations of native speech categories are continuous between infants' early speech perception, infants' early lexical representations and adults' mature lexical representations.

This question is investigated from two perspectives in the current thesis:

a) continuity between infants' speech perception and infants' early lexical representations and b) continuity between infancy to adulthood in the architecture of distinctive feature representations in speech and the lexicon.

By the end of the first year of life, infants have already acquired a majority of their native speech categories. However, existing theories that differentiate between phonetic representations and phonological representations consider speech category learning a disconnected process from lexical representations. According to this view, while infants can successfully differentiate /b/ and /d/ for example, in a speech perception task, they would not be able to do so in word recognition tasks until they know that /**d**uck/ and /**b**uck/ refer to two different animals. In this way, infants' early lexical representations should not be as detailed as adults' mature lexical representations until infants have a large vocabulary size. Such considerations thus predict that infants' early lexical representations should be less specified in terms of distinctive features than mature lexical representations.

Using psycholinguistic experiments and computational modeling, this thesis has challenged hypotheses of developmental **dis**continuity. The results indicate that infants' early lexical representations include very detailed and even adult-like information in terms of distinctive features, suggesting substantial developmental continuity between early and mature lexical representations. Moreover, experimental results have indicated that developmental continuity exists both in the *amount* and the *nature* of feature specifications in infants' and adults' lexicon. Finally, the computational model suggests that from infancy to adulthood, the architecture of distinctive feature representations is also continuous in speech perception.

GENERAL INTRODUCTION

Human development poses an interesting paradox. While it is often characterized by *change*, continuity is also fundamental for the maintenance of knowledge and skills behind the apparent changes over time. How is continuity achieved in the context of developmental change? This question leads directly to the problem of developmental continuity, a central and long-standing issue in developmental science.

The question of developmental continuity asks “Where do new skills come from?” Proponents of developmental continuity claim that new skills grow from the seeds of prior accomplishments. Therefore, to examine whether developmental continuity exists in a certain psychological domain, people usually address the following two problems: First, do old skills (or knowledge) that are acquired earlier map onto new skills (or knowledge) that are acquired later in development to accomplish the same goals? Second, do different stages of development share any similarity in the mechanism used for information processing? This thesis is designed to address these two problems in the domain of language acquisition.

Language provides a unique tool for understanding human developmental continuity. Not only because it is the one skill that defines us as human but also that it starts much earlier in life compared with many other cognitive skills. From the moment we are born (or even before that), we have language in our lives; and this skill is maintained lifelong for most people. Most importantly, language is also a very sophisticated skill. As the PRIMIR model of language acquisition suggests (Werker & Curtin, 2005), “...there is rich information available in the speech input and that the system picks up and organizes this information along a number of different dimensions”.

These dimensions “are not “given” at birth. Rather, they are the joint product of the biases the infant brings, the regularities in the input, and statistic learning”. Yet children achieve such goals easily with minimum effort in a daunting speed. Thus, understanding developmental (dis)continuity in the domain of language acquisition will significantly better our understanding of how human learning occurs.

In order to investigate developmental continuity in language acquisition, this thesis focuses on infants’ and adults’ representations of *distinctive features*. Following Jakobson (1941), distinctive features are defined as the phonological components that are used to *differentiate between phonemes in any world language*. For example, a consonant phoneme can be described using three distinctive features: place of articulation, voicing and manner of articulation. For example, the phoneme /b/ is [+labial, +voiced, +stop] and the phoneme /k/ is [+velar, -voiced, +stop]. /b/ and /k/ thus differ by two distinctive features: place of articulation and voicing. To address the question of developmental continuity in distinctive feature representations, this thesis will investigate the following two specific problems.

First, does acquisition of native phonemes as indicated by distinctive feature representations (earlier skill) map onto word learning (later skill)? This question asks whether there is continuity in small steps of development.

Second, do infants and adults use similar architecture to process distinctive features in word recognition? This question asks whether there is continuity in the long term of development.

Continuity in small steps: Mapping distinctive features onto words

For several decades, the question of developmental continuity between the acquisition of native phonemes, which usually happens during the first year of life, and word learning, which usually happens during the second year of life, has been an enduring source of debate in the literature of language acquisition.

As will be shown in Chapter 1, at the beginning of life, infants exhibit a similar pattern of phonetic perception regardless whether they are from the same or different language learning environment. They discern differences between the phonetic units of many different languages, including the target language that they are learning and languages that they have never heard before (e.g. Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Trehub, 1976; Werker, Gilbert, Humphrey & Tees, 1981). For example, in contrast to English speaking adults, 6-month-old English-learning infants are able to differentiate between the Hindi retroflex contrast [ɖ]-[tʌ] though this contrast is not distinctive in English (Werker & Tees, 1984). Similarly, 4-month-old English-learning infants are able to differentiate between different Cantonese tones (Yeung, Chen & Werker, 2013) though tonal contrast is not a distinctive dimension in English. These findings suggest that infants are born with the universal sensitivity to a global set of speech contrasts and that such sensitivity prepares them for learning any native language(s) in the world (see Kuhl, 2004 for review).

By the end of the first year of life, infants' speech perception of phonetic categories has been significantly altered by native linguistic experience (Werker & Tees, 1984; Kuhl, Williams, Lacerda, Stevens & Lindblom, 1992). By around 10-months of age, English-learning infants can no longer differentiate between the Hindi retroflex contrast [ɖ]-[tʌ]; and by around 9-month of age, English-learning infants can no longer

differentiate between different Cantonese tones (Yeung, Chen & Werker, 2013).

However, sensitivities to native speech contrasts are maintained. This learning process is referred to as *perceptual narrowing*.

It is our natural assumption that acquisition of native phonemes (i.e. native phonetic categories) serves as a basis for word learning. For example, knowing that /b/ and /d/ are two different phonetic categories in English, infants should be able to differentiate between *ball* and *doll* as referring to two different objects (Fennell & Werker, 2003). Thus, there should be developmental continuity between phoneme learning and word learning. However, as will be seen in Chapter 2, studies do not always suggest that this is the case. Existing theories that differentiate between phonetic representations in speech and phonological representations in words consider phoneme learning and word learning two disconnected processes. In particular, unlike what developmental continuity predicts, some studies have shown that distinctive features represented at the lexical level may be different from the phonetic details represented at a purely speech level. For example, 14-month-old English-learning infants were able to differentiate between *bih* and *dih* only when these two sounds were NOT used to refer to meanings. When *bih* and *dih* were presented in a referential context as novel word labels, they failed to differentiate between the two sounds (Stager & Werker, 1997). Therefore, as suggested by many earlier studies (e.g. Charles-Luce & Luce, 1990, 1995; Storkel, 2002; Pisoni, Lively, & Logan, 1994; Walley, 1993, 2005), it is possible that two words are distinctive from each other at the lexical level only when they are used to refer to different objects. Thus, before knowing that *buck* is the name of another animal, for example, infants might represent the word *buck* simply as an exemplar of the word *duck*.

Therefore, unlike perceiving differences between phonetic categories in speech perception, phonetic detail in early lexicon should be related with the number of different words whose phonological representations are contrastive. Given the fact that infants have a very small vocabulary size, early lexical representations are not likely to have adult-like sophistication in terms of distinctive features. Based on this view, phonetic categories acquired in the first year of life do not seem related to lexical learning, and there should not be developmental continuity between early acquisition of native phonemes and early lexical representations.

Chapters 1 and 2 will discuss the mapping between *distinctive features (old skill)* maps and *words (new skill)* in more detail. In particular, more recent findings will be reviewed that directly examine the relationship between the acquisition of native distinctive features and infants' early representations of familiar words. Also, two experiments will be presented that were conducted to address this question in further depth. As will be seen in these chapters, unlike findings in the earlier mentioned studies, the two experiments together with more recent findings in the literature suggest that acquisition of native phonemes does map directly onto word learning, and that there is developmental continuity between early speech and early lexical representations.

Long term Continuity: Similarity in infants' and adults' representations of distinctive features

Besides the relationship between early phoneme learning and early word learning, another critical question regarding developmental continuity in language acquisition is whether infants and adults represent distinctive features using the same mechanism. To

address this question, this thesis will be focusing on place of articulation features in stop consonants, with a particular focus on the feature of [coronal] (sounds articulated with the tongue tip or blade).

Phonological theories, have proposed that the coronal place of articulation has a special status in the lexicon. Unlike other distinctive features, it is not specified in the lexicon (Clements, 1987a; Steriade, 1987; Kiparsky, 1982; Archangeli, 1984; Archangeli, 1988; Pulleyblank, 1986). Such a claim predicts an asymmetry in word recognition. For example, /d/, as a coronal sound, is unspecified (or unmarked) for place of articulation in the lexicon; thus, words that contain coronal stops, /d/, may be perceived as equivalent to words with other places of articulation, e.g. /b/ or /g/. / . In contrast, /b/, as a labial sound, is fully specified in the lexicon, and hence it is not likely to be misperceived as other places of articulation. As a consequence, the word *buck* may be perceived as an exemplar of *duck* but *duck* cannot be an exemplar of *buck*. According to nativists' point of view, the special status of the coronal place of articulation is determined by universal grammar (UG), and is thus independent of experience with a specific language (see Lahiri & Reetz, 2010).

Indeed, such asymmetries have been shown in adults. For example, at temporally early stages of speech perception as measured by Event-Related Potentials (ERP), German-speaking adults display asymmetric discrimination for mispronunciations of familiar words with coronal or noncoronal onsets (Friedrich, Lahiri, & Eulitz, 2008), word internal consonants (Friedrich, Eulitz, & Lahiri, 2006; Cornell, Lahiri & Eulitz, 2013), codas (Lahiri & van Coillie, 1999), and vowels (Lahiri & Reetz, 2010; Cornell, Lahiri, & Eulitz, 2011). Similar effects have also been found with English-speaking

adults (Roberts, Wetterlin & Lahiri, 2013). Other studies using gating and lexical decision procedures also have demonstrated putative effects of underspecification in Bengali, English, and German (Gaskell & Marslen-Wilson, 1996; Lahiri & Marslen-Wilson, 1991; Lahiri & Reetz, 2010; Wheeldon & Wasker, 2004).

These findings on adults' asymmetrical perception of words containing coronal and noncoronal features naturally lead to the question of developmental continuity: Do infants also represent coronal and noncoronal features asymmetrically in the early lexicon? In other words, is there continuity in the mechanism used for lexical processing over development?

Studies on pre-lexical infants did find evidence for early appearance of asymmetries as predicted by underspecification. Tsuji, Mazuka, Cristia & Fikkert (2015) tested 4-month-old Dutch-learning infants and Japanese-learning infants on their discrimination of [ɔmpa] (labial) and [ɔnta] (coronal) in a habituation task. They found that for both language groups, 4-month-olds were able to discriminate [ɔmpa] and [ɔnta] only when they are habituated with [ɔmpa]. Therefore, [ɔnta] cannot be [ɔmpa] but [ɔmpa] can be [ɔnta]. Similarly, Dijkstra & Fikkert (2011) tested 6-month-old habituated Dutch-learning infants to either repeated *taan* or *paan* tokens and then tested their ability to discriminate these stimuli using trials in which one or the other stimuli was repeated or trials in which the two stimuli alternated. Whereas infants habituated to *paan* discriminated the two types of trials, infants habituated to *taan* did not. Both studies interpreted their findings in terms of underspecification: when the standard of comparison was a coronal feature, place of articulation of the consonants was unspecified, so that both coronal and noncoronal features were compatible with the standard, but when the

standard of comparison was noncoronal, place of articulation of the onset was specified, so that only noncoronal was compatible with the standard, and coronal and noncoronal features were discriminable. However, although the presence of an asymmetry in pre-lexical infants is not inconsistent with the predictions of the underspecification hypothesis, these studies focused on phonetic not lexical discrimination. Without direct evidence on whether infants' early lexical processing is also asymmetrical, it is unclear in what way these results with pre-lexical infants can be translated into the *lexical* level representations as assumed in underspecification.

To date, previous empirical research on word learning has rarely considered possible asymmetrical processing in early lexical representations, but has assumed that all distinctive features, no matter coronal or noncoronal, are equally specified in the early lexicon. Thus, it remains unknown whether underspecification as a mechanism underlying lexical access holds true over development.

This dissertation will address this question in Chapters 3 through 6. By examining whether different directions of mispronunciations (i.e. coronal to noncoronal and noncoronal to coronal) asymmetrically affect word recognition for infants and adults, as predicted by underspecification, this dissertation will contribute a critical piece of evidence to our understanding of whether the special status of coronal place of articulation is reflected continuously from infancy to adulthood. More importantly, using the coronal distinctive feature as leverage, the findings from these chapters will significantly better our understanding as to whether there is developmental continuity in the architecture of lexical processing mechanisms from infancy to adulthood.

This dissertation is organized as follows: Chapter 1 reviews previous empirical work on pre-lexical infants' acquisition of native phonemes. Chapter 2 presents two experiments on toddlers' representations of segmental and supra-segmental information. Chapter 3 discusses the special status of the feature [coronal] and its psycholinguistic implications. Chapters 4 and 5 present a series of experiments examining toddlers' and adults' sensitivities to different directions of mispronunciations in familiar word recognition tasks, respectively. Chapter 6 presents a computational model using Bayesian inference to provide an alternative explanation of underspecification of distinctive features. Chapter 7 summarizes this dissertation by discussing its implications and limitations.

**PART I: CONTINUITY IN SMALL STEPS:
MAPPING DISTINCTICE FEATURES ONTO WORDS**

CHAPTER 1

ACQUISITION OF NATIVE PHONETIC CATEGORIES

This chapter reviews work on phonetic category acquisition. Two types of evidence are considered in this domain. The first type of evidence comes from experimental findings that show that infants become more sensitive to native segmental categories, and the second type of evidence comes from experiments in which infants are shown to be more sensitive to native supra-segmental categories over development. Although native segmental category learning and supra-segmental phonetic category learning have typically been considered separately, this chapter compares the acquisition of lexical segmental and supra-segmental categories in infancy, so as to provide a comprehensive overview to the question of developmental continuity.

1.1 Acquisition of native segmental categories

One of the first tasks infants face during native speech acquisition is to learn which sounds are distinctive in their native language. Studies examining perceptual changes relating to sound category learning have demonstrated that learning native phonetic contrasts begins quite early, during the second half of the first year of life.

Very young infants initially show the ability to discriminate a set of universal contrasts that occur in many different languages (e.g. Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Trehub, 1976; Werker & Tees, 1984; see Saffran, Werker & Werner (2006) for a review). For example, Morse (1972) found that American English-learning infants at the age of 40 to 54 days could differentiate [ba] and [ga] in synthesized speech. Trehub (1976) tested French-speaking adults and 7-15 week old French-learning infants in Montreal on their discrimination of a French vowel contrast in [pa]-[pã], and a Czech

consonant contrast in [za]-[řa]. The infants were found to be able to discriminate both contrasts, regardless of whether these contrasts are present in their native language or not. Conversely, the adults were only able to discriminate [pa]-[pã] contrast but were unable to discriminate [za] and [řa] contrast.

Between six and twelve months, discrimination of consonants that are contrastive in foreign languages, but not in the native language, declines. Initial evidence for this developmental trajectory for native consonants primarily came from studies by Werker and colleagues (Werker, Gilbert, Humphrey & Tee, 1981; Werker & Tee, 1983; Werker & Tee, 1984; Werker & Lalonde, 1988). In a series of studies, Werker et al. tested English adults, English infants and Hindi adults on their ability to discriminate the dental, retroflex speech contrasts [t'a]-[d'a] and glottalized velar and glottalized uvular sounds in Thompson ([k']-[q']). In both the cross-sectional studies and the longitudinal studies, they found that English infants aged 6-8 months could discriminate these sounds in similar ways that native adult speakers of Hindi or Thompson do (Werker et. al., 1981; Werker & Tee, 1984), suggesting that between 6-8 months, infants still maintain their ability to discriminate nonnative contrasts regardless of whether they are distinctive in their native languages. However, 10-12-month old English-learning infants, like English-speaking adults, failed to discriminate these nonnative contrasts, suggesting that between 8 and 10 months, infants lost their sensitivities to speech contrasts that are not distinctive in their native languages. Werker and colleagues further found that infants being raised to speak Hindi or Thompson sounds could still discriminate the contrasts at 10-12 months (Werker & Tee, 1984), suggesting that infants' capacity of discrimination to *native* speech contrasts is maintained over development.

A similar developmental trajectory occurs to vowel perception. Available evidence suggests that developmental change begin even earlier for vowels than for consonants. Kuhl, Williams, Lacerda, Stevens & Lindblom (1992), for example, tested both American English-learning and Swedish-learning infants on their perception of English vowel /i/ and Swedish vowel /y/. Using the head turning preference procedure (HPP), they found that 6-month-old infants exhibited a strong magnet effect only for native-language phonetic prototypes, suggesting that language specificity in vowel perception in English and Swedish learning infants started as early as six months of age. Bosch & Sebastian-Galles (2003) tested Spanish-learning infants on their ability to discriminate vowel contrast [e] and [ɛ], which are distinctive in Catalan but not in Spanish. Their results indicated that Spanish-learning infants' discrimination between these two vowels declines between the ages of four and eight months. Also, using a habituation paradigm, Polka & Werker (1994) suggested that English-learning infants lose sensitivity to a backness contrast in German high rounded vowels [u] and [y] by six to eight months.

Although the decline in the discrimination between nonnative segmental categories is often found to be true, not all contrasts fit this developmental profile. For example, some nonnative segmental contrasts remain highly discriminable across development. Best, McRoberts and Sithole (1988), for example, found that 12–14 month-old English-learning infants as well as English-speaking adults could still discriminate non-native place of articulation contrasts in clicks (sounds produced by releasing a pocket of air trapped by the tongue in the mouth, found in some languages of southern Africa). They argued that the characteristics of click contrasts were sufficiently different from

English contrasts so that they were discriminable by older infants even in the absence of native-language experience. Certain non-native vowel contrasts also remain discriminable across development. Polka & Bohn (1996), for example, found that the German front-back high vowel contrast /y/-/u/ remained discriminable across English-learning infants' development. This finding failed to replicate Polka and Werker's (1994) demonstration of language-specific reorganization for these vowels by 6–8 months. Polka and Bohn (1996) suggested that the contrast remained discriminable for English-learning infants for acoustic reasons: the formant differences distinguishing the /y/-/u/ contrast in their study were substantially larger than the formant differences observed in the stimuli used by Polka and Werker (1994).

Besides, some contrasts are not discriminated by young infants and appear to require months of language experience before they are discriminated by native listeners. For example, at 10–12 months but not at 6–8 months did Filipino-learning infants, whose native language includes the syllable-initial alveolar /n/ versus velar /ŋ/ contrast, show discrimination of native [na]–[ŋa] (Narayan, Werker & Beddor, 2010). Similarly, by the age of 12 months, experience with English improves English-learning infants' discrimination of the native /r/–/l/ contrast (Kuhl, Stevens, Hayashi, Deguchi, Kiritani & Iverson, 2006; also see Sebastian-Galles, 2006, for discussion). Similarly, by the end of the first year of life, experience with Mandarin improves Mandarin-learning infants' discrimination of the native affricate-fricative /tʃ/-/tʃ̥/ contrast (Tsao, Liu & Kuhl, 2006). Language experience facilitates discrimination of the /d/-/ð/ contrast into early childhood and adulthood for English-learning children (Polka, Colantonio & Sundara, 2001; Sundara, Polka & Genesee, 2006). In all of these studies demonstrating facilitation (Polka

et al., 2001; Kuhl et al., 2006; and Tsao et al., 2006), the discrimination performance of young infants on the native contrast was good (significantly above chance), but not as good as the performance of older infants whose results approached the performance of adults on the same contrasts. Therefore, infants' discrimination between some native segmental contrasts is not always inborn, but can be learned and enhanced over development as the infant receives more language exposure.

Although the precise timing and extent of this perceptual narrowing process depend on factors specific to the segments in question, it is clear that infants begin with considerable ability to discriminate language-general phonetic contrasts in early infancy, and then rapidly begin to develop a language-specific perceptual sensitivity which gets refined further across development (see Holt & Lotto, 2008).

1.2 Acquisition of native supra-segmental categories

Studies reviewed in the previous section have built a rich understanding of how the acquisition trajectories of native phonetic categories unfold over development. However, their findings remain limited in the sense that phonetic categories under investigation in these studies were almost exclusively phonetic *segments* (i.e. vowels and consonants). Much less is known about how infants learn to perceive native *supra-segmental* categories, such as lexical stress, duration, pitch accent and lexical tones. While the use of most of these supra-segmental categories is relatively rare in Indo-European languages (though some languages do make use of lexical stress), these supra-segmental categories play essential roles in differentiating meanings in most other language families in the world, such as East Asian languages, African languages and some Polynesian languages. Moreover, it has been argued that infants respond first to

supra-segmental patterns in the speech around them (Lewis, 1951; Crystal, 1970), perhaps even before they were born (Fifer and Moon, 1988), that their babbling acquires adult-like intonation patterns before the appearance of true lexical items (Kaplan & Kaplan, 1971), and that the first word may be defined supra-segmentally rather than segmentally (Engel, 1973). Therefore, knowing of the developmental trajectories of native supra-segmental categories is of significant importance for us to see the nature of human language acquisition. For the purpose of the current dissertation, findings in how infants acquire native lexical supra-segmental categories will significantly facilitate our understanding of developmental continuity in language acquisition.

1.2.1 Acquisition of Native Lexical Stress

As in the case of consonant and vowels, though very young infants can differentiate between different stress patterns, their preferences for native-language-specific patterns of lexical stress do not emerge until after 5 months of age in the first year of life.

Young infants demonstrate delicate sensitivities to the acoustic differences between words carrying different lexical stress patterns. In a classic study, Spring & Dale (1977) tested 2-month-olds on their discrimination of two disyllabic words that only differed by their stress patterns (babà vs. bàba). Using the habituation procedure by applying the high-amplitude sucking paradigm (HAS), they found that all infants tested were dis-habituated when they heard the word carrying a novel stress pattern, suggesting that by 2 months of age, infants already have the capacity of discriminating between different lexical stress patterns. This result was replicated by Jusczyk & Thompson (1978) for disyllabic words badà vs. bàda in which the stressed syllables in the two words

also differed in consonants. Sansavini, Bertoncini & Goivanelli (1997) replicated the findings of both Spring & Dale (1997) and Jusczyk & Thompson (1978) with 2-day-old Italian-learning newborns. Using a similar design as in Spring & Dale, the authors found that newborn subjects successfully discriminated disyllabic (mamà vs. màma or dagà vs. dàga) and trisyllabic (tàcala vs. tacàla) words that differed in their stress patterns.

Although very young infants can *discriminate* between different kinds of lexical stress patterns in speech perception tasks, they do not show *preference* towards the predominant stress patterns of their native languages until sometime in the second half year of life. At this time, infants are able to make use of such preferences in continuous speech segmentation. In a pioneering study, Jusczyk, Cutler, & Redanz (1993) found that 9- but not 6-month-old English learning infants listened significantly longer to words with trochaic stress patterns, suggesting that such native language preference may develop sometime between 6- to 9-month-old of age. Such findings were replicated by Echols, Crowhurst & Childers (1997), in which the author found that 9- but not 7-month-old English-learning infants prefer trochaic stress patterns. Adults showed similar preference as 9-month-olds, suggesting that representations of native predominant stress patterns are preserved to adulthood.

Moreover, Echols, Crowhurst & Childers (1997) also found that 9-month-olds used the trochaic stress pattern in lexical level segmentation tasks. Subjects distinguished a trochaic sequence that had previously been embedded in a four-syllable string from a novel trochaic sequence, suggesting that they recognized the previously heard trochaic unit as familiar. No such discrimination was present for iambic targets, suggesting that infants failed to recognize the previously heard iambic sequence. Therefore, it is likely

that by 9 month of age, infants have already started to store lexical stress in their memory as an important supra-segmental element of lexical representations. Curtin, Mintz, & Christiansen (2005) further investigated this hypothesis and tested 7-month-old English learning infants. In one of the experiments, they first familiarized the infant subjects with a continuous speech that was comprised of trisyllabic novel words carrying a strong-weak-weak stress pattern (e.g. DObita). Then, they tested infants on real English sentences that contained a target word either from the familiarization stream (e.g. I like your DObita), or a control item that differed in stress pattern from the target word (e.g. I like your DOBITA). The findings suggested that infants preferred the target items over the control items. Therefore, infants' representation of stress information at the lexical level can start as early as seven months of age.

Recent studies using more advanced technology or simpler experimental design suggested that infants' acquisition of the predominant stress patterns in their native language can start from four months of age for German-learning infants and French-learning infants (Herold, B., Höhle, B., Walch, E., Weber, T., & Obladen, M. , 2008). Although the timing for infants to develop native stress bias might be language specific and can vary due to the rhythmic properties of the language in acquisition (Höhle, Bijeljac-Babic, Herold, Weissenborn, & Nazzi, 2009), it is consistent that before the end of the first year of life, infants already have good command of the predominant patterns of lexical stress in their native languages (Skoruppa et al., 2009; Pons & Bosch, 2010; Skoruppa, Cristia, Peperkamp, & Seidl, 2011, Segal & Kishon-Rabin, 2012) and that they make use of such knowledge to facilitate native phonetic category learning (Mattys, et. al., 1999), speech segmentation (Polka & Sundara, 2012) and even grammar learning (Shi

& Werker, 2001).

Therefore, demonstrable developmental continuity has been suggested between early speech learning and later language acquisition. It thus naturally follows that knowledge of native lexical stress patterns may serve as the foundation for lexical learning, as hypothesized back in 1970s by Crystal (1970), Engel (1973) and Wieman (1976). In particular, lexical stress, as one of the most important supra-segmental lexical cues, should guide the representation of native lexical structure (Sansavini, 1994) because infants appear to encode stress information as part of their proto-lexical representations (Curtin, Mintz & Christiansen, 2005).

However, recent studies have found mixed evidence for the phonetic details encoded in infants' early lexical representations in terms of lexical stress. On one hand, 20- to 24-month-old English-learning children could differentiate minimal pairs of newly learned words only when the minimal pairs occur in stressed syllables (Flocchia, Nazzi, Austin, Arreckx & Goslin, 2011), meaning that infants at such an age do encode stress information in lexical processing. On the other hand, adults can easily recognize familiar words (e.g. bunny, banana) in a visual world online word recognition task, when stress was either indicated normally with convergent cues (pitch, duration, amplitude, and vowel quality) or was manipulated such that only pitch was different. However, 2.5-to-5-year-old English-learning children could only be successful in the normal stress condition. This indicates possible developmental change progressing toward the adulthood, which starts later, well past infancy. Therefore, it is still an open question whether and when infants represent lexical stress in the lexicon using the same mechanism as adults do, though it is clear that in purely speech perception task infants' sensitivities to lexical

stress patterns is adult-like by the end of the first year of life. It is thus open to know whether there is developmental continuity between early speech learning, early lexical learning and mature lexical representations in terms of lexical stress representations.

1.2.2 Acquisition of Native Pitch Accent

Another lexical supra-segmental element is pitch accent. Although very few studies have investigated the acquisition trajectory of pitch accent, existing findings have suggested that lexical contrasts of pitch accent develop on a similar schedule as most vowel and consonant contrasts.

Using the high-amplitude sucking procedure, Nazzi, Floccia, & Bertoncini (1998) tested newborns French-learning infants with two lists of phonetically varied Japanese words that differed in pitch-accent contours. They found that these infants were able to discriminate the pitch-accent difference between HL and LH (Nazzi, Floccia, & Bertoncini, 1998), even though pitch-accent is not used as a lexical supra-segmental element in French.

Sato, Sogabe, & Mazuka (2010a) tested 4- and 10-month-old Japanese-learning infants on their sensitivities to lexical pitch-accent changes in disyllabic words. Their behavioral results indicated that both age groups were able to detect the pitch accent changes. However, results from fNIR measures suggested that brain responses to the pitch change within words differed between 4- and 10-month-old infants in terms of functional lateralization, with left hemisphere dominance for the perception of the pitch change embedded in words only seen in the 10-month-olds. These results suggest that there is a developmental change in Japanese lexical pitch–accent perception between four and ten months of age, as reflected by the switch in brain functional lateralization from

bilateral to left hemisphere dominance. Given the findings by Sato, Sogabe, & Mazuka (2007) that left language-related regions contribute to the processing of lexical pitch-accent in native Japanese adult speakers, we can conclude that Japanese-learning infants have developed adult-like lexical representations of pitch accent by 10 months of age.

To investigate whether infants make use of their knowledge of native pitch-accent to facilitate early word learning in the second year of life, future studies need to look at whether changes in pitch-accent impair infants' word recognition. Japanese adults were found to make use of lexical pitch-accent information to constrain their word activation in spoken-word recognition tasks (Cutler & Otake, 1999). If similar patterns are found for infant learners in word recognition tasks, such findings will significantly better our understanding whether there is developmental continuity between early speech, early lexical representations and mature lexical representations in terms of pitch-accent.

1.2.3 Acquisition of Native Phoneme Duration

The previous sections indicated that infants begin to develop native sensitivities to lexical stress and pitch-accent on a similar schedule as the development of sensitivities to most native vowel and consonant contrasts. Nevertheless, as suggested by Yeung, Chen & Werker (2013), vowels and consonants occur within individual syllables, while lexical stress and pitch-accent are defined across multiple syllables. Such difference would make direct comparison between lexical segments and supra-segment difficult and less meaningful. To make a more closely matched comparison, the segments and supra-segments under investigation should be defined along similar time scales (i.e., within syllables). Native phoneme duration is a possible supra-segment that can provide such direct comparison, because variations in phoneme duration only occur on a single

phoneme.

Duration is used in many languages to distinguish words (e.g., Japanese, Arabic, Dutch, & Berber). As in the case of lexical stress, very young infants can also differentiate between words that only differ by vowel durations. In the second experiment of Spring & Dale (1977), the authors found that 2-month-old infant subjects could successfully differentiate between disyllabic words that only differed in vowel duration.

However, recent studies that investigated the acquisition trajectories of native phoneme duration have shown very different developmental patterns than vowels and consonants or lexical stress and pitch-accent. For example, using the visual habituation procedure, Sato, Sogabe, & Mazuka, (2010b) tested Japanese-learning infants' abilities to discriminate Japanese long and short vowels embedded in two-syllable words (/mana/ vs. /ma:na/). The results revealed that neither the 4-month-old nor the 7.5-month-old Japanese infants were able to discriminate the contrast, whereas 9.5-month-olds showed the discrimination ability. When 4-month-olds were tested on their discrimination between contrasts introduced by phonemic changes (/mana/ vs. /mina/), they showed successful discrimination. Such findings indicate that Japanese infants do not acquire native sensitivity to long-short vowel contrasts until sometime between 7.5 and 9.5 months of age and that the developmental course of the phonemic category by the durational changes is different from that by the phonemic change. Similar findings were replicated in a follow-up study (Sato, Kato, & Mazuka, 2012). Using fNIRs combined with a habituation task, Minagawa-Kawai, Mori, Naoi, & Kojima (2007) tested Japanese learning infants on their discrimination of phonemic duration contrasts (e.g. /mama/ vs. /mama:/). Their findings suggested a U-shaped developmental trajectory. While 6-to-7-

month-old infants could differentiate cross-category changes of phoneme duration, such discrimination was found to be lost at 10-to-11 months but resumed and maintained after 13 months of age. Moreover, even by 18-month-old of age, Japanese-learning infants could not discriminate vowel duration in an adult-like fashion. Using the switch task, Mugitani et. al. (2009) for example, tested 18-month-old Japanese-learning infants on their discrimination of paired nonsense words that contained a short or a long vowel, in which the contrast between the words was solely the length of this vowel (i.e., /taku/ vs. /ta:ku/). They found that Japanese-learning infants at this age detected only the change from long to short vowel, but not the change in the opposite direction. In a different experiment, the authors found that native adult Japanese speakers, unlike Japanese 18-month-old infants who are presumably still developing phonological perception, symmetrically discriminated both ways of vowel changes (i.e. the change from long to short vowel and the change from short to long vowel).

Therefore, both behavioral and brain-imaging results suggested a markedly protracted developmental trajectory for duration categories, which may not be acquired by going through the regular curve of perceptual narrowing.

As in the case of pitch accent, there are few existing studies investigating whether infants make use of their knowledge of native phonemic duration to facilitate early word learning in the second year of life. As far as I am concerned, the only study in the literature that investigated how phonemic duration affects early lexical processing is Dietrich, Swingley, & Werker (2007). Using the “switch task”, the author tested Dutch- and English-learning 18-month-olds’ on their sensitivities to newly learned words that varied in vowel duration or vowel quality. They found that Dutch learners interpreted

vowel duration as lexically contrastive, but English learners did not, in keeping with properties of Dutch and English, suggesting that half way through the second year of life, infants can use native phoneme duration information in lexical processing. However, it is uncertain whether such timing applies to learners of other languages.

To further explore the question of developmental continuity between early speech and early lexical representations, future studies may at least need to 1) test Dutch-learning infants' early acquisition of native phonemic duration cues at a purely phonetic level; and 2) test Japanese-learning infants on their sensitivities to phonemic duration in lexical processing tasks.

1.2.4 Acquisition of Native Lexical Tone

Studies reviewed in the previous parts have shown how acquisition of some native supra-segments unfolds over development. However, due to their cross-syllabic nature, lexical stress and pitch accent cannot be good representatives of native supra-segmental categories for establishing a direct comparison with native segmental categories. Neither is phonemic duration because the developmental trajectory of which is still quite elusive. Therefore, to make such a direct comparison possible, lexical tone is probably the best candidate.

Lexical tones play fundamental roles in differentiating meanings in a great proportion of languages, and are used by the largest population in the world. Therefore, knowing the developmental trajectories of native lexical tones has important implications for us to see the nature of human language acquisition. Given the fact that it is currently the optimal candidate for us for making comparisons between supra-segmental categories and segmental categories, findings in how infants acquire native lexical tones will

significantly facilitate a comprehensive understanding of developmental continuity in language acquisition from both segmental and supra-segmental perspectives.

Tones are primarily defined by F_0 variations within single syllables, and are found in most native American languages, Sub-Saharan Africa languages, as well as East and Southeast Asian languages (Yip, 2002). For example, Mandarin has four lexical tones. The same syllable carrying different types of tones differs in meaning, for example, “ma¹”-*mother*, “ma²”- *hemp*, “ma³”-*horse*, and “ma⁴”-*scold*. (Duanmu, 2007). Thus, tone constitutes an important factor in lexical retrieval for native Mandarin-speaking adults (Cutler & Chen, 1997).

Very few reports (reviewed below) have examined the perception of tones in infancy. Among those few studies, however, it has been consistent that tone acquisition also follows a similar pattern of perceptual narrowing, but it occurs on an earlier timetable than that of consonant or vowel contrasts.

English-learning infants at four months of age can differentiate between different Tai (Mattock, Molnar, Polka, & Burnham, 2008) or Cantonese lexical tones (Yeung, Chen, & Werker, 2013, Experiment 1) in the same way as age-matching Cantonese-learning infants do. However, such discrimination ability decreased by 9 months of age (Mattock, 2006; Mattock, Molnar, Polka, & Burnham, 2008). Like English-learning infants, Cantonese-learning infants showed discrimination to Cantonese lexical tones at 4 months of age. However, unlike English-learning infants, by 9 months, they have maintained their discrimination of native lexical tonal contrasts (Yeung, Chen, & Werker, 2013, Experiment 2). Yorùbá infants of 6 months were more attentive to Yorùbá tones than were English infants (Harrison, 2000), revealing early native enhancement in lexical

tone perception. Such findings have suggested a typical perceptual narrowing process for lexical tone acquisition: learners whose native languages are non-tonal lose their sensitivities to tonal contrasts, while learners whose native language is tonal maintain their sensitivities to native tonal contrasts over development.

Also, like consonant and vowel perception, studies have suggested that learners of both tonal and non-tonal languages rely heavily on the acoustic distances between different lexical tones in tasks of tonal discrimination. For example, 8-to-15-month-old Dutch-learning infants are less sensitive to Mandarin tonal contrasts when the F_0 values between the tones are closer to each other (Liu & Kager, 2014). For tonal language learners, Shi, Gao & Li (2010) suggested that 7-to-11-month-old, but not 6-month-old, Mandarin-learning infants were able to discriminate the acoustically similar Mandarin tonal contrast Tone 2 vs. Tone 3. Moreover, in a tone perception task, using the conditioned head-turn procedure, Tsao (2006) found that 12-month-old Taiwan Mandarin-learning infants were sensitive to different levels of acoustic similarities among the four Taiwan Mandarin tones: infants in this study were better in discriminating a tonal contrast with greater acoustic distance (e.g. Tone 1 vs. Tone 3) than a tonal contrast with closer acoustic distance (e.g. Tone 2 vs. Tone 3).

Given the similarity between the developmental trajectory of lexical tones and consonant and vowels, lexical tone may provide a good test base for the question of developmental continuity between infants' speech perception and word recognition; that is, like consonants and vowels, perceptual reorganization of lexical supra-segments during early infancy is also to serve the function of word learning. This question will be further explored in Chapter 2.

1.3 Chapter Conclusion

This Chapter reviewed studies on how infants acquire native phonetic categories during the first year of life, including the acquisition of native segmental and supra-segmental categories. The precise timing of acquisition highly depends on factors specific to the segmental and supra-segmental categories in question. Nevertheless, findings of the studies reviewed in these sections suggest that there is a common trajectory in development underlying the variety of experience-related profiles: most of the time, infants begin with successful discrimination of language-general phonetic contrasts in early infancy, which then develops into language-specific perceptual sensitivity. This step of language acquisition serves a foundation for the next step of language acquisition: word learning. Importantly, it provides a means for testing the first question of developmental continuity: does acquisition of native speech categories (earlier skill) map onto word learning (later skill)? The answer to this question will address whether there is continuity in small steps of development.

CHAPTER 2

SEGMENTAL AND SUPRASEGMENTAL DETAIL IN EARLY LEXICON

The previous chapter reviewed findings that indicated that by the end of the first year of life, infants' speech perception is tuned by language-specific phonetic structures (e.g., Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Werker, Gilbert, Humphrey, & Tees, 1981; Werker & Tees, 1984; Werker & Lalonde, 1988). Therefore, according to these studies, by 12 months, infants have aligned their perception with many of the phonetic categories of their native languages.

It has been assumed that learning native phonetic categories in infancy lays a good foundation for the following steps of language acquisition (see Morgan & Demuth, 2014 for a review). However, in recent decades, whether there is such a relation of developmental continuity between phonetic learning and word learning has been a central debate in the field of language acquisition. The current chapter emphasizes research that concerns the relationship between the two learning processes, that is: Do infants make use of their knowledge of native speech categories in facilitating word learning?

In each of the following sections, I first review findings in the literature that investigated this research question. Then, I introduce my experiment that was designed to further address this question. Two kinds of native speech categories were considered in this Chapter: Segmental categories and supra-segmental categories.

2.1 Segmental detail in early lexical representations

2.1.1 The holistic hypothesis

Findings in the 1990s often suggested that infants' early lexical representation should not be credited with too much richness in terms of phonological detail (Hallé & de

Boysson-Bardies, 1996; Charles-Luce & Luce, 1990, 1995; Storkel, 2002; Walley, 1993, 2005). For example, using the switch procedure, Stager & Werker (1997) found that 14-month-old English-learning infants could differentiate between the minimal pair *bih* and *dih* only in speech perception tasks where these two novel sounds are not used to label objects. When *bih* and *dih* were used to label novel objects, infants at this age failed to differentiate between the two sounds. Therefore, it seems that that phonetic category learning achieved during the first year of life might not be directly related to phonological contrasts in the lexicon. In particular, learning the phonetic categories of a language during the first year of life is purely a speech learning process that is based on infants' ability to distinguish a universal set of sounds in early infancy.

However, learning the phonological system of a language is closely tied to learning vocabularies: Two sounds are phonologically contrastive in a language only when they help to differentiate words (e.g. Charles-Luce & Luce, 1990, 1995; Storkel, 2002; Pisoni, Lively, & Logan, 1994; Walley, 1993, 2005). As a result, in a small lexicon, little detail needs to be represented. A strong version of this holistic view predicts that children should only be sensitive to very large mispronunciations, perhaps involving the deletion or alteration of multiple segments (see Hallé & de Boysson-Bardies (1996) for data that are consistent with this prediction). Thus, early development in speech-sound categorization need not imply that young toddlers represent words in memory as a set of adult-like, detailed, and sophisticated phonological units. In this vein, children's early representations of words are holistic (not formed from discrete sound categories) and vague (missing some of the detailed information that is necessary for differentiating words) (Metsala & Walley, 1998), and thus the specificity with which

individual words are represented is increased as a consequence of the development of vocabulary size. Therefore, through word learning, the lexicon representation becomes more and more detailed.

2.1.2 The detailed hypothesis

An alternative point of view is that when infants begin to learn new words, their early phonetic categories are highly detailed. This view is based on the rationale that young infants are already able to use their analytical abilities to perform a distributional analysis of the speech signal, categorize native sounds, and differentiate between linguistic and non-linguistic dimensions, before word learning begins. When facing the task of word learning, as suggested in the PRIMIR model of language acquisition, once infants have established representations of native speech categories (phonemes), it will guide them in information pick-up and act as a beacon in directing attention to just that information that is criterial in distinguishing phonetically similar words such as “bih” and “dih.” (see Werker & Curtin (2005), page 220).

In recent years, studies using more sensitive measures have shown increasingly supportive evidence for this point of view (e.g. Bailey & Plunkett, 2002; Ballem & Plunkett, 2005; Swingley & Aslin, 2000, 2002). Swingley & Aslin (2000), for instance, tested 18-23-month-old English-learning infants on their sensitivities of familiar word mispronunciations using an intermodal preferential looking procedure (IPLP). In each trial, the authors presented the infants with two horizontally opposed familiar objects on the screen (Target and Distractor, e.g. a baby and a car). Then the audio stimuli were presented asking the infant to find the target object. Labels of familiar object were either correctly pronounced or mispronounced. The mispronunciations involved replacement of

one segment with a similar segment (e.g. **b**aby is mispronounced as **v**aby). They found that although children recognized the spoken words in the correct (e.g. **b**aby) and the mispronunciation (e.g. **v**aby) conditions, recognition was significantly poorer when words were mispronounced (e.g. **v**aby). Similar results were replicated when 14-month-old babies were recognizing highly familiar words (Swingly & Aslin, 2002) and newly learnt words (Ballem & Plunkett, 2005).

Findings by White & Morgan (2008) provided even stronger evidence on this view that early lexical representations are detailed. The authors conducted a set of studies on the sensitivity of 19-month-old infants to the severity of familiar word onset mispronunciation. Unlike the IPLP procedure used by Swingly & Aslin (2002), White & Morgan (2008) paired a familiar target object with an *unfamiliar* distractor object in each set of displays to increase the sensitivity the task. In this way, if mispronunciations are effective in impairing word recognition, they can be referred to as the labels of the unfamiliar objects displayed on the screen. Their findings suggested that 19-month-old infants display linearly graded sensitivity to word onset mispronunciations deviating by one (e.g. ball-gall), two (e.g. ball-kall), or three features (e.g. ball-shall) from known forms, directing their attention to the familiar object less and less as the severity of the mispronunciation increases. Given that adults are affected by the degree of acoustic–phonetic mismatch in word recognition and semantic priming tasks (Connine, Titone, Deelman, & Blasko, 1997; Milberg, Blumstein, & Dworetzky, 1988), the authors concluded that at least with respect to word consonant onsets, infants’ early word representations are detailed, with adult-like sophistication in terms of phonetic features.

In a similar task, Mani & Plunkett (2011) replicated the graded sensitivity with 2-year-old British English-learning toddlers on their perception of vowel mispronunciations.

2.1.3 Coda representations in early lexicon

In spite of these previous findings arguing against the holistic hypothesis, most previous research has been largely limited to the study of onset consonants, and a few have examined vowels with monolingual (Mani & Plunkett, 2007, 2011) and bilingual infants (Ramon-Casas et al., 2009). Thus far, word codas have been rarely investigated. Without evidence for word codas, there are a variety of good reasons to suspect that infants' early lexical knowledge is still less complete.

For one thing, robust evidence in the literature of word recognition has shown that both adults and infants have a strong asymmetry in their perception of word onset and coda representations. In English for example, segments in the coda position tolerate more variability than in the onset position, rendering changes in the coda segment less easy to detect. For example, Levelt (2009) habituated 14- and 18-month-old Dutch infants to CVC syllables like [pak]. When they tested the infants with syllables like [pa], both age groups of infants failed to detect the coda deletion. The author suggested that codas are not firmly stored in the lexicon.

Also, listeners can begin “activating” a cohort of words only based on word-initial phonetic information in online word recognition (Marslen-Wilson, 1975; Dahan & Tanenhaus, 2005). Therefore, it is possible that this information can override, delay, or weaken the effects of later-occurring phonetic information, such as that provided by word codas. In this way, infants may devote more attention to the beginning parts of words than the ending parts so as to make word recognition more efficient and less effortful.

Such an onset-oriented strategy may in turn make word coda learning more difficult for infants. Therefore, it is very likely that early lexical representations of codas are less mature than their representations of onsets.

Swingley (2009) showed that 14- to 21-month-old infants are able to detect one-feature (place or voice) mispronunciations of codas of familiar words. Nazzi and Bertoncini (2009) found that 20-month-old infants are able to learn pairs of words differing in coda consonant voicing. However, it is still not known whether infants represent and judge codas holistically in a match or mismatch manner, or in a more sophisticated feature-graded fashion as White & Morgan found for onsets.

To address this question, the following experiment, Experiment 1, was designed by replicating White & Morgan's procedure on coda mispronunciations.

EXPERIMENT 1

This experiment was designed to examine whether 19-month-olds exhibit feature-graded sensitivities to different severities of mispronunciations of familiar word codas. Toddlers were tested using an Intermodal Preferential Looking procedure in which looking time to referents of familiar words versus objects without known labels was measured.

Methods

Participants

Forty-six 19-month-olds were recruited from the Providence, Rhode Island area using public birth records. Seventeen subjects were excluded from the analyses due to fussiness or crying (7), not attentive to the stimuli (4), equipment failure (2), parental

interference (1), over 10% alternative language exposure (3). This left 29 subjects, 15 females and 14 males, mean age = 585.625 days.

Stimuli

Familiar stimuli comprised a set of words that are understood by the majority (>50%) of infants by 14 months, according to the MacArthur CDI norms ((Dale & Fenson, 1996). Thus, on average, the majority of our 19-month-olds were likely to have had exposure to the familiar words for well over five months prior to the experimental session.

In each trial, infants saw two images, one depicting a referent of a familiar word, the other depicting a referent of an unknown (to 19-month-olds) word. Both the familiar and unfamiliar objects selected for the study were real objects, similar in visual complexity and drawn from the same set of categories (e.g. artifacts, living things, body parts, etc.). An example stimulus pair is depicted in Figure 1. Pairings of familiar and unfamiliar objects remained constant across subjects.

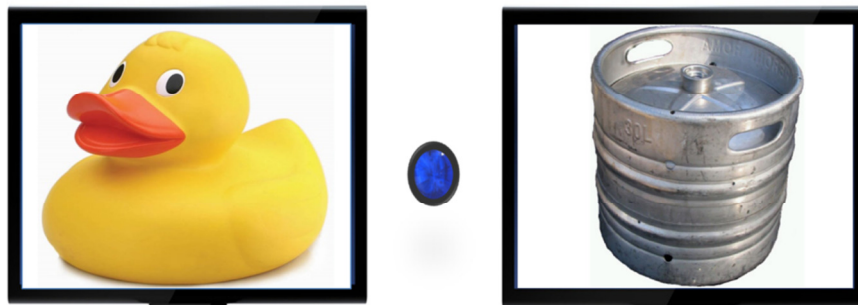


Figure 1. Sample visual stimulus pair in Experiment 1.

Infants were tested in 18 trials. Each subject received three trials in which the familiar object's label was pronounced correctly (e.g. duck), three trials in which the familiar object's label was pronounced with a one-feature change (place) in the coda consonant (e.g. dut), three trials in which the familiar object's label was pronounced with

a two-feature change (place + voicing) in the coda consonant (e.g. dud), three trials in which the familiar object's label was pronounced with a three-feature change(place + voicing + manner) in the coda consonant (e.g. duz), two familiar fillers in which the familiar object was named (e.g. hand) and four novel fillers in which the unfamiliar object was named (e.g. keg). All mispronunciations resulted in non-words or in words judged unlikely to be familiar to toddlers at this age. Examples of mispronunciations are given in Table 1. Novel labels were phonologically distinct from familiar labels. Our inclusion of an unfamiliar object was motivated by the hypothesis that this would be a more sensitive context for evaluating toddlers' detailed lexical knowledge. It was, therefore, important to indicate to them that the unfamiliar object was a possible referent and in turn to reduce the possibility that subjects would adopt a strategy of always looking at the familiar object. More importantly, the presence of an unfamiliar object in each trial provided a baseline from which we can measure looking behavior in the context of a completely unfamiliar word. All the words tested in the current study were mono-syllabic.

Table 1

Sample stimuli for coda mispronunciations in Experiment 1

Audio Stimuli Samples	(Mis)pronunciation	Features
<i>Where's the duck?</i>	Correct	NA
<i>Where's the dut?</i>	1-feature	place
<i>Where's the dud?</i>	2-feature	place and voicing
<i>Where's the duz?</i>	3-feature	place, voicing and manner

As in White & Morgan (2008), each item was presented only once during the session to increase the sensitivity of the paradigm.

A complete list of stimuli and conditions is provided in *Appendix A*.

All stimuli were naturally produced by a trained female speaker of American English who produced the utterances in infant-directed speech with positive affect. The mean length of target items across all conditions was 588.05 ms (SD = 78.58 ms); lengths of target items did not differ significantly across the four pronunciation conditions, $F(3, 92) = 0.036$, $p = 0.991$. While pairings of objects remained constant, the assignment of these stimulus pairs to pronunciation condition was counterbalanced across subjects (see *Appendix A*), with the exception of six trials (two correct fillers and four novels) that were constant across subjects.

Procedure

Testing was conducted in a sound-treated laboratory room. Parents sat with children on their laps and listened to instrumental music over noise-cancelation headphones to mask the audio stimuli. Approximately 90cm in front of the child were two 51cm television monitors mounted side-by-side, together subtending approximately 55 degrees of visual angle. A speaker was located centrally between the two television monitors behind a pegboard panel. At the subjects' eye level, a blue light was mounted on the panel between the two television monitors. The subjects were monitored over a closed-circuit video system and recorded on a digital camcorder at 30 fps for later off-line coding. Speech stimuli were played at conversation level (70 dB).

We adopted the intermodal preferential looking procedure (IPLP) used in White & Morgan (2008). Each trial began with the blue light flashing until the subject fixated at midline. At that point, the experimenter turned off the center light and initiated the salience phase. During the salience phase, one object with a known label and a second

object with unknown label were simultaneously displayed on two screens (see Figure 1). The two objects were displayed silently for 4 seconds to establish baseline looking preference. After the salience phase, the two monitors went dark, and infants' attention was then recaptured to midline to avoid contingencies between side of fixation between the end of the salience period and the beginning of the test period.

After recentralization, the experimenter initiated the test phase. During the test phase, the audio stimulus (*Where's the X?*) was played. Immediately after the offset of the target word, the two visual stimuli were presented. Three seconds after the audio stimuli, a second audio stimulus (*Find the X!*) was presented to maintain the infant's interest. The test phase lasted 9 seconds in total. Following an interval of at least 1 second, the next trial commenced. The side of presentation of the familiar object was randomized between trials by the customized experimental software, but was consistent across salience and test phases within each trial.

The dependent measure was the change in proportional looking to the familiar object between the (silent) salience phase and the test phase. Of interest was whether looking behavior would differ as a function of the severity of mispronunciation. The order of presentation was pseudo-random with the constraint that the first two trials always contained one correct filler and one novel trial. Following the experimental session, the parent completed a questionnaire on his/her toddler's comprehension and production of the stimulus items.

Results and discussion

Results from the parental questionnaire indicated that the items we selected were appropriate for this sample of 19-month-olds. On a scale of 1 (neither say nor understand)

2 (not say, not sure understand or not), 3 (not sure whether say or understand), 4 (not say, understand), 5 (not sure say or not, understand), 6 (both say and understand) items used as familiar words received an average score of 5.20 (SD = .15), indicating that they were very familiar to this group of toddlers. Items used as unfamiliar words received an average score of 1.24 (SD = .23).

Looking behavior was coded off-line frame-by-frame (1 frame = 33 ms) using in-house coding software. For the salience phase, looking behavior was coded for the entire 4 second duration of the phase. For the test phase, looking behavior was coded only for the first 3 seconds following the offset of the first occurrence of the target word. This was done in order to include only initial response to the target word, before the onset of the second audio stimulus. For each phase, the proportion of looking towards each of the objects was computed over the total time the subject spent looking at both objects for that phase. For each subject, a difference score was computed for each condition. This difference score measured the change in looking toward the familiar object after the target was named, and was computed using the following formula, where averages are computed over trials in a single condition:

$$\% \text{Looking(Familiar)}_{\text{Test}} - \% \text{Looking(Familiar)}_{\text{Salience}}$$

Comparison across test and salience phases allowed us to use each stimulus pair as its own control, thereby controlling for any inherent preference for a particular stimulus in each pairing. Subjects exhibited a general preference for the familiar object in the salience phase. The average looking proportion for familiar objects across all conditions was .57(SD=0.268). A single-sample t-test comparing mean salience proportions (across conditions) to 0.5 was significant ($t(20) = 2.774$, $p = 0.012$). This

finding replicated Schafer, Plunkett and Harris (1999) and White & Morgan (2008). Both studies showed that toddlers (17-month-olds and 19-month-olds) prefer to look at familiar objects even in the absence of audio input. Because of such biased preference in the baseline, trials in which subjects did not look at both objects during the salience phase were not included in the analysis. Across all subjects, 25 trials were discarded for this reason (9.02% of trials). One more subject was discarded because he/she did not have response scores left in the correct condition due to trial deletion for this reason. In addition, subjects who had more than two large negative ($<-.15$) difference scores in the correct condition were not included in the analysis, an inclusion criterion also used by White & Morgan (2008). Across all correct trials, seven subjects were discarded for this reason. This criterion is based on the requirement of the current experiment that children should have familiarity with the referents of the words. Subjects with low negative scores in the correct condition did not fit this assumption, because they were either not engaged in the task or unfamiliar with the words. In this way, data from 21 subjects were used in the analyses. Scores of difference in proportional looking to target are depicted in *Figure 2*.

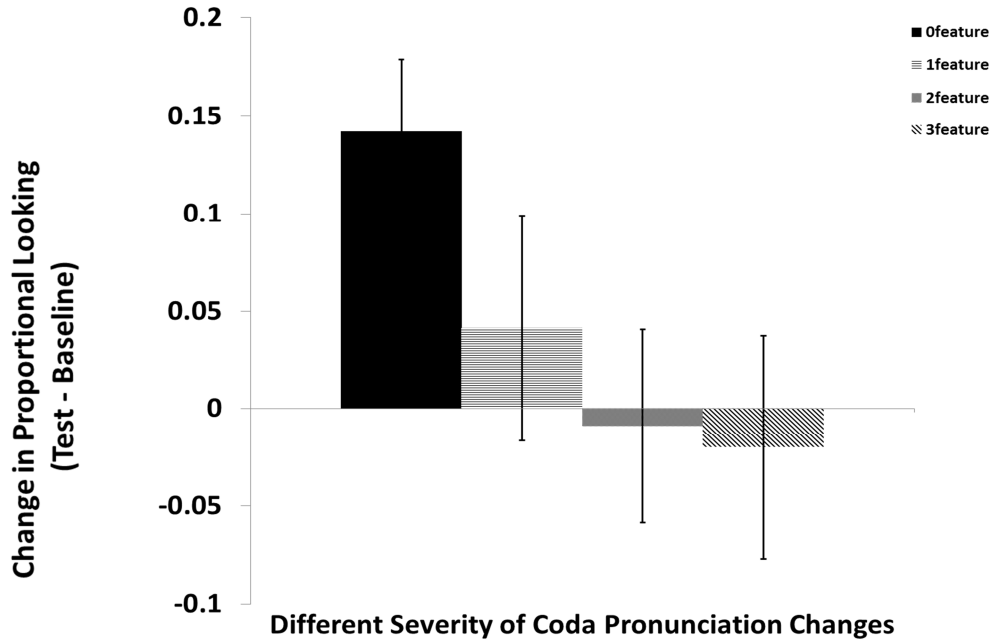


Figure 2. Looking proportion and standard errors, Experiment 1.

To establish whether there was an effect of mispronunciation on responses in general, one-way MANOVAs across the 4 pronunciation conditions (excluding novel trials) were conducted using repeated measures (Wilk's Lamada). Different patterns of looking across the conditions were significant ($F(3, 18) = 4.811, p = 0.012$, partial $\eta^2 = 0.445$). Most importantly, to explore whether toddlers have graded sensitivity to various degrees of mispronunciations, a within-subject trend analysis on the four pronunciation conditions (excluding novel trials) was performed. A striking linear trend in difference scores was found ($F(1, 20) = 9.461, p = 0.006$, partial $\eta^2 = 0.321$), while the quadratic and cubic trends were not significant and both negligible. Thus, toddlers were significantly affected by the severity of mispronunciations. Because preferences in the salience phase were similar across conditions ($F(3, 18) = 0.517, p = 0.676$, partial $\eta^2 = 0.079$), differences in looking behavior can be attributed to effects of the audio stimulus at test.

These results replicate White & Morgan's findings for onsets. Nineteen-month-

old toddlers have graded sensitivity to different severities of familiar word coda mispronunciations, even though codas are more variable and perceptually less salient than onsets. Therefore, together with White & Morgan's finding for onsets and Mani & Plunkett's finding for vowels, we can conclude that early lexical representations of native **segments** are specified with great amount of phonetic detail.

The next section of this chapter investigates whether early lexical representations of **supra-segments** have a similar amount of phonetic detail.

2.2 Supra-segmental detail in early lexical representations

As reviewed in Chapter 1, supra-segments play an equally important role in word recognition. For example, lexical stress is used to differentiate words in Persian languages; pitch accent is used in Japanese and lexical tone is used in Chinese-Tibetan languages. Despite their importance, lexical supra-segments have been paid much less attention in the developmental literature than lexical segments. This is perhaps due to the fact that lexical supra-segments are perceptually less salient than lexical segments.

In the case of lexical stress, Cutler (1986) found that minimal stress pairs with no segmental difference (such as TRUS^{ty} or trus^{TEE}; FOR^{bear} or for^{BEAR}) are effectively homophonous in the earliest stages of word recognition, suggesting that the suprasegmental distinction plays no role at this stage. Fear, Cutler & Butterfield (1995) found that vowel quality differences (***a**udience vs. **a**ddition*) are perceptually more important than stress differences (*audience vs. auditorium*). Furthermore, in studies of

stress misplacement, it has been consistently found that mispronunciation of lexical stress has no adverse effect on word recognition in noise (Small, Simon, & Goldberg, 1988), unless vowel quality is also changed (Cutler & Clifton, 1985; Koster & Cutler, 1997).

Similarly, tonal cues have also been found much less prominent than segmental cues. Tong, Francis & Gandour (2008), for example, asked Mandarin-speaking listeners to classify words based on their shared dimensions in consonants (**ba**²-**bu**²), vowels (**ba**²-**da**²) and tones (**ba**²-**bu**²) respectively, while ignoring the variation along the irrelevant dimensions. Their findings suggested that the tone dimension was more likely to be ignored than segmental dimensions. Cutler & Chen (1997) reported a similar processing disadvantage of lexical tones with Cantonese-speaking listeners: Subjects accepted a nonword as a word more often when the only difference between the nonword and the word was in tone. Same-different judgments by these listeners were also slower and less accurate when the only difference between two syllables was in tone. Ye & Connine (1999) found that Mandarin-speaking participants had faster responses for vowel mismatch stimuli than for tone mismatch stimuli. Similarly, in two priming experiments, Sereno & Lee (2014) found that for Mandarin-speaking participants, priming effects only occurred when both tonal and segmental information matched or when only segmental information overlapped. No priming occurred for the tone-only overlap condition.

Taking together these findings, it is evident that supra-segmental units are not always as reliably informative for adult listeners as segmental units in word recognition. While such difference in saliency between segmental and supra-segmental information may be the reason that lexical supra-segments are less noticeable by researchers¹, it also

¹ Or the reason might be that mostly that the field has been dominated by English speaking researchers. Thank Dr. Linda Polka for suggesting such a possibility.

presents strong reasons for us to suspect that early lexical representations of tones are less complete.

However, as reviewed in Chapter 1, infants who are acquiring tone languages have aligned their perception with most of the native lexical supra-segments by the end of the first year of life. Therefore, if there is developmental continuity between early speech and early lexical representations, as suggested in section 2.1, and if the developmental continuity equally applies to lexical supra-segments, then we would expect that supra-segments are also represented with sophisticated phonetic detail in early lexicon.

Unlike studies on segmental representations, however, little research in the literature has directly investigated the developmental continuity for lexical supra-segments. However, among the rare investigations that tested supra-segments at the lexical level with infants, it has been suggested that early lexical representations of supra-segments include very sophisticated phonetic detail.

2.2.1 Early lexical representations of lexical stress

Early word recognition is influenced by stress starting from 11-months. Changes in the stress pattern of a familiar word do not affect word recognition with 11-month-olds, but it does slow down processing (Vihman et al., 2004). Mismatching the stress pattern of Sw words affects recognition of familiar words by 3-year-olds, but mismatching the stress pattern of wS words does not (De Bree, Van Alphen, Fikkert, & Wijnen, 2008). These findings suggest that although lexical stress is not the strongest factor that affects word recognition, infants are still encoding the stress information and this influences familiar word recognition.

Recent studies have found that infants are more sensitive to lexical stress information in novel word recognition tasks. For example, changing the stress pattern of a newly encountered word affects 12-month-olds' word recognition (Curtin, 2010b), and 12-month-olds can learn two new word-object pairings even when the novel words differ only in their stress patterns (e.g., *XYZ* vs. *xYZ*) (Curtin, 2009). Using the Switch procedure (Werker et al., 1998), Curtin (2010) found that 14-month-old English-learning infants were able to learn two novel words and pair them with two novel objects when the two novel words only differ in their stress patterns (e.g. *BEdoka* vs. *beDOka*). Taken together, studies with familiar and novel words seem to indicate that stress information is encoded in detail in very early lexical representations.

2.2.2 Early lexical representations of lexical tones

As indicated in Chapter 1, a direct comparison between lexical stress and vowels and consonants is difficult, because the former is defined across multiple syllables while the latter occurs within individual syllables. Therefore, a comparison between segments and supra-segments defined along similar timescales (i.e., within syllables) would provide a more closely matched comparison. Lexical tone in Asian languages is such a good supra-segmental candidate given that it is only imposed on a single segment (usually the vowel) in a syllable. However, there is only one study so far which directly tested infants' sensitivities to native lexical tones in tasks related to word recognition: Tao & Xu (2013), using the IPLP paradigm as in Swingley & Aslin (2002), tested 12-month-old Mandarin-learning infants on their sensitivities to tonal mispronunciations of familiar words. Their findings replicated Swingley & Aslin (2002) for consonants, indicating that at the end of the first year of life, Mandarin-learning infants are sensitive

to lexical tonal mispronunciations when the correct tone is changed into an incorrect tone.

To obtain a clearer picture on the amount of phonetic detail early lexical representations include in terms of supra-segmental information, and to better compare segmental and supra-segmental details in early lexicon, we need to know whether like in White & Morgan (2008), infants are also sensitive to different severities of tonal mispronunciations. The following experiment investigates this question.

By using the same paradigm and design as in Experiment 1, it examines whether by the middle of the second year of life, infants process segmental and supra-segmental components in similar mechanism in familiar word recognition.

EXPERIMENT 2

This experiment is intended to examine whether by the middle of the second year of life, infants are sensitive to different severities of mispronunciations of native lexical tones in on-line word recognition.

Methods

Participants

Twenty-one monolingual Mandarin-learning toddlers (mean age = 637.3 days) were tested in Beijing, China. All babies were raised in families where Mandarin Chinese has been the only language spoken in multiple generations. Toddlers whose families had dialect- or foreign-language-speaking members were not recruited for the current study. Five subjects were excluded from the analyses due to fussiness or crying (2) and equipment errors (3). This left 16 subjects distributed across the four counterbalancing groups (11 females and 5 males, mean age = 632.7 days).

Stimuli

As discussed in Chapter 1, in Mandarin Chinese lexical tones are linguistically distinctive in the same way as consonants and vowels. Mandarin has four lexical tones. The same syllable carrying different types of tones differs in meaning, for example, “ma¹”-*mother*, “ma²”-*hemp*, “ma³”-*horse*, and “ma⁴”-*scold*. (Duanmu, 2007).

Phonetically, the four lexical tones can be described as high level [55] (Tone 1), high rising [35] (Tone 2), low falling rising [214] (Tone 3), and high falling [51] (Tone 4), respectively (Chao, 1968; Howie, 1976). The numbers brackets represent the relative levels of the pitch, with five levels in total: 1 as the lowest and 5 as the highest. Perceptual data on Chinese tones indicate that variation in F₀ values yields high levels of recognition for isolated tones (Howie, 1976; Tseng, 1990). Other acoustic cues including amplitude (Whalen & Xu, 1992) and duration (Fu, Zeng, Shannon, & Soli, 1998; Xu, Tsai, & Pfingst, 2002) also influence tonal perception.

In continuous speech, Tones 3 and 4 are subject to sandhi in certain contexts (Yip, 2002; Chen, 2000; Duanmu, 2007). Tone 3 is most susceptible to alteration. If Tone 3 precedes Tone 1, Tone 2 or Tone 4, it will not be fully produced but changed into a deviant form of [21] (e.g. *yi*³_[214]*zi* → *yi*_[21]*zi*, “chair”). If Tone 3 precedes another instance of Tone 3, it will be changed to Tone 2, i.e. the rising tone [35] (e.g. *xiao*³_[214]*gou*³_[214] → *xiao*²_[35]*gou*³_[214], “dog”) (see Chen, Liu & Kager (2015) for perceptual evidence of tone sandhi). In addition, when Tone 4 is followed by another instance of Tone 4, it will be pronounced as [53] instead of [51] (e.g. *dian*⁴_[51]*shi*⁴ → *dian*_[53]*shi*⁴, “TV”).

Phonologically, the four lexical tones are represented with high (H) and low (L) features: high level /HH/ (Tone 1), high rising /LH/ (Tone 2), low falling rising /L(H)/ (Tone 3), and high falling /HL/ (Tone 4), respectively (Duanmu, 2007, p.237). The (H) of the low falling rising tone /L(H)/ (Tone 3), can be interpreted as a floating tone that follows the syllable (Milliken, 1989), and may be realized if the syllable is alone or if a neutral tone syllable follows². The expected feature-based perceptual distance matrix among the four lexical tones is shown in Table 2. The numbers in Table 2 indicate distance from the correct pronunciation, with 0 indicating no distance, 1 close distance, 2 farther distance and 3 farthest distances.

For example, if Tone 2 is mispronounced as Tone 3 (i.e. LH is changed to L(H)), it involves either one feature change if L(H) surfaces as L or no feature change if L(H) surfaces as LH. Therefore, phonologically, L(H) is the closest mispronunciation for Tone 2, coded as 1. If Tone 2 is mispronounced as Tone 1 (i.e. LH is changed to HH), it always involves one feature change, i.e. the starting feature. Therefore, compared with Tone 3, Tone 1 is a farther mispronunciation of Tone 2, coded as 2. If Tone 2 is mispronounced as Tone 4, (i.e. LH is mispronounced as HL), it always involves two feature changes, i.e. the starting feature and the ending feature. Therefore, compared with Tone 3 and Tone 1, Tone 4 is the farthest mispronunciation of Tone 2, coded as 3 in Table 2. The same rationale of coding applies for other mispronunciation changes in Table 2.

² We thank Dr. Larry Hyman for his insightful suggestions on using this phonological scheme in analyzing Mandarin tonal features.

Table 2. Expected feature-based perceptual distance matrix³.

		Mispronunciation Changes			
		T1 /HH/	T2 /LH/	T3 /L(H)/	T4 /HL/
Correct Tones	T1 /HH/	0 none	1 start	2 start + end (or start)	1 end
	T2 /LH/	2 start	0 none	1 end (or none)	3 start + end
	T3 /L(H)/	2 start + end (or start)	1 end (or none)	0 none	3 start + end
	T4 /HL/	1 end	2 start + end	2 start + end	0 none

Note: 0=no mispro; 1=close mispro; 2=farther mispro; 3=farthest mispro

As shown in Table 2, only Tone 2 and Tone 3 could create mispronunciations that were increasingly distant from the correct form. However, Tone 3 is not a suitable candidate for our mispronunciation manipulations because it is subject to tone sandhi in almost all tonal contexts as we noted previously. Therefore, in this experiment I used only words with Tone 2 bearing stems as correct pronunciation forms.

To select an appropriate set of familiar words and unfamiliar words for the experiment; I searched the Chinese Contemporary Dictionary (Lv & Ding, 2005, 4th Edition) word by word and selected words with picturable referents that were possibly known by 19-month-olds as well as a few picturable words that were likely unknown by 19-month-olds. Based on these words, I designed a questionnaire which was administered in a telephone-based survey requesting parents to judge their toddlers' familiarity with the selected lexical items. The words included in this questionnaire and their judged familiarities are shown in Appendix B.

³ We thank Dr. Duanmu San for his comments on this table.

Thirty-four parents of 19-month-old toddlers (mean age=595.5 days) from Beijing district consented to participate in the survey. No family who had participated in the survey was later recruited to participate in the experiment. On a scale of 3 (1 = only visually familiar; 2 = visually familiar and label known; 3 = visually familiar, label known and produced), I selected 18 picturable words for the experiment from the questionnaire. Twelve words were used as the test stimuli. These words all have Tone 2 bearing stems, could create nonword mispronunciations and were highly familiar (>1.65) to 19-month-olds. Four words that were highly unfamiliar (<0.65) to 19-month-olds were used as the unfamiliar fillers. The remaining two words that were highly familiar (>1.95) to 19-month-olds were used as the familiar filler. A complete list of stimuli and conditions is provided in Appendix C.⁴

The 12 test stimuli included 4 monosyllabic words (e.g. *chuang*², “bed”), 4 disyllabic words with Tone 2 bearing stems on the first syllable and the “-zi” suffixes on the second syllable (e.g. *hou*²*zi*, “monkey”); and 4 disyllabic words with “xiao-” prefixes on the first syllable and Tone 2 bearing stems on the second syllable (e.g. *xiaoxiong*², “bear”). Such manipulation of the tonal context controls for the fact that the exact pitch of a Mandarin lexical tone can be affected by the preceding tone and the following tone (e.g. Tsao, 2010). As in Experiment 1, to reduce the cognitive load of the task and control for the amount of lexical information carried by each word, all twelve words selected as testing stimuli were mono-morphemic.

⁴ To confirm that the twelve testing stimuli that we had selected were most familiar to 19-month-olds, I further compared these words with the words that are understood by the majority (>50%) of 16-month-old Mandarin infants as shown in the Vocabulary Inventory for Mandarin Chinese (Hao et. al, 2008)⁴. In this CDI, 7 mono-morphemic words carried tone 2 in stems, had picturable referents, and yielded nonce words when mispronounced (Appendix D). Five of the 7 words were included in our questionnaire except *tang*², “candy or sugar” and *chong*²*zi*, “worm, insect, or bug”. These words were not used in our study because they are polysemants and can be pictured as different objects.

The design of Experiment 2 is similar to that of Experiment 1. Infants were tested in 18 trials with 6 fillers and 12 testing trials. The testing trials can be either correctly pronounced or mispronounced. All mispronunciations in the current experiment involved changes in pitch values of lexical tones.

Among the twelve test trials, three trials had the familiar object's label pronounced correctly as Tone 2 (/LH/) (e.g. *hou²zi*, “monkey”). Three trials had the familiar object's label pronounced with a close mispronunciation change, i.e. Tone 3 (/L(H)/) (e.g. *hou³zi*) Three trials had the familiar object's label pronounced with a farther mispronunciation change, i.e. Tone 1 (/HH/) (e.g. *hou¹zi*). The remaining three trials had the familiar object's label pronounced with a farthest change i.e. Tone 4 (/HL/) (e.g. *hou⁴zi*). All mispronunciations resulted in non-words or in words judged unlikely to be familiar to toddlers of this age. Examples of mispronunciations are given in *Table 3*. Each item was presented only once during the experiment.

Table 3
Sample stimuli for tone mispronunciations in Experiment 2

Audio Stimulus Samples	Phonological Features	(Mis)pronunciation	Feature changes
<i>Hou²zi zaina*?</i>	LH	correct	none
<i>Hou³zi zaina?</i>	L(H)	close change	(possibly) ending feature
<i>Hou¹zi zaina?</i>	HH	further change	starting feature
<i>Hou⁴zi zaina?</i>	HL	furthest change	starting & ending feature

* *Where's the monkey?*

Among the six fillers, two were familiar (e.g. “*xiang¹jiao¹*”, “banana”) in which the familiar object was named correctly; and four were unfamiliar in which the unfamiliar object was named correctly (e.g. *huang²guan⁴*, “crown”).

In each trial, infants saw two images, one depicting a referent of a familiar word, the other depicting a referent of an unfamiliar (to 19-month-olds) word. Both the familiar and unfamiliar objects selected for the study were real objects, similar in visual complexity and category (e.g. artifacts, living things, body parts, etc.). An example stimulus pair is depicted in Figure 3. Pairings of familiar and unfamiliar objects remained constant across subjects.

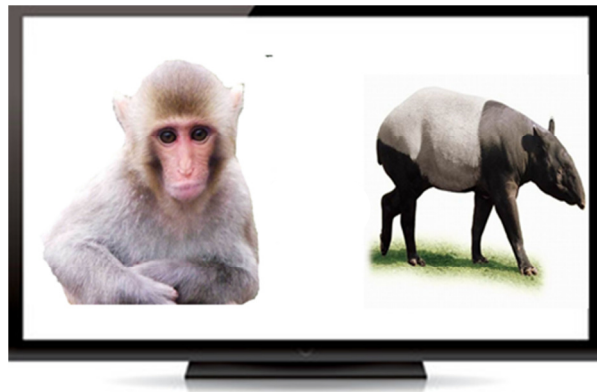


Figure 3. Sample visual stimulus pair in Experiment 2.

All stimuli were naturally produced by a trained female speaker of Mandarin Chinese who produced the utterances with positive infant-directed affect. The mean length of target items across all conditions was 638.05ms (SD = 70.58ms); lengths of target items did not differ across conditions, $F(3, 92) = 0.036$, $p = 0.991$. Pairings of familiar and unfamiliar objects remained constant across subjects. However, the assignment of these stimulus pairs to pronunciation condition was counterbalanced across subjects (see Appendix C), with the exception of six trials (two familiar fillers and four unfamiliar fillers) that were constant across subjects.

Procedure

The procedure was similar to that of Experiment 1. I adapted the intermodal preferential looking procedure to the single-screen display in the lab in Beijing (see

Figure 3). In addition, I used a short video clip of Elmo dancing accompanied by the laughing sound of Woody Woodpecker as the attention getter. This is because the distance between the baby seat and the screen was twice more distant in the lab in Beijing than in the lab I used in the United States for Experiment 1. Moreover, the testing room in Beijing was also three times bigger than the testing room in the United States. Perhaps because of this larger space and greater distance from the screen, babies tested in Beijing were more easily distracted and less likely to pay attention to the task.

Following the experimental session, the parent completed a questionnaire on his/her toddler's comprehension and production of the stimulus items (familiar and unfamiliar). The vocabulary questionnaire was written in Chinese characters; and all the parental reports in the questionnaire were also in Chinese.

Results and discussion

Results from the post-experiment parental questionnaire indicated that the selected items were appropriate for this sample of 19-month-olds. On a scale of 0 (neither say nor understand) 1 (not say, not sure understand or not), 2 (not sure whether say or understand), 3(not say, understand), 4 (not sure say or not, understand), 5 (both say and understand) items used as familiar words received an average score of 4.2 (SD = .15), indicating that they were very familiar to this group of toddlers. Items used as unfamiliar words received an average score of 1.24 (SD = .23).

The dependent measure was the same as in Experiment 1: change in proportional looking to the familiar object between the (silent) salience phase and the test phase. Looking behavior was coded off-line frame-by-frame (1 frame = 33ms) using in-house video coding software. For the salience phase, looking behavior was coded for the entire

4 second duration of the phase. As in Experiment 1, for the test phase, looking behavior was coded only for the first 3 seconds following the onset of the first occurrence of the target word. For each phase, the proportion of time looking towards each of the objects was computed over the total time the subject spent looking at both objects for that phase. As in Experiment 1, for each subject, a difference score was computed for each condition using the following formula:

$$\% \text{Looking (Familiar)}_{\text{Test}} - \% \text{Looking (Familiar)}_{\text{Salience}}$$

Unlike Experiment 1, toddlers exhibited a general preference for the unfamiliar object in the salience phase. Average looking proportion for familiar objects across all conditions was .424 during the salience phase. A single-sample t-test comparing mean salience proportions (across conditions) to chance (95% CI: .33-.47) was significant ($t(15) = 3.211$, $p = .006$). Since novelty preference is often considered a cognitively more advanced behavior (Hunter & Ames, 1988), and the toddlers tested in the current experiment were slightly older than those in Experiment 1, preference to unfamiliar objects during the salience phase may have reflected developments in general cognitive abilities between 19 and 21 months of age.

Across all subjects, there were 70 trials where toddlers only looked at one of the two images in the visual stimuli. Following the same rationale as in Experiment 1, these trials were excluded from further statistical analyses (18.2% of trials). Meanwhile, there was no correctly pronounced trial in which subjects had large negative values when ($<-.30$) (mean=0.09, range=-0.12~0.42). Differences in proportional looking to target objects across (mis)pronunciation conditions are depicted in Figure 4.

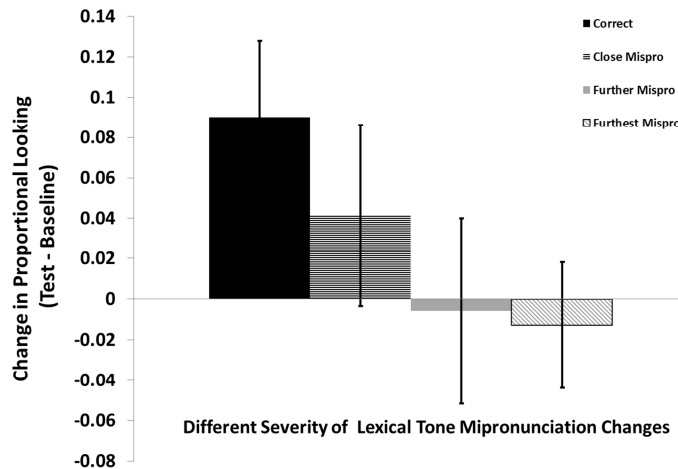


Figure 4. Looking proportion and standard errors, Experiment 2.

As in Experiment 1, two primary analyses were conducted. First, to establish whether there was an effect of mispronunciation on responses, a repeated measure ANOVA across the 4 pronunciation conditions was conducted. Different patterns of looking across the conditions were significant ($F(3, 13) = 3.427, p = 0.049$). To further examine the effect of mispronunciations, planned comparisons were conducted between different mispronunciation conditions. The correct condition and the 1-feature mispronunciation condition were not significantly different ($t(15) = 1.787, p = .094$). Planned comparisons between the correct condition and the 2-feature mispronunciation condition were marginally significant ($t(15) = 2.031, p = .060$); but planned comparisons between the correct condition and the 3-feature mispronunciation condition were significant ($t(15) = 2.848, p = .012$).

Second, to explore graded sensitivity to the three degrees of mispronunciation severity, a repeated measure trend analysis on the four pronunciation conditions was performed. A significant linear trend in difference scores was found ($F(1, 15) = 5.896,$

$p=.028$) while the quadratic and cubic trends were both negligible, replicating the findings in Experiment 1 for codas.

In general, the analyses indicated that toddlers were affected by the severity of mispronunciations of lexical tones. Because preferences in the salience phase were similar across conditions ($F(3, 13) = 0.710$, $p = 0.551$), differences in looking behavior can be attributed to effects of the audio stimuli during the test phase. Therefore, as one of the first attempts in testing the role of lexical tones on early word recognition, this experiment replicated findings for onsets in White & Morgan (2008), vowels in Mani & Plunkett (2012) and codas in Experiment 1.

As discussed in the introduction, access to the information carried by lexical tones can be fragile in adult word recognition, even by speakers of tonal languages. Hence, toddlers might be expected to be less apt to detect mispronounced lexical tones. However, the findings of the current experiment suggest that small phonological changes in lexical tones had similar effects as segmental changes. When presented in a more sensitive reference context where both familiar and unfamiliar objects are target candidates, toddlers displayed graded sensitivities to different degrees of deviations in lexical tone pronunciations. By the middle of the second year of life, toddlers' lexical representation of lexical tones includes sophisticated details.

However, the current experiment only investigated Tone 2 mispronunciations because Tone 2 is the only tone that does not subject to tone sandhi and has graded degrees of mispronunciations (Table 2). Future studies should test whether our results can be generalized to mispronunciations of other tonal categories. For example, we might test whether Tone 2 is a more acceptable mispronunciation form of Tone 1 than Tone 3 for

both Mandarin-speaking adults and toddlers, as predicted in Table 2.

2.3 Discussion of Chapter 2

Research on early lexical representation has been largely focusing on Indo-European languages. In these languages, fundamental frequency (pitch) variation is primarily non-phonemic while phonemic variation is primarily driven by formant height and contour (Howie 1976). As a result, more emphasis has been put on the distinction between the roles of segments and suprasegments in language acquisition and speech perception.

However, like segments, suprasegments/tonemes are also highly frequently used as phonemic features in many languages. Therefore, it is equally important to study the similarity between segments and suprasegments in lexical representations. In particular, comparison of supra-segmental representations between that of segmental representations in infants 'early word recognition is essential for the understanding of the mechanism infants use to cope with various dimensions of information in the speech input.

However, little study in the developmental literature has directly evaluated, compared and incorporated the roles of segmental information and supra-segmental information, especially at the lexical level. The present chapter aims to achieve this goal by investigating English-learning and Mandarin-learning toddlers' sensitivities to different severities of mispronunciations of word codas and tones using a modified intermodal preferential looking procedure. Given that both codas and tones may be less perceptually salient in lexical processing tasks, studying these units can further provide a clearer picture of how much phonetic detail early lexical representations include and whether they are as detailed as in mature lexicon.

The findings of the present study replicated White & Morgan (2008). In particular, for both codas and tones, toddlers' proportional looking to the referents of familiar words decreased with the increase of the degree of mispronunciation severities. Therefore, together with previous findings in the literature, it seems that half way through the second year of life, toddlers' lexical representations comprises very detailed information in both segmental and supra-segmental components, for both prominent lexical elements like onsets (White & Morgan, 2008) and vowels (Mani & Plunkett, 2011) and less prominent elements like codas and tones.

Findings in this study also suggest developmental continuity between early speech perception, early lexical representation and mature lexical representation. The graded sensitivity with mispronunciations found with Chinese toddlers in the present study resembles the discriminative patterns found with 10-to-12 month old Mandarin-learning infants in tone perception in a non-referential task (Tsao (2008) as reviewed in Chapter 1, i.e. with the increase of the phonetic distances in tonal changes, infants are more and more aware of the distinction between the target tone and the deviant tone. It is likely that such sensitivities are carried over from infancy to the second year of life and facilitate word learning. Future studies should test this hypothesis with younger infants (e.g. 12-month-olds) on their graded sensitivities to different severities of tonal mispronunciations using lexical processing tasks similar to the IPLP procedure we used in current study.

The graded sensitivities found with toddlers in our study also resemble findings with native Mandarin adult listeners. For example, using the oddball paradigm, Chandrasekaran, Krishnan, & Gandour (2007), tested Mandarin adults on their discrimination of similar Mandarin tones (T2-T3) and dissimilar tones (T1-T3). The

mean MMN amplitude was found larger for the latter group than the former group.

Findings in Experiment 2 with toddlers suggested that such language-specific perceptual pattern can be traced back at least half way through the second year of life.

The findings of the present chapter may also indicate that toddlers utilize a similar mechanism in processing both segmental and tonal information at the lexical level. Such mechanism may comprise the following properties. First, toddlers estimate the perceptual distance between the correct form and the deviant forms to cope with variations in word recognition. With the increase of the perceptual distance, the degree of lexical activation is decreased in a graded fashion. Second, mispronunciations from different acoustic dimensions (F0 for lexical tones, F1, F2 and F3 for vowels and consonants) can have similar effects in interfering lexical access for 19-month-old toddlers.

To sum up, studies reviewed in this Chapter provide further evidence that the holistic hypothesis is not true. Instead, infants' early lexical representations are very detailed in terms of both segments and suprasegments, even for word units that are not perceptually salient, such as codas and lexical tones. In addition, halfway through the second year of life, it appears that infants utilize a similar mechanism in processing segmental and suprasegmental information in word recognition tasks. In general, we found a clear picture of what toddlers know about the sounds of familiar words, a question of central attention in studies of early lexical representation.

Part II: CONTINUITY IN THE LONG TERM:
FEATURE (UNDER)SPECIFICATION IN INFANTS' AND ADULTS' LEXICON

CHAPTER 3

ASYMMETRICAL PERCEPTION OF DISTINCTIVE FEATURES

The previous part of this dissertation has shown striking continuity between native phoneme acquisition and word learning. Starting from the second half of the first year of life, infants have already acquired native phonological dimensions that are used for lexical identity, i.e. native distinctive features. Toddlers focus on these critical dimensions of contrast (e.g. segments and tones) to differentiate words. Findings in previous chapters answered a critical question of developmental continuity (i.e. continuity in small steps): freshly tuned phonetic categories do map onto new skills that emerge later in life. Thus, continuity exists between these early (but important) steps of language acquisition.

This part of the dissertation addresses another critical question of developmental continuity: Do different stages of language development share similarities in the architecture used for lexical processing? In particular, do infants and adults share the same mechanism in representing distinctive features in the lexicon?

3.1 The special status of coronal place of articulation

The coronal place of articulation is one of the most special distinctive features in the lexicon. Articulated with the tongue tip or blade, it is subjective to many special phonological processes such as place of articulation assimilation, deletion and vowel fusion. In Catalan (Mascaro, 1976), regressive place of articulation assimilation occurs only for *coronal* stops, as in (1):

(1) Stop assimilation in Catalan:

se[t] ‘seven’

<i>se[m] mans</i>	<i>‘seven hands’</i>
<i>se[p] forcs</i>	<i>‘seven fires’</i>
<i>se[b] beus</i>	<i>‘seven voices’</i>
<i>se[d] dones</i>	<i>‘seven women’</i>
<i>se[k] cases</i>	<i>‘seven houses’</i>

Labial and velar stops do not undergo such assimilation, as in (2):

(2) Stop assimilation in Catalan :

<i>ca[p]</i>	<i>‘no’</i>
<i>ca[p] signe</i>	<i>‘no sign’</i>
<i>po[k]</i>	<i>‘few’</i>
<i>po[k] pa</i>	<i>‘few bread’</i>
<i>po[k] sol</i>	<i>‘few sun’</i>

Such differences between coronal and noncoronal phonemes in phonological assimilation are also found in many other languages, such as English place assimilation (Gaskell & Marslen-Wilson, 1996) and Maltese vowel-consonant assimilation (Hume, 1996).

Besides assimilation, phonological research has found that vowels separated by coronals, but not other consonant type, often behave as if they were adjacent. Such a process is referred to as vowel fusion. Guere, for example, is a tonal language spoken in the West African country of Côte d'Ivoire. In this language, intervening coronals between vowels can be optionally deleted, whether both flanking vowels are identical or not. As shown in (3) below, [l] is a free intervocalic variant [d] in Guere.

(3) Guere coronal consonant deletion (Paradis & Prunet, 1989)

Original	Variant	Variant	English Gloss
----------	---------	---------	---------------

<i>zida</i>	<i>zila</i>	<i>zia</i>	“Separate!”
<i>kide</i>	<i>kile</i>	<i>kie</i>	“Laugh!”
<i>zide</i>	<i>zile</i>	<i>zie</i>	“Escape”
<i>vide</i>	<i>vile</i>	<i>vie</i>	“Throw away”

Similar vowel fusion processes also occur in many other languages such as Fula, a language spoken in Central and Western Africa. In this language, imperfect suffixes geminate by always spreading rightward across the intervening coronal consonant and such germination does NOT occur when the intervening consonant does not have a coronal place of articulation (see (Paradis & Prunet (1989) for more detail).

3.2 The nativist approach: Underspecification of distinctive features in the lexicon

To provide explanations for why pronunciations of nominally coronal segments in words are more subject to special phonological processes than are pronunciations of noncoronal segments, theories of phonology have proposed that the distinctive feature of coronal place of articulation is *the natural class of consonants* (Paradis & Prunet, 1989) or the *least marked* (Kean, 1975) in terms of theoretical phonology.

Chomsky & Halle (1968) proposed that in world languages, features should be filled-in the phonological representations due to markedness conventions (SPE; Kean 1975), with marked distinctive features filled in and unmarked features not filled in in the phonological representations. In applying these rules to the lexical representations, the underspecification hypothesis (Archangeli, 1986) claims that unlike other distinctive features of place of articulation, the [+coronal] feature is not filled in the phonological lexicon. If a feature is not filled in the phonological lexicon, this feature is

underspecified/unspecified (Avery & Rice, 1989; Kiparsky, 1982). In this sense, the coronal place of articulation is thus underspecified/unspecified.

Since early 1980s, the underspecification hypothesis has been illustrated under different phonological frameworks. For example, in the framework of feature geometry (Clements, 1985), coronal place of articulation is underspecified in the way that “unmarked coronals universally lack a Place node” (Paradis & Prunet, 1989). According to Optimality Theory (Prince & Smolensky, 1997), coronal place of articulation is underspecified in the way that it acts as a universal constraint and ranked highly in the mental phonological grammar. These proposals of underspecification are all couched in a principles-and-parameters approach similar to that proposed by Chomsky (1987) for syntax, and that proposed by Kaye et. al. (1985), Paradis (1989), Paradis & Prunet (1989a & b) and Piggott (1987, 1988), among others, for phonology. In such a nativist approach, the underspecified status of coronal place of articulation is defined as a universal constraint (Prince & Smolensky, 1997) or convention. In this vein, like all other principles determined by UG, underspecification is innate and thus independent of experience with a specific language.

3.3 Psycholinguistic predictions of underspecification

The claim that the coronal place of articulation is not specified in the lexicon predicts an asymmetry in word recognition. For example, /d/, as a coronal sound, is unspecified (or unmarked) for place of articulation in the lexicon; thus, words that contain stops with other places of articulation may be perceptually equivalent to /d/. In contrast, /b/, as a labial sound, is fully specified in the lexicon, and hence it is not likely to be misperceived as other places of articulation. As a consequence, the word *buck* may be an exemplar of

duck but *duck* cannot be an exemplar of *buck*. A word-recognition model incorporating these properties – the Featurally Underspecified Lexicon (FUL) – has been developed by Lahiri and colleagues (Lahiri & Marslen-Wilson, 1991; Lahiri & van Coillie, 1999).

Psycholinguistic studies, using a variety of tasks, have adduced evidence supporting predictions of the FUL model. For example, at temporally early stages of speech perception as measured by Event-Related Potentials (ERP), German-speaking adults display asymmetric discrimination for mispronunciations of familiar words with coronal or noncoronal onsets (e.g. ***Br**achen --**Dr**achen [dragon] vs ***D**renze -- **G**renze [border]) (Friedrich, Lahiri, & Eulitz, 2008), word internal consonants (e.g. ***Hor**be--**H**orde [horde] vs ***Pro**de--**P**robe [sample]) (Friedrich, Eulitz, & Lahiri, 2006; Cornell, Lahiri & Eulitz, 2013), and codas (e.g. ***Bah**n-- **Bah**m [railway] vs ***baum**--**baun** [tree]) (Lahiri & van Coillie, 1999), supporting the notion of underspecification. Similar effects have also been found with English-speaking adults. Roberts, Wetterlin & Lahiri, (2013), for example, conducted two experiments using lexical decision with semantic priming in British English. As predicted by underspecification, results from both experiments, including reaction time latencies and N400 amplitudes, exhibit asymmetries in the acceptance of mispronunciations for coronal vs. noncoronal consonants (e.g. ***te**mor--**ten**or vs ***in**age--**im**age). Earlier studies using gating and lexical decision procedures also demonstrated putative effects of underspecification in Bengali, English, and German (Gaskell & Marslen-Wilson, 1996; Lahiri & Marslen-Wilson, 1991; Lahiri & Reetz, 2010; Wheeldon & Waskler, 2004).

3.4 Underspecification and Developmental Continuity

Findings on adults' asymmetrical perception of coronal and noncoronal features naturally direct to the question of developmental continuity: Do infants also represent and process coronal features and noncoronal features asymmetrically in the early lexicon or earlier in their attunement to phonetic categories?

To date, existing empirical research on infants' early word recognition has rarely directly considered such possible asymmetrical processing in early lexical representations. Moreover, certain theoretical claims may predict that effects of underspecification might be observed more strongly in infants' early lexicon than in adults' mature lexicon, suggesting developmental **discontinuity**. Language experience may cause effects of underspecification to become attenuated over time. Several empirical accounts of early lexical representation have indicated that the degree of phonological specification correlates positively with the toddler's vocabulary size (Charles Luce & Luce, 1990, 1995; Storkel, 2002; Walley, 1993, 2005). According to the Lexical Restructuring Model (Metasala & Walley, 1998), for example, both speech sound categorization and the concomitant increase in the specificity with which individual words are represented are caused by word learning itself. As the child's vocabulary expands, the lexicon comes to contain more and more sets of similar sounding words. According to such views, toddlers, whose vocabularies are markedly smaller, are more likely to have an underspecified lexicon than adults. In this way, it is possible that not only coronal place of articulation is underspecified; but that other distinctive features are all likely to be underspecified as well, making infants not sensitive to *any directions* of sound changes in word recognition tasks. Indeed, in a recent study, for example, van der Feest & Fikkert (2015) tested 20-month-old Dutch-learning infants on their sensitivities

of familiar word mispronunciations in voicing (e.g. /boom/ is mispronounced as /poom/ or /pop/ is mispronounced as /bop/). They found that 20-month-olds failed to differentiate between the correct and the mispronounced forms in either direction of mispronunciations.

Interestingly, findings on pre-lexical infants' asymmetrical sensitivities to different directions of sound changes do not seem to support any of the above predictions on developmental **discontinuity**. Instead, studies do find evidence for early appearance of asymmetries in infants' speech perception tasks as predicted by underspecification. Tsuji, Mazuka, Cristia & Fikkert (2015) tested 4-month-old Dutch-learning infants and Japanese-learning infants on their discrimination of [ɔmpa] (labial) and [ɔnta] (coronal) in a habituation task. They found that for both language groups, 4-month-olds were able to show that they discriminate [ɔmpa] and [ɔnta] only when they are habituated with [ɔmpa]. Therefore, [ɔnta] cannot be [ɔmpa] but the opposite is possible. Similarly, Dijkstra & Fikkert (2011) tested 6-month-old habituated Dutch-learning infants to either repeated *taan* or *paan* tokens and then tested their ability to discriminate using trials in which one or the other stimuli were repeated or trials in which the two stimuli alternated. Whereas infants habituated to *paan* discriminated the two types of trials, infants habituated to *taan* did not. The authors of both studies interpreted their findings in terms of underspecification. For example, in the context of Dijkstra & Fikkert (2011), since the place of articulation of the onset /t/ of the nonword *taan* is unspecified in lexical representation; mispronunciations of the onset of *taan* such as *paan* are not incompatible with 6-month-old infants' underlying representations of *taan*. By contrast, the onset /p/ of the word *paan* is specified as [+labial] in the underlying lexical representation, so

mispronunciations of the onset of *paan* such as *taan* are incompatible with 6-month-old infants' underlying representations of *paan*. Therefore, when the standard of comparison was a coronal feature, place of articulation of the consonants was unspecified, so that both coronal and noncoronal features were compatible with the standard, but when the standard of comparison was noncoronal, place of articulation of the onset was specified, so that only noncoronal was compatible with the standard, and coronal and noncoronal features were discriminable.

As findings in Chapter 2 suggest, infants' representations of native speech categories can be carried over into early word recognitions. Therefore, if perceptual asymmetry is due to the differences in which features are specified in different speech categories, it is likely that infants carry such asymmetrical representations to word recognition tasks. Therefore, we should expect that infants are also asymmetrically sensitive to different directions of mispronunciations in word recognition.

Indeed, studies seem to find evidence for early appearance of asymmetries in infants' lexical processing as predicted by underspecification. In a novel word learning task, Fennel (2007), for example, found such asymmetry with older infants in novel word learning. Using the "Switch Task", the author habituated 14-month-old English-learning infants with a novel object associated with either a nonword starting with a coronal consonant /**d**in/, or a different nonword starting with a labial consonant: /**b**in/. After habituation, two types of test trials were presented: the "Same" trial that involved the same word-object combination, and the "Switch" trial that used the same object but was labeled by a different sound. For example, infants who were habituated with /**b**in/ were presented with /**d**in/ in the "switch" trial, and infants who were habituated with /**d**in/ were

presented with /**bin**/ in the “switch” trial. The findings were similar to those of Dijkstra & Fikkert (2011), with 14-month-old English-learning infants only detecting a labial-to-coronal (/b**in**/-to-/d**in**/) switch and failing to detect a coronal-to-labial (/d**in**/-to-/b**in**/) switch. To interpret the findings in the framework of underspecification, the author suggested that since the place of articulation of the onset /d/ of the nonword *din* is unspecified in lexical representation, mispronunciations of the onset of *din* such as *bin* are not incompatible with 14-month-olds’ lexical representations of the nonword *din*. By contrast, the onset /b/ of the word *bin* is specified as [+labial] in the underlying lexical representation, so mispronunciations of the onset of *bin* such as *din* are incompatible with 14-month-olds’ lexical representations of the nonword *bin*.

However, it is worth noting that in this novel word-learning task, 14-month-old infants may have found it more difficult to establish the association between the novel object with the coronal onset label /d**in**/ than with the noncoronal onset lab /b**in**/. Studies have suggested that this alternative explanation is quite plausible. On one hand, the switch task is highly cognitively demanding, and successful word learning in such a context is only possible in very restricted circumstances (Yoshida et. al., 2009; Fennell, & Waxman, 2010). On the other hand, as indicated previously, coronal sounds are more diverse in their pronunciations than noncoronal sounds. Studies have suggested that categories with higher variability are more difficult to learn than categories with lower variability. For example, as suggested in Goldstone (1998), a general finding has been that increasing the variability of instances within the categories to be discriminated increases the amount of training time needed to reach a criterion level of accuracy, but also yields better transfer to novel stimuli (Posner & Keele 1968). Therefore, to establish

whether infants represent coronal and noncoronal categories asymmetrically in their early lexicon, we may need to use a task that is less demanding and better resembles infants' daily word recognition tasks.

Van der Feest & Fikkert (2015) seems to have achieved such a goal. In a familiar word recognition task, the authors tested 20- and 24-month-old Dutch-learning infants on their sensitivities to coronal to noncoronal and noncoronal to coronal mispronunciations of familiar words using the intermodal preferential looking procedure similar to the one used in Chapter 2. They found that both groups of infants could only discriminate noncoronal to coronal mispronunciations (e.g. /boom/ is mispronounced as /d**o**om/) but failed to discriminate coronal to noncoronal mispronunciations (e.g. /**d**uif/ is mispronounced as /**b**uif/). As the authors suggested, such findings supported the underspecification hypothesis that coronal sounds such as [d] do not have the feature of place of articulation specified while noncoronal sounds such as [b] do have the feature of place of articulation specified in toddlers' lexicon.

However, scrutiny of this study reveals an important issue on their experimental design: each subject in their experiment was exposed to multiple mispronunciation conditions of a single testing item. For example, for the word *boom* (tree), a subject hears the correct pronunciation /boom/, the mispronunciation /d**o**om/ and a different kind of mispronunciation /**p**oom/ which they used to test a different research question, i.e. mispronunciation in voicing, as reviewed above. Whereas hearing *boom* once in a specific (mis)pronunciation condition may interrogate the degree to which it activates the word - a perceptual task - hearing it in more than one pronunciation might be construed as asking whether *doom*, for example, is a plausible mispronunciation of *boom*. This

latter is a metalinguistic task; subjects may use their (implicit) knowledge that coronals frequently assimilate in place to preceding or following segments whereas noncoronal segments do so rarely if at all to judge that *doom* is not a plausible mispronunciation of *boom* but *peen* may be a plausible mispronunciation of *teen*. Tasks in which responses are influenced by supra-lexical, metalinguistic knowledge cannot provide decisive evidence on lexical underspecification. Thus, the question of whether coronal-vs.-noncoronal asymmetry appears in lexical representations remains uncertain.

To better comprehend in what way results with pre-lexical infants can be immediately translated into *lexical* representations, Chapter 4 and 5 will examine whether different directions of mispronunciations (i.e. coronal to noncoronal and noncoronal to coronal) asymmetrically impair word recognition by presenting subjects each testing item for only once. Since underspecification is construed as reflecting properties of *lexical* representations, if the FUL model holds true in feature specification of early lexical representations, it is predicted that infants should be asymmetrically sensitive to coronal-vs.-noncoronal mispronunciations with noncoronal to coronal mispronunciations more detectable than the reversed direction. Such finding would suggest developmental continuity: infants and adults use similar architecture in processing words.

3.5 Underspecification Re-Evaluation

The previous sections have reviewed studies examining coronal-vs.-noncoronal asymmetry as predicted by the Featurally Underspecified Lexicon (FUL) model. While underspecification seems to have provided an explanation for asymmetrical perception found in many studies reviewed above, it is nevertheless not the optimal explanation for

coronal-vs.-noncoronal asymmetries for the following reasons as will also be shown in later chapters of the thesis.

First, it is important to note that the underspecification hypothesis is construed as reflecting properties of *lexical* representations. While it is true that young infants can form representations that have at least some lexical properties (e.g., they can use their own names and other highly familiar forms as anchors for segmentation (Bortfeld, Morgan, Golinkoff, & Rathbun, 2005), and they can associate some form of reference with well-known words (Bergelson & Swingley, 2012)), there is no evidence that 6-month-olds can do any of these things with the repeated, isolated forms characteristic of habituation stimuli as in the settings of Dijkstra & Fikkert (2011), or that the tasks have tapped on to *lexical* representation in the settings of Fennell (2007) where nonwords were used. Thus, the conservative interpretation is that infants treat such stimuli as phonetic sequences, and that the asymmetry observed in those habituation tasks may simply reflect a certain kind of perceptual bias.

Moreover, it is important to note that underspecification fails to account all coronal vs. noncoronal asymmetries. One example comes from Fennell, van der Feest, & Spring, 2010). In this study, the authors tested 14-month-old English-learning infants on their discrimination of coronal-to-velar (/d/-to-/g/) changes using the same design as in Fennell (2007). In particular, infants who were habituated with /gin/ were presented with /din/ in the “switch” trial, and infants who were habituated with /din/ were presented with /gin/ in the “switch” trial. However, unlike findings in Fennell (2007), 14-month-olds who were habituated with /din/ detected the coronal-to-velar (/din/-to-/gin/) switch and the other group of 14-month-olds who were habituated with /gin/ failed to detect the

coronal-to-velar (/d**in**/-to-/g**in**/) switch. This pattern of asymmetry opposes the prediction of the underspecification hypothesis.

To account for this reversed asymmetry, Fennell, van der Feest, & Spring (2010) proposed a frequency-based account. In particular, after analyzing the first three formants (F1, F2 and F3) at six points in the first 40ms of the vowel in all the /bin/, /din/, /gin/ sounds that were used in their auditory stimuli, the authors found that the labial consonants in their stimuli were the least variable speech sounds in terms of these acoustic parameters, coronals are more variable than labials, and velars are the most variable among the three categories. Therefore, they suggested that the observed perceptual asymmetry may result from the difference in the acoustic variability of these consonants. In particular, the more variable the acoustic properties of a certain speech sound, the harder it is to detect change.

Chapter 6 will discuss this variability-based account in further detail. As will be shown in later chapters, this frequency-based account is more convincing than underspecification. Not only because it is built upon concrete acoustic measures such as formant transitions rather than the hypothetical term of “specification”, but also that it parallels with accounts that have been derived to explain similar perceptual asymmetries found in other cognitive domains (e.g. Mullennix & Pisoni, 1990; Quinn, Eimas, & Rosenkrantz, 1993; Eimas, Quinn, & Cowan, 1994). For example, Quinn, Eimas, & Rosenkrantz (1993) found that 3-to-4-month-olds show an asymmetry in their discrimination of changes between dogs and cats. Infants habituated with dogs failed to discriminate dogs from cats while infants habituated with cats could still discriminate dogs from cats (Quinn, Eimas, & Rosenkrantz, 1993). To account for such an asymmetry,

in several follow up studies (e.g. Eimas, Quinn, & Cowan, 1994), the Quinn and colleagues analyzed the facial features of their dog and cat stimuli, and found that the dogs they used in Quinn, Eimas, & Rosenkrantz (1993) were more variable in their appearance than cats. The authors further proposed that changes from less variable to more variable categories are more detectable than vice versa. Therefore, infants found cat-to-dog changes more detectable than dog-to-cat changes, resembling the variability-based account proposed by Fennell et. al. (2010). Moreover, research in speech perception has also suggested that variability plays an important role in perceptual learning and robust category formation (e.g. Lively, Logan & Pisoni, 1993).

Therefore, instead of underspecification, it is possible that certain statistical properties of speech categories are the real causes of perceptual asymmetry. Further research is needed to explore this possibility. In particular, it needs to determine why statistical properties such as variability can affect category formation and perception both in speech and other cognitive domains. To achieve such a goal, Chapter 6 will present a computational model in the framework of Bayesian Inference to formally and systematically explain how stimulus variability could statistically introduce perceptual asymmetry.

3.6 Stop versus Fricative Asymmetry

As is implied by the name, the Stop versus Fricative Asymmetry refers to the phenomenon that listeners are asymmetrically sensitive to different directions of changes between stop consonants [e.g. /b/, /p/] and their fricative counterparts [e.g. /v/, /f/] that share the same/similar features of place of articulation (e.g. [+labial]) as the stop sounds but differ in manner.

In a habituation (switch) task, Altvater-MacKensen & Fikkert (2010) for example, tested Dutch-learning 14-month-olds on their discrimination between the non-word minimal pair /paap/ and /faap/. Infant subjects in this study noticed the switch when they were habituated with /faap/, but did not notice the switch when they were habituated with /paap/. Following the underspecification hypothesis, the author suggested that stop consonants are underspecified in terms of manner while fricatives are specified in manner. Therefore, according to the FUL model, switch from the more specified sound /faap/ to underspecified sound /paap/ is more detectable than vice versa.

Nevertheless, as discussed in the previous section, underspecification is construed as reflecting properties of *lexical* representations. Thus it is not a good account in explaining asymmetries found in pure speech perception tasks. Moreover, as suggested earlier, if stop-vs.-fricative asymmetry is due to the differences in which features are specified in different speech categories, infants should carry such asymmetrical architecture to word recognition tasks and we should expect an asymmetry in tasks where infants are to discriminate different directions of mispronunciations in familiar word recognition. To examine this question, in a later study, Altvater-MacKensen et al. (2014) tested Dutch-learning 25-month-olds on their detection of familiar word mispronunciations using the intermodal preferential looking paradigm similar to the one used in this thesis. However, contrary to the prediction of underspecification, they found that the 25-month-olds successfully detected *both* types of mispronunciation.

Nam & Polka (under review), for example, examined in English- learning and French-learning 5-6-month-olds on their asymmetrical sensitivities to changes between /b/ and /v/ sounds. Both sounds carry the labial feature in place of articulation but differ

in manner (stop vs. fricative). Findings suggested that infants from both language groups noticed the difference when /vas/ changed to /bas/ but not when /bas/ was changed to /vas/. Instead of relying on featural underspecification, the authors interpreted this asymmetry as a perceptual bias, which is a natural consequence of the acoustic-phonetic patterns generated by stop and fricative productions: The stops have a more rapid, abrupt amplitude shift, an acoustic property that is perceptually more salient to the auditory nerve system. Due to the perceptual saliency, the stops serve as natural referent consonants of the fricatives in infants' perception of consonant manners. As will be introduced in the following section, this hypothesis is called the natural reference hypothesis and the perceptually more salient speech categories serve as the reference categories. Changes from the reference to the nonreference categories are not as detectable as vice versa.

By interpreting perceptual asymmetries as phonetic/acoustic biases, this hypothesis naturally falls into the theoretical frame of PRIMR (Werker & Curtain, 2005). Similar to the predictions of PRIMIR, the natural reference hypothesis suggests that infants are born with initial biases such as asymmetrical perception. However, as infants receive more exposure to their native languages, such bias will be ultimately overridden by native phonetic contrasts. In this way, following the rational of PRIMIR, the natural reference hypothesis explains findings by Altvater-MacKensen et al. (2014) that the 25-month-olds symmetrically detected *both* types of mispronunciation (i.e. stop-to-fricative and fricative-to-stop).

3.7 Perceptual Asymmetry in Vowels

Like in consonant perception, directional asymmetries in vowel perception were found very early in infancy. Bohn & Polka (2001) tested 6-12-month-old German-learning infants on their discrimination of the three German contrasts: [i]–[e], [e]–[I] and [o]–[U]. Using a head-turning preference procedure, they found that the [e]–[i] change was more detectable than the [i]–[e] change; the [I]–[e] change was more detectable than the [e]–[I] change and the [U]–[o] change was more detectable than the [o]–[U] change, respectively. Similar results were later replicated in studies with infants from different language backgrounds and using different experimental procedures (see Polka & Bohn, 2003, 2011 for a systematic review).

To account for these asymmetries in vowel perception, Polka & Bohn (2003) proposed that directional asymmetries in vowel perception can be predicted by the relative position of each vowel in the vowel space as defined by dimensions of F1 (vowel height) and F2 (front/back). Similar to the acoustics-based account in Nam & Polka (under review), more peripheral vowels are perceptually more salient. Thus, they act as natural referent vowels. Therefore, vowel changes from less salient/more central ([e], [I], [U]) to more salient/more peripheral vowels are consistently easier to discriminate than vice versa. This account is called the Natural Referent Vowel hypothesis (NRV). As suggested in the preceding section, NRV assumes that infants' preference for peripheral vowels is innate.

To further examine this hypothesis, Polka & Bohn (2011) tested Danish-learning infants aged 6-to-12 months on their discrimination of the contrast between the second and the third highest front unrounded Danish vowels, [e] and [ɛ]. Consistent with the prediction from NRV, they found that changes from [ɛ] to [e] were significantly more

detectable higher than vice versa. Moreover, to examine whether the perceptual asymmetry was affected by the age of the infants, the authors separately analyzed the data from the youngest infants aged between 6:10 and 9:02 months, and the data from the oldest infants aged between 10:14 and 11:28 months. They found that younger infants aged between 6:10 and 9:02 months could differentiate changes from /ε/ to /e/ significantly better vice versa, as predicted by NRV. However, older infants aged between 10:14 and 11:28 could successfully detect *both* directions of changes in the same level of significance. The authors indicated that the significant developmental change might be due to infants' experience with a native vowel contrast, which overrode their initial perceptual (phonetic) bias. Polka, Bohn, & Molnar (2005) also reported a similar directional asymmetry in German-learning infants between the German contrast: [u]–[y] (Polka & Bohn, 1996; Polka & Werker, 1994). Also consistent with the prediction of NRV, infants were more sensitive to [y] to [u] changes than vice versa. In addition, their results also showed a significant effect of language experience. German adults tested with the same task and stimuli did not show directional asymmetries; instead, their performance was essentially errorless in *both* directions of vowel changes.

Pons et. al. (2012), using a visual habituation procedure, tested Catalan-learning and Spanish-learning infants on their discrimination of the vowel contrast [i]–[e] which occurs in both languages. Their results also confirmed the prediction of NRV: At 4 month of age, infants from both language groups found [e] to [i] changes more detectable than vice versa. However, by 12 month of age, while Catalan-learning infants still looked significantly longer on switch trials when the direction of change went from the vowel /e/ to the vowel /i/, Spanish-learning infants' looking behaviors were flipped: they looked

significantly longer on switch trials when the direction of change went from /i/ to /e/. The author interpreted this discrepancy in responses between Catalan-learning infants and Spanish-learning infants in terms of the effect of language exposure. In particular, in Spanish, the medial vowel /e/ is more frequent than the vowel /i/ (Alcina & Blecua, 1975); whereas in Catalan, the opposite case is observed, the vowel /i/ is more frequent than vowel /e/ (Rafel, 1980). Therefore, findings with the older group of infants confirmed a significant role of language experience with changes from less frequent to more frequent categories are more detectable than vice versa. This account will be examined further in later chapters.

While NRV model successfully describes most asymmetries discovered by Polka and colleagues in their works, one obvious challenge to this model is that we do not know how to utilize NRV to account for asymmetries between vowels differing in dimensions beyond F1 & F2. For example, Polka & Bohn (2011) reported a study that tested 6- and 11-month-old Danish-learning infants on their discrimination between the Danish contrasts [e]–[ø]. The two vowels differ primarily in their roundness (F3). Also, [e] is more peripheral than [ø]. However, contrary to NRV, infants at both age groups found changes from the more peripheral [e] to the less peripheral [ø] significantly more discriminable than vice versa. Perhaps in future research, F3 dimension is also needed to be incorporated into the NRV framework to make it a more powerful account.

Another challenge to NRV arises from Japanese long and short vowels. These vowels are the same in qualities. The only difference is the vowel duration. In a recent study, Mugitani, et. al. (2009) found that 18 month-old Japanese-learning infants discriminated a long to short vowel change but not a short to long vowel change. The

authors interpreted the finding using a *frequency-based account* similar to the one proposed by Pons et. al. (2012). In particular, short vowels are about ten times more frequent than long vowels in occurrence in Japanese. Thus, it is likely that long vowels with a low frequency of occurrence might have not yet been categorized as long vowels by 18-month-olds, but rather as atypical, longer exemplars of short vowels. In contrast, concentrated exposure to short vowels enhances the salience of short vowels relative to long vowels and hence confers a processing advantage for short vowels as the referent anchor in perception. In this way, the authors indicated that asymmetries found in early vowel perception might be explained within the framework of Natural Language Magnetic Effect, as proposed by Kuhl (1991), which suggested that changes from prototypical exemplars to non-prototypical exemplars are less detectable than vice versa. Thus, according to Mugitani, et. al. (2009), this frequency-based account successfully explains the asymmetries found in native vowel perception by predicting the acquisition order of different native vowel categories (e.g. short vowels are acquired earlier than long vowels by Japanese-learning infants due to their high frequency of occurrence in Japanese).

This frequency-based account suggests that distributional information (such as frequency of occurrence and category variability) could be a cause for certain asymmetries in speech perception, with the more frequent categories the reference category. Such a hypothesis is consistent with the majority of findings that infants use the distributional information of their native speech categories in speech perception (e.g. Saffran, Aslin & Newport, 1998). In addition, it provides an alternative account for many perceptual asymmetries found in the literature. In particular, besides the [ɛ]-vs.-[e]

asymmetry in Catalan and the long-vs.-short vowel asymmetry in Japanese, this hypothesis it can also be used to account for coronal-vs.-noncoronal asymmetry in consonant perception. As will be discussed in Chapter 6, coronal categories are more frequent crosslinguistically than noncoronal categories. According to the hypothesis proposed in Pons et. al. (2012), changes from the less frequent vowel to the more frequent vowel is more detectable than vice versa. Therefore, listeners found it easier to discriminate changes from noncoronal to coronal sounds than vice versa.

This frequency-based account will be further explored in Chapter 6 using Bayesian Inference. Instead of relying on acquisition order as indicated by Mugitani, et. al. (2009), the computational model proposed in Chapter 6 will suggest that any categories with higher frequency of occurrence will be more salient in perception due to their statistical properties. Unlike NRV which claims that the asymmetrical sensitivity reflects an innate bias, the frequency-based approach completely relies on the role of language experience.

3.8 Chapter Conclusion

In general, this Chapter reviewed studies on directional asymmetries in infant early speech perception and their developmental changes. Asymmetries in consonant perception were reviewed primarily in the framework of Feature Underspecification Lexicon (FUL). However, it has been suggested that underspecification may not be the optimal explanation. Alternative accounts have been proposed, such as the variability-based account that relies on the acoustic variability; and the natural reference account that relies on the acoustic saliency of different consonants categories.

Asymmetries in vowel perception are reviewed primarily in the framework of Natural Referent Vowel (NRV). The Natural Reference Vowel hypothesis has been very successful in explaining a majority of perceptual asymmetries found in the literature of infant vowel perception. In addition, recent studies (Nam & Polka, under review) have extended this account to explain asymmetries in consonant perception. However, NRV is facing challenges. Given that NRV focuses on innate perceptual biases, orthogonal models are needed to explain asymmetries that are possibly due to language exposure (e.g. Japanese long-vs.-short vowel). Therefore, the frequency-based account has been suggested: perceptual asymmetries can be due to differences in frequency of occurrence, with changes from categories of higher frequency to categories of lower frequency less detectable than vice versa.

Studies reviewed in the current chapter have raised a series of open issues in terms of infants' early lexical structure.

First, it is still to be determined whether perceptual asymmetries occur at the phonetic level or the lexical level. According to the underspecification hypothesis (or FUL), asymmetries arise from differences in featural specification. Therefore, they are construed lexical rather phonetic. However, based on other framework (i.e. NRV and the variability-based account and the frequency-based account), asymmetries can be purely phonetic effects that arise from differences in phonetic properties of various speech categories. Such phonetic properties can be either acoustic (e.g, NRV) or statistical/distributional (e.g. variability-based account and frequency-based account) or both.

Whether perceptual asymmetries occur at the phonetic level or the lexical level

will have very different predictions on infants' early lexical representations. In particular, as discussed in preceding sections, the FUL model is construed as a lexical model that depicts feature architectures in infants and adults' lexicon. Therefore, if the lexical architecture is continuous between infancy to adulthood (i.e. developmental continuity), infants' early lexical should also asymmetrically specify coronal and noncoronal places of articulation. Therefore, in word recognition tasks, both infants and adults should be asymmetrically sensitive to coronal-to-noncoronal and noncoronal-to-coronal mispronunciations, with the latter more discriminable than the former. However, if asymmetries occur at a purely phonetic level, we would not expect asymmetries to occur in word recognition tasks. According to the PRIMIR model of language acquisition (Werker & Curtain, 2005), phonetic biases, such as perceptual asymmetry, that are not necessary in lexical representation will be stripped away in word recognition tasks, leaving information that is most relevant to features in the lexical representations to be used to identify lexical items. Therefore, we can easily predict that if certain perceptual asymmetries are due to phonetic biases, such bias to occur should not occur in word recognition tasks if they are not lexically relevant. In this way, we should expect symmetrical sensitivities to different directions of familiar word mispronunciations, similar to findings in Altvater-MacKensen et al. (2014).

Second, it is debated whether experience with native phonological contrasts can override perceptual asymmetries over development. The underspecification hypothesis and other accounts for perceptual asymmetries make very different predictions in this respect. Coming from the nativist approach, the underspecification hypothesis predicts that coronal sounds are innately underspecified in the lexicon while noncoronal sounds

are specified in terms of place of articulation. Moreover, according to such a hypothesis, if certain features are underspecified, it should be underspecified over development, at different stages of lexical processing and in different languages. Thus, we would expect that both toddlers and adults are similarly asymmetrically sensitive to different directions of familiar word mispronunciations as suggested in van der Feest & Fikkert (2015).

However, other accounts for perceptual asymmetry incorporate the important role of language exposure. In particular, although the NRV hypothesis conceives perceptual asymmetries as initial perceptual bias, it recognizes the fact that as infants receive more exposure to their native language, such biases may be attenuated and ultimately disappears (see Polka & Bohn, 2011). The variability-based account (Fennell, 2007) and the frequency-based account (Pons et. al., 2012) are two models that completely rely language experience. Therefore, these model predict that in word recognition tasks, toddlers and adults may be sensitive to different both directions mispronunciations (e.g. coronal-to-noncoronal and noncoronal-to-coronal) due to their experience with native phonetic contrasts.

The following chapters are designed to address these questions. Chapter 4 and 5 will examine whether toddlers and adults are (a)symmetrically sensitive to different directions of coronal-vs.-noncoronal mispronunciations. Chapter 6 will introduce a computational model that further investigates the possible causes of asymmetries in speech perception.

CHAPTER 4

UNDERSPECIFICATION IN TODDLERS' LEXICAL REPRESENTATIONS

The current chapter presents a series of studies examining possible asymmetries in English-learning toddlers' detection of correct and incorrect pronunciations of familiar words. Tested were coronal-to-labial, labial-to-coronal, coronal-to-velar, and velar-to-coronal mispronunciations in an online word recognition task using the visual world paradigm.

As discussed in Chapter 3, if toddlers' lexical representations are unspecified for coronal place of articulation as predicted by the underspecification hypothesis, then we expect to see significant asymmetries with larger effects for labial-to-coronal and velar-to-coronal mispronunciations than for coronal-to-labial and coronal-to-velar mispronunciations. By contrast, if all distinctive features are specified in infants' early lexicon, including the coronal place of articulation, then different directions of mispronunciations should equally affect infants' word recognition. This chapter presents two experiments designed to evaluate these two competing accounts that make divergent predictions for the patterns of results.

As mentioned in Chapter 2, White & Morgan (2008) introduced several refinements to the preferential looking mispronunciation task pioneered by Swingley & Aslin (2000). First, rather than displaying two referents with known labels in each trial, White & Morgan paired a referent with a known label and a referent without a known label. The rationale for this modification was that known labels might reduce the sensitivity of the paradigm by repelling mispronunciations towards the correct forms. For example, if a toddler were to see pictures of a baby and a car (both known labels to the

baby) and hear a mispronunciation *vaby*, because that pronunciation is not a possible label for the car, changes in looking time relative to the correct pronunciation *baby* might be limited.

In contrast to Swingley and Aslin (2000), White & Morgan also tested each item on only a single trial. This modification is of importance for studies of asymmetries in speech perception because testing items repeatedly raises possibilities of carry-over effects as well as adaptation and may change the perception of the stimuli. In particular, whereas hearing *vaby* once may provide evidence for whether it is an acceptable pronunciation of *baby* (a perceptual task), hearing these two pronunciations multiple times might be construed as asking whether *vaby* is a plausible mispronunciation of *baby*. To find a solution to such a question, infants might need to draw on implicit metalinguistic knowledge. In this way, tasks with repetitions of stimuli may tap into different processing levels than on-line lexical processing.

Collectively, the modifications introduced by White & Morgan served to increase the power of the mispronunciation task. Whereas earlier studies (Bailey & Plunkett, 2002; Swingley & Aslin, 2002) had failed to find graded effects depending on the numbers of features altered in mispronunciations, White & Morgan showed a strong linear effect of the number of features altered in onset changes on toddlers' looking times. Similarly, Chapters 2 of this thesis showed graded effects as well for consonant codas as well as suprasegmental tones. Here, I use White & Morgan's procedures to investigate possible asymmetries in word recognition.

EXPERIMENT 3

Experiments 3a and 3b were designed to establish whether 19-month-olds exhibit asymmetrical sensitivities to mispronunciations depending on whether the mispronunciations are from coronal to noncoronal segments or vice-versa. To address this question, Experiment 3a tested mispronunciations in consonantal onsets and Experiment 3b tested mispronunciations in codas, respectively.

Methods

Participants

Fifty-five 19-month-old English-learning infants were recruited from monolingual English-speaking families in Rhode Island and Massachusetts. Twenty-seven of these infants were tested on onset mispronunciations in Experiment 1a and twenty-eight were tested on coda mispronunciations in Experiment 1b. Nine participants were excluded from data analysis due to fussiness (5) or crying (3) or over 50% alternative language exposure (1). This left twenty 19-month-olds (mean age = 1;7;17) in Experiment 1a and twenty-six 19-month-olds (mean age = 1;7;23) in Experiment 1b, respectively.

Stimuli

Familiar labels were taken from a set of words that are comprehended by the majority of infants by 14 months according to the MacArthur CDI norms (Dale & Fenson, 1996). In each trial, infants saw two images, one depicting a referent of a familiar label, the other depicting a referent of an unfamiliar label (to 19-month-olds). An example stimulus pair is depicted in Figure 5; a list of displayed objects is given in Appendix E for onset mispronunciations and in Appendix F for coda mispronunciations.

Infants were tested in 18 trials. These included three trials with correctly pronounced coronal stops (e.g. onsets in Experiment 3a *dog* or codas in Experiment 3b *cat*), three mispronounced coronal stops (e.g. *gog* or *cak*), three correctly pronounced noncoronal stops (e.g. *cat* or *dog*), three mispronounced noncoronal stops (e.g. *tat* or *dod*), three correctly pronounced familiar fillers (e.g. *hand*) and three novel fillers (e.g. *wrench*). Examples of (mis)pronunciations are given in Table 4 for onsets and Table 5 for codas. Mispronunciations involved only single-feature changes in place of articulation, and all mispronunciations resulted in non-words or in words judged unlikely to be familiar to infants at this age. Most of the words tested in the current study were monosyllabic, with only two exceptions (*baby* and *table*) for onsets in Experiment 3a due to the restricted vocabulary size of babies at 19 months. All stimuli were naturally produced by a trained female speaker of American English who produced the utterances with positive infant-directed affect. The mean length of target items across all conditions was 588 msec (SD = 79 msec); lengths of target items did not differ significantly across conditions, $F(3, 92) = 0.036$, $P = 0.991$. A complete list of words used in the current study and their accompanying images are provided in Appendix E for onsets and Appendix F for codas.



Figure 5. Sample visual stimulus pair in Experiments 3a & 3b.

Table 4
Sample stimuli for onset mispronunciations

Audio Stimuli Samples	(Mis)pronunciation	Place of articulation
<i>Where's the duck?</i>	Correct	Coronal
<i>Where's the guck?</i>	Mispronounced	Noncoronal
<i>Where's the cat?</i>	Correct	Noncoronal
<i>Where's the tat?</i>	Mispronounced	Coronal

Table 5
Sample stimuli for coda mispronunciations

Audio Stimuli Samples	(Mis)pronunciation	Place of articulation
<i>Where's the cat?</i>	Correct	Coronal
<i>Where's the cak?</i>	Mispronounced	Noncoronal
<i>Where's the duck?</i>	Correct	Noncoronal
<i>Where's the dut?</i>	Mispronounced	Coronal

Pairings of familiar and unfamiliar objects remained constant across subjects. Assignment of stimulus pairs to pronunciation condition was counterbalanced across subjects (filler trials were constant across subjects). Half of the coronal items were mispronounced with a corresponding labial stop, and half were mispronounced with a velar stop. These were counterbalanced so that one sub-group heard two labial mispronunciations and one velar mispronunciation, while the other heard one labial mispronunciation and two velar mispronunciations. Similarly, the noncoronal items were evenly divided between labials and velars; one sub-group heard one correct labial pronunciation and two correct velar pronunciations, while the other heard two correct

labial pronunciations and one correct velar pronunciation. Each item was presented to each infant one time only, i.e. either the correct form or the mispronounced form. Order of presentation was pseudo randomized for each infant with the constraint that the first two trials always contained one familiar filler trial (e.g. car) and one novel filler trial (e.g. pump).

Procedure

Testing was conducted in a sound-treated laboratory room. The parent sat with the child on his/her lap, while listening to instrumental music over noise-cancellation headphones to mask the audio stimuli. Approximately 90cm in front of the child were two 51cm flat-panel monitors mounted side-by-side, together subtending approximately 55 degrees of visual angle. A speaker was located centrally between the two monitors behind a pegboard panel. At the subjects' eye level, a blue light was mounted on the panel between the two monitors. The subjects were observed over a closed-circuit video system and recorded on a digital camcorder at 30 fps for later off-line coding. Speech stimuli were played at conversational level (70 dB).

An intermodal preferential looking procedure (IPLP) similar as that used in White & Morgan (2008) was used during the experiment. Each trial began with the blue light flashing until the subject fixated at midline. At that point, the experimenter turned off the center light and initiated the salience phase. During the salience phase, one object with a known label and a second object with unknown label were simultaneously displayed on the two monitors (see *Figure 5*). The two objects were displayed silently for 4 seconds to establish baseline looking preference. After the salience phase, the two monitors went dark, and infants' attention was then recaptured to midline to avoid contingencies

between side of fixation at the end of the salience period and at the beginning of the test period.

After recentralization, the experimenter initiated the test phase. During the test phase, the audio stimulus (*Where's the X?*) was played, and immediately after that, the two visual stimuli were presented for 8 seconds. Following an interval of at least 1 s, the next trial commenced. Side of presentation of the familiar object was randomized between trials by the customized experimental software, but was consistent across salience and test phases within each single trial. The dependent measure was the change in proportional looking to the familiar object between the (silent) salience phase and the test phase. Of interest was whether looking behavior would differ as a function of the directions of mispronunciation.

Following the session, the parent completed a questionnaire on his/her toddler's comprehension and production of the stimulus items (familiar and unfamiliar).

Results

Results from the parental questionnaire indicated the test stimuli were appropriate for this sample of 19-month-olds. On a scale of 3 (1 = only visually familiar; 2 = visually familiar and label known; 3 = visually familiar, label known and produced), items used as familiar words received an average score of 2.462 for the words used for onset mispronunciations ($SD = .242$) and 2.462 for the words used for coda mispronunciations ($SD = .273$), indicating that they were highly familiar to toddlers. Items used as unfamiliar words received an average score of 0.24 for the words used for onset mispronunciations ($SD = .235$) and 0.301 for the words used for coda mispronunciations ($SD = .208$), indicating that they were highly unfamiliar to toddlers.

Looking behavior was coded off-line frame-by-frame (1 frame = 33 msec) using in-house coding software. For the salience phase, looking behavior was coded for entire the 4 s duration of the phase. For the test phase, looking behavior was coded only for the 3 s following the onset of the first occurrence of the target word. This was done in order to include only subjects' initial response to the target word, before the onset of the second audio stimulus. For each phase, proportional looking towards each of the objects was computed over the total time the subject spent looking at both objects for that phase. Then, a difference score of proportional looking was computed using the following formula:

$$\%Looking(Familiar)_{Test} - \%Looking(Familiar)_{Salience},$$

which measures the change in looking toward the familiar object after the target was named. Such difference scores allowed us to use each stimulus pair as its own baseline, controlling for differences in visual salience or inherent preference for a particular stimulus in each pairing. Overall summaries of the data for onsets in Experiment 3a and codas in Experiment 3b are shown in Figures 6 and 7, respectively.

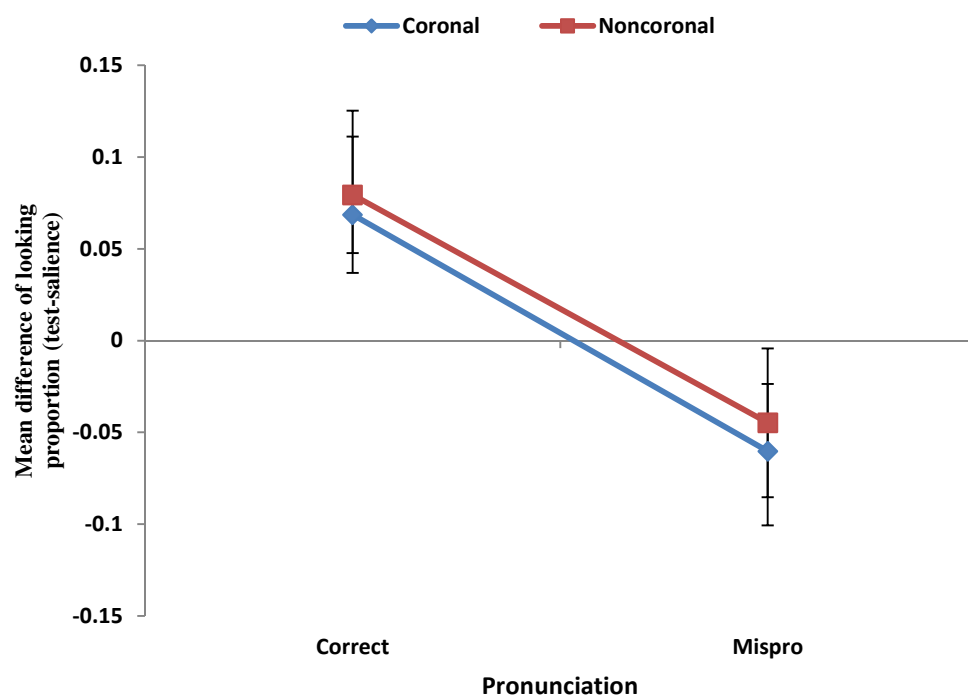


Figure 6. Experiment 3a: Toddlers' sensitivities to different directions of onset mispronunciations.

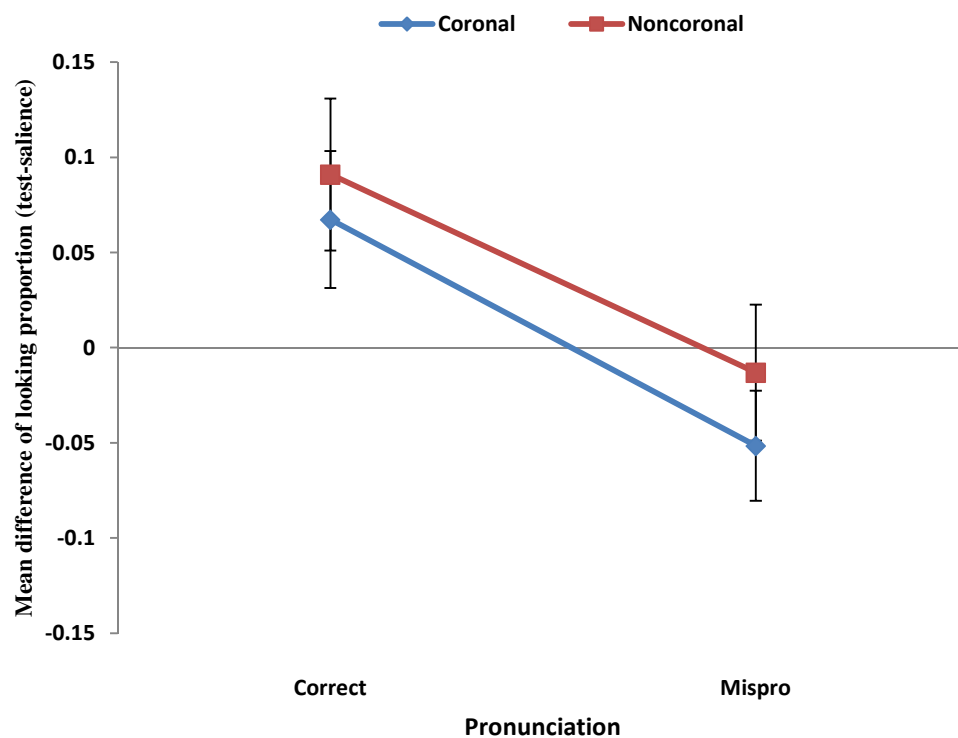


Figure 7. Experiment 3b: Toddlers' sensitivities to different directions of coda mispronunciations.

For both onsets and codas, mean proportional looking time did not differ across the two counterbalancing groups, either by subjects or by items, as shown by one-way ANOVAs. For onsets, the mean difference between the two counterbalancing groups was 0.05 (95% CI: -0.043, 0.143), $F_s(1, 19) = 0.525$, $p = 0.478$, partial $\eta^2 = 0.028$; $F_i(1, 23) = 1.247$, $p = .276$, partial $\eta^2 = 0.054$. For codas, the mean difference between the two counterbalancing groups was 0.017 (95% CI: -0.0684, 0.103), $F_s(1, 25) = 0.170$, $p = .684$, partial $\eta^2 = 0.007$; $F_i(1, 23) = .098$, $p = 0.758$, partial $\eta^2 = 0.004$. Thus, mean proportional looking times from the two counterbalancing groups are collapsed into a single group for further statistical analysis.

Two (pronunciation: correct vs. mispronunciation) by two (place of articulation: coronal vs. noncoronal) repeated measures ANOVAs were used for analyzing by-subject data of both experiments. Two by two mixed ANOVAs were used for item-based analyses, with place of articulation treated as a between-items factor. Effects of pronunciation were significant both by subjects and by items for both onset (Experiment 3a) and coda (Experiment 3b). For onsets, the mean difference between correctly pronounced words and their mispronounced counterparts was 0.137 (95% CI: 0.065, 0.210), $F_s(1, 19) = 8.817$, $p = .008$, partial $\eta^2 = 0.317$; $F_i(1, 10) = 16.019$, $p = 0.003$, partial $\eta^2 = 0.616$. For codas, the mean difference between correctly pronounced words and their mispronounced counterparts was 0.118 (95% CI: 0.054, 0.181), $F_s(1, 25) = 11.077$, $p = .003$, partial $\eta^2 = 0.307$; $F_i(1, 10) = 15.399$, $p = .003$, partial $\eta^2 = 0.606$. Therefore, 19-month-olds were sensitive to one-feature (place of articulation) mispronunciations of familiar words (see *Figure 6* for onsets and *Figure 7* for codas).

For both onsets and codas, no significant effects of place of articulation were found either by subject or by item. For onsets, the mean difference between words with coronal initial consonant and words with noncoronal initial consonants was 0.122 (95% CI: -0.075, 0.100), $F_s(1, 19) = 0.118$, $p = .734$, partial $\eta^2 = 0.006$; $F_i(1, 10) = 0.096$, $p = .763$, partial $\eta^2 = 0.010$. For codas, the mean difference between words with coronal final consonants and words with noncoronal final consonants was 0.026 (95% CI: -0.051, 0.102), $F_s(1, 25) = 1.008$, $p = .325$, partial $\eta^2 = 0.039$; $F_i(1, 10) = 0.577$, $p = .473$, partial $\eta^2 = 0.053$. Therefore, 19-month-olds in this study did not have a preference between words with coronal onsets/codas and words with noncoronal onsets/codas.

Most critically, to establish whether 19-month-olds were asymmetrically sensitive to different directions of familiar word mispronunciations, pronunciation by place of articulation interactions were analyzed. Contrary to predictions from underspecification, there was no significant interaction found in either experiment. For onsets, $F_s(1, 19) = 0.003$, $p = 0.96$, partial $\eta^2 = 0.000$; $F_i(1, 10) = 0.214$, $p = .654$, partial $\eta^2 = 0.021$. For codas, $F_s(1, 25) = 0.044$, $p = .836$, partial $\eta^2 = 0.002$; $F_i(1, 10) = 0.016$, $p = .901$, partial $\eta^2 = 0.002$. Therefore, based on the statistical analysis, it is obvious that the interactions predicted by the underspecification account did not materialize. 19-month-olds were equally sensitive to both directions of familiar word mispronunciations. Changes from coronal sound to noncoronal sound and vice versa equally impaired 19-month-olds' word recognition.

To further evaluate which theoretical account, i.e. the underspecification hypothesis and the equivalent specification hypothesis, best explains the findings above, Bayesian model comparisons are conducted to test which of these two accounts are more

probable in predicting the current patterns of results. By specifying priors that are appropriate for each account, Bayesian data analyses (Kruschke, 2010) can be computed that yield estimates of the likelihood of the observed data given each of the alternative accounts. The ratio of these likelihoods – the *Bayes factor* – provides a principled quantitative estimate of the relative credibility (i.e. posterior probability) of the two accounts.

First, priors for the main effects of (mis)pronunciation and place of articulation and the interaction between the two factors were all set as normally distributed centered on zero⁵. This is similar to null hypothesis significance testing with 2x2 repeated-measures ANOVAs. JAGS code for each of these analyses is given in Appendix G. Posterior distributions for two main effects and the interaction are shown for Experiments 3a and 3b in Figures 8 and Figure 9, respectively. For both experiments, the HDIs of the posterior distributions for the main effect of (mis)pronunciation exclude zero; in both instances, this main effect was credibly non-zero. By contrast, zero was not excluded from the HDIs of the posterior distributions for the main effect of place of articulation or the (mis)pronunciation by place of articulation interaction for either experiment: these effects are not credibly different from zero.

⁵ Standard deviations of normal distributions were set to follow a half gamma distribution with the rate parameter and the shape parameter both set as 0.2. .

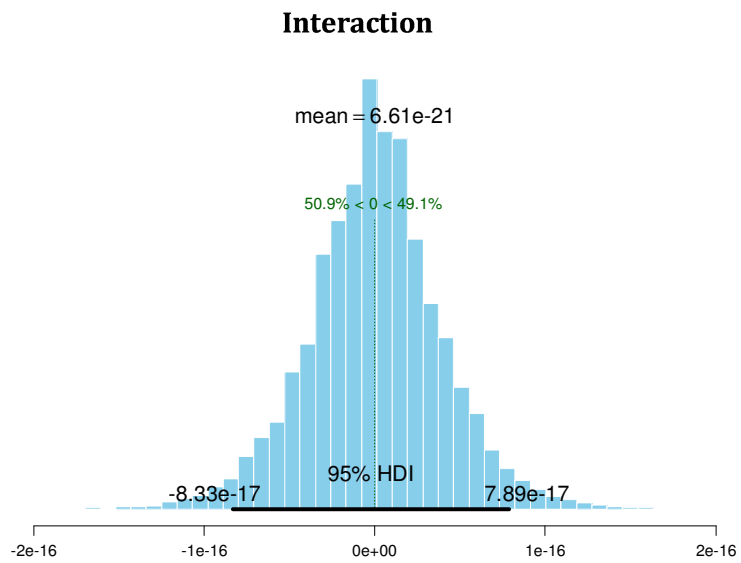
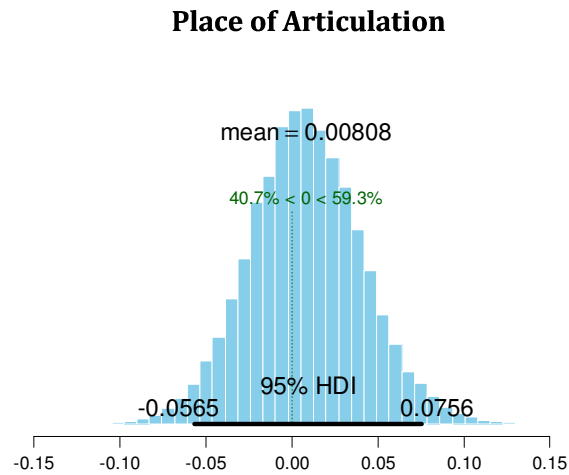
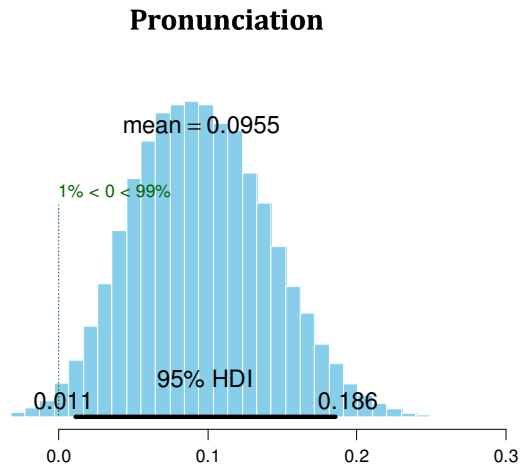


Figure 8. Planned contrasts for toddlers' proportional looking time in Experiment 3a

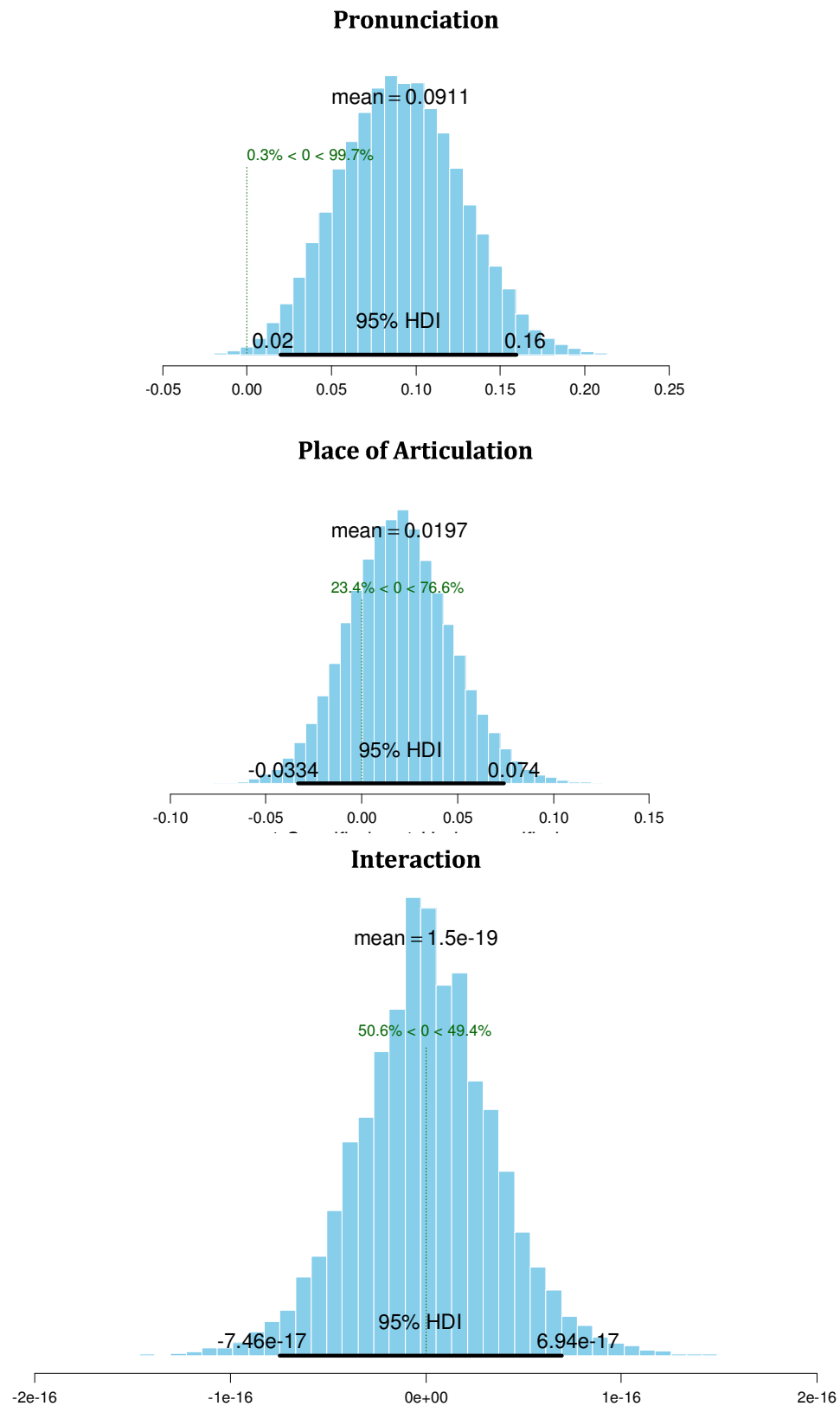


Figure 9. Planned contrasts for toddlers' proportional looking time in Experiment 3b.

Next, to compare the underspecification and equivalent-representation accounts, an informed set of priors was established according to the following set of principles: a) The marginal means for correct and incorrect pronunciations were estimated to be the same as those observed for correct pronunciations and one-feature mispronunciations in similar studies (White & Morgan, 2008 for onsets; Experiment 1, for codas). b) The cell means for correct pronunciations of items with coronal and noncoronal segments were assumed to be equal; any differences between the observed means could be attributed to differences in the lexical items chosen (e.g., word frequency). c) In the underspecification model, the cell means for correct and incorrect pronunciations of coronal segments were assumed to be equal; under this model, all of the effect of mispronunciation is assumed to reside in the noncoronal items. These assumptions allowed us to compute the expected mean for the noncoronal-mispronunciation cell as well as the place of articulation marginal means. d) In the equivalent-representation model, it was assumed that there was no (mis)pronunciation by place of articulation interaction; the expected cell means and place of articulation marginal means were computed accordingly.

Means were z-standardized, and Maximum Likelihood Estimation (MLE) was used to estimate regression coefficients. Specific values for priors in each experiment are given in Table 6.

Table 6

Priors for the standardized coefficients

	Effects	Infant Onset	Infant Coda
Model 1	Pronunciation	0.45	0.21
	Place of Articulation	0	0
	Interaction	0	0
Model 2	Pronunciation	0.45	0.21
	Place of Articulation	0.45	0.21
	Interaction	0.45	0.21

Given the onset data in Experiment 3a, the posterior likelihood of the model based on parameters derived from the underspecification account ($P(M_U|D)$) is 0.007, whereas the posterior likelihood of the model based on parameters derived from the equivalent representation account ($P(M_{ER}|D)$) is 0.993 according to the MCMC estimation. Based on Bayes' rule,

$$P(M_i|D) = \frac{P(D|M_i)P(M_i)}{\sum_{i=1}^2 P(D|M_i)P(M_i)}, \text{ where } i \text{ is the model index}$$

Therefore,

$$P(D|M_i) = \frac{(\sum_{i=1}^2 P(D|M_i)P(M_i)) P(M_i|D)}{P(M_i)} \quad (1)$$

Expanding i , we have:

$$P(D|M_U) = \frac{(\sum_{i=1}^2 P(D|M_i)P(M_i)) P(M_U|D)}{P(M_U)} \quad (2)$$

$$P(D|M_{ER}) = \frac{(\sum_{i=1}^2 P(D|M_i)P(M_i)) P(M_{ER}|D)}{P(M_{ER})} \quad (3)$$

Thus, dividing equation (3) by equation (2), we obtained the Bayes factor K :

$$K = \frac{P(D|M_{ER})}{P(D|M_U)} = \frac{P(M_{ER}|D)}{P(M_U|D)} * \frac{P(M_U)}{P(M_{ER})}$$

Assuming the prior probability of the two sets of parameters to be equal ($P(M_U) = P(M_{ER}) = 0.5$), the Bayes factor reduces to the ratio of the two posteriors, hence

$$\frac{P(D|M_{ER})}{P(D|M_U)} = \frac{P(M_{ER}|D)}{P(M_U|D)} = \frac{0.993}{0.007} = 141.86$$

A Bayes factor of this magnitude is typically interpreted as “very strong” (Kass & Raftery, 1995) or “decisive” (Jeffreys, 1961); for the data in Experiment 3a, the equivalent representation parameters provide a credible account, whereas the underspecification parameters do not.

Given the coda data in Experiment 3b, the posterior likelihood $P(M_U|D)$ is 0.331, whereas the posterior likelihood $P(M_{ER}|D)$ is 0.669. This yields a Bayes factor of 2 in favor of the equivalent representation account. A Bayes factor of this magnitude is unremarkable. In this instance, neither set of priors provides a very good fit to the data. The problem appears to be that the mispronunciation effect in Experiment 1, on which estimates of the priors were based, was substantially weaker ($\alpha_1=0.21$) than that observed here ($\alpha_1=0.298$) perhaps due to differences in infants’ familiarity with the particular items used in the two studies. If using a prior for the effect of mispronunciation based on the data observed here (an adjustment that is unorthodox but nonetheless neutral to the critical question of whether the interaction predicted by underspecification actually materialized), we would obtain posterior likelihoods of $P(M_U|D) = 0.98$ and $P(M_{ER}|D) = 0.02$, yielding a Bayes factor of 49 in favor of the equivalent representation account. A Bayes factor of this magnitude is typically interpreted as strong to very strong.

Discussion

In both Experiments 3a and 3b, 19-month-old English-learning toddlers were found *equally* sensitive to all directions of mispronunciations, no matter whether the

mispronunciation changes were from coronal to noncoronal phonemes or vice versa. Therefore, both directions of mispronunciations equally impaired 19-month-olds' word recognition. The Bayesian model comparisons also failed to provide high credibility for the underspecification hypothesis. Instead, strong Bayes factors favor the equivalent representation account. In this way, the asymmetries in effects of mispronunciations on coronal and noncoronal segments predicted by underspecification failed to materialize. Thus, findings in this chapter have thus suggested that both coronal and noncoronal features are equally specified in 19-month-olds' early lexicon.

Such findings did not provide evidence for developmental continuity. Recall that previous studies found that adults from various language backgrounds (e.g. Lahiri & Reetz, 2010) were only sensitive to noncoronal to coronal mispronunciations, suggesting that the coronal place of articulation is underspecified in their minds. However, in the current experiments, 19-month-old English-learning infants do not differentiate between different directions of mispronunciations. Therefore, it seems that both coronal and noncoronal features are equally specified in 19-month-olds' early lexicon. Thus, perhaps early lexical representations are **more specified** in terms of distinctive features for both early native speech and mature lexical representations. If this is the case, findings in the present chapter suggest a developmental **discontinuity** from infancy to toddlerhood and from toddlerhood to adulthood in the specification nature of the coronal place of articulation.

Such discontinuity seems possible. As discussed in Chapter 3, although traditional models of the adult mental lexicon assumed that all surface details that are not lexically relevant are stripped away via normalization processes (Blumstein & Stevens, 1979;

Gerstman, 1968; Shankweiler, Strange & Verbrugge, 1977; Stevens, 1972); 7.5-month-old infants, unlike adults, are unable to normalize certain surface properties (e.g. pitch change) but to rely more on surface matches (Singh, White, & Morgan, 2008). Therefore, it is likely that toddlers store more information in the lexicon than adults, rendering early lexical representations more specified than mature lexical representations.

However, before any conclusion on discontinuity is made, it worth noting that the inconsistency between our findings and those reported for infants and adults may in part be due to the use of online measurements of processing of correct and incorrect pronunciations in an intermodal preferential looking procedure. Previous studies of underspecification in adults have typically used other tasks to examine effects of mispronunciations. It is thus unclear whether the reported results reflect procedural differences or developmental differences in lexical representation between infants' early lexical representation and adults' mature lexical representation. To disentangle this, the next chapter presents three experiments with adults, which employed a visual world paradigm, mimicking the intermodal preferential looking procedure used for infants, to test adults on their responses to mispronunciations.

CHAPTER 5

UNDERSPECIFICATION IN ADULTS' LEXICAL REPRESENTATION

In the previous chapter, 19-month-old toddlers were found symmetrically sensitive to changes from coronal to noncoronal and noncoronal to coronal mispronunciations in online familiar word recognition tasks as presented using the visual world paradigm. Such a finding may suggest that toddlers have more detailed feature specification in their early lexicon than adults.

However, previous studies with findings that adults have asymmetrical sensitivities to different directions of sound changes used very different experimental methods than the online word recognition task used in Chapter 4. Some of these studies used nonce words (Cornell, Lahiri & Eulitz, 2013) rather than real words. A majority of these studies tested subjects in purely speech perception tasks without a referential context (Cornell, Lahiri & Eulitz, 2011 & 2013; Scharinger et. al., 2012; Eulitz & Lahiri, 2004). Also, most of these studies used judgment tasks such as lexical decision (Roberts, Wetterlin & Lahiri, 2013; Lahiri & Reetz, 2011; Friedrich, Eulitz, & Lahiri, 2006; Wheeldon & Wasker, 2004; Gaskell & Marslen-Wilson, 1996), priming (Lahiri & Reetz, 2011; Friedrich, Lahiri, & Eulitz, 2008), or gating (Lahiri & Marslen-Wilson, 1991). It is therefore possible that these methods tapped into a different level of representations than the visual world paradigm. For this reason, the discrepancy in findings between Chapter 4 and existing findings with adults that supported the notion of underspecification could simply manifest methodological differences.

Indeed, like Experiment 3, other studies using different modified versions of the visual world paradigm have also found symmetrical representations of coronal and

noncoronal features in adults' lexicon. In four eye-tracking experiments, Mitterer (2011), asked Dutch-speaking adult listeners to match the words they heard with the words printed on the screen by clicking the printed words. The auditory stimuli (the words they heard) and the visual stimuli (the printed words) were coronal vs noncoronal minimal pairs (e.g. *pin-tin*) or the words had an onset overlap (e.g. *peacock-teacake*). According to the notion of underspecification, mismatching auditory input (*pin*) should activate printed lexical entries with underspecified coronal stops (*tin*), but not vice versa. However, their results failed to find evidence for the predicted asymmetries. Instead, symmetrical effects of activation were found for both directions of matching. In light of such evidence, it is necessary to test adult participants using the same online word recognition task as the one used for toddlers in Chapter 4 before drawing any conclusion on developmental (dis)continuity.

The two experiments reported in the current chapter were designed to accomplish such a goal. Experiments 4a and 4b addressed this question for consonantal onsets and codas, respectively. The studies reported in this chapter differ from those of Mitterer's in several ways, most notably in using images rather than printed words and examining effects of mispronunciations. Ramifications of these differences will be discussed at the end of Experiment 4.

EXPERIMENT 4

Methods

Participants

Fifty-two mono-lingual English-speaking adults aged from 19 to 37 were

recruited from students and staff members at Brown University. Thirty-one adults were tested in Experiment 4a for onsets; and twenty-one tested in Experiment 4b for codas. Four participants were excluded from data analysis due to failure in the eye tracker calibration (3) and discomfort with the experimental settings (1). This left twenty-eight participants in Experiment 4a for onsets; and twenty participants in Experiment 4b for codas.

Stimuli

Familiar stimuli were comprised of a set of highly frequent objects with familiar names; the distractors were comprised of a set of rare or fantastical objects, with their names unlikely to be known. An example stimulus pair is depicted in Figure 10.

One hundred and eight trials were used, with eighteen correct coronal trials (e.g. duck/cat), eighteen mispronounced coronal trials (e.g. guck/cak), eighteen correct noncoronal trials (e.g. cat/duck), eighteen mispronounced noncoronal trials (e.g. tat/dut), eighteen familiar fillers (e.g. *hand*), and eighteen novel fillers (e.g. *pruse*). According to the Corpus of Contemporary American English (Davies, 2011), mean log frequencies were 3.70 for items with coronal onsets, 3.93 for items with noncoronal onsets, 3.97 for items with coronal codas, and 3.61 for items with noncoronal codas. Differences in frequencies were not significant for onset items ($t(69) = -1.56, p > .10$), but items with coronal codas occurred more frequently than items with noncoronal codas ($t(69) = 2.36, p < .05$). Because the selection of items was constrained by imageability and main effects of place of articulation were not of central interest, these differences in item frequencies are acceptable.

All stimuli were naturally produced by a trained female speaker of American English who produced the utterances with positive affect. The mean length of target items across all conditions was 564.05 msec (SD = 70.84 msec); lengths of target items did not differ significantly across conditions, $F(3, 140) = 1.036, p = 0.890$. The sets of non-filler items were counterbalanced across two subgroups: the items that one subgroup heard correctly pronounced, the other group heard mispronounced, and vice-versa. Complete lists of words used in the current studies are shown in Appendix H for onsets and Appendix I for codas; accompanying images are provided in the supplementary materials.



Figure 10. Sample visual stimulus pair in Experiments 4a & 4b.

As in Experiment 3, mispronunciations only involved changes in place of articulation. All mispronunciations resulted in non-words in English. Following the rationale described in Experiments 3, each item was presented to each participant one time only.

Procedure

The procedure was similar to that used in Experiments 3a and 3b. The carrier phrase “*Where’s the*” was eliminated from the test phase; participants heard only the isolated label. Participants’ eye-movements were recorded by an ASL eye-tracker. Regions of interest were defined based on the locations of the two images shown in each trial and were used to automatically code the eye-tracker output.

Results

The dependent measure was participants' looking proportion to the target object. Overall summaries of the data for onsets in Experiment 4a and codas in Experiment 4b are shown in Figures 11 and 12, respectively. It is obvious from inspection of the graphs that the underspecification account was again not well supported.

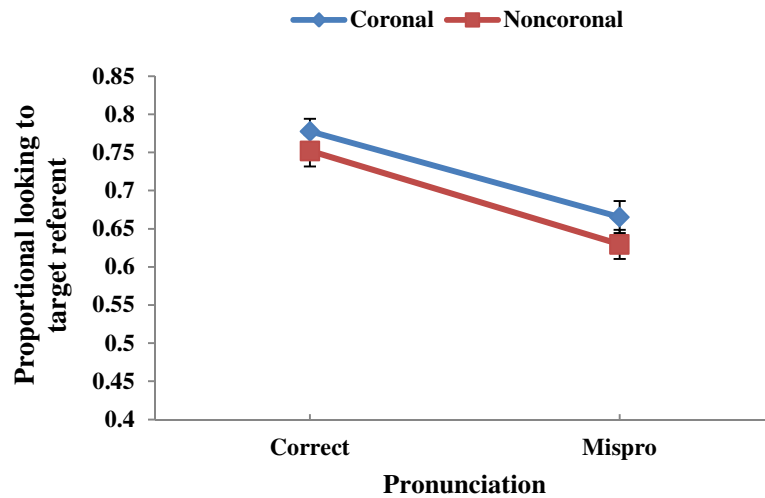


Figure 11. Experiment 4a: Adults' sensitivities to different directions of onset mispronunciations.

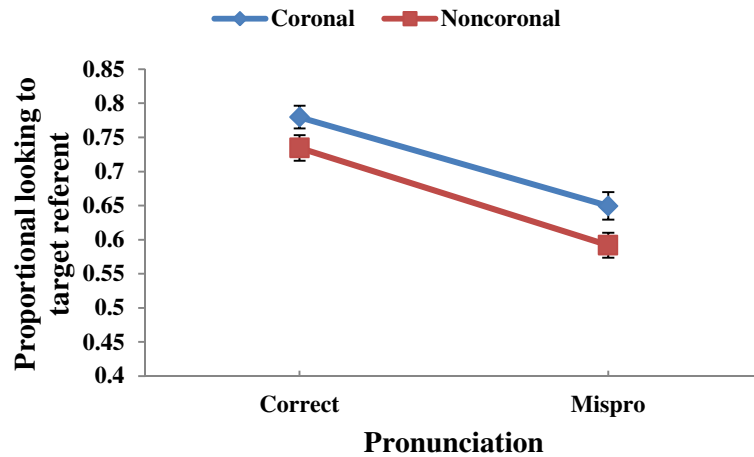


Figure 12. Experiment 4b: Adults' sensitivities to different directions of coda mispronunciations.

Like the infant data, mean proportional looking times did not differ across the two counterbalancing groups in both experiments either by subject or by item, as shown by one-way ANOVAs. For onsets, the mean difference between the two counterbalancing groups was 0.0104 (95% CI: -0.053, 0.073), $F_s(1, 27) = 0.116$, $p = 0.737$, partial $\eta^2 = 0.004$; $F_i(1, 71) = 1.490$, $p = 0.226$, partial $\eta^2 = 0.021$. For codas, the mean difference between the two counterbalancing groups was 0.024 (95% CI: -0.020, 0.067); $F_s(1, 19) = .507$, $p = 0.486$, partial $\eta^2 = 0.027$; $F_i(1, 71) = 1.180$, $p = .281$, partial $\eta^2 = 0.017$. Thus for each experiment we collapsed the mean proportional looking times from the two counterbalancing groups into a single group for further statistical analysis.

The same ANOVAs as those used in Experiment 3 were applied to the data from Experiment 4. Significant effects of pronunciation were found for both onsets and codas and by both subjects and items. For onsets, the mean difference between correctly pronounced words and their mispronounced counterparts was 0.123 (95% CI: 0.108, 0.127), $F_s(1, 27) = 57.217$, $p < .001$, partial $\eta^2 = 0.679$, $F_i(1, 70) = 93.568$, $p < .001$, partial

$\eta^2=0.579$. For codas, the mean difference between correctly pronounced words and their mispronounced counterparts was 0.136 (95% CI: 0.106, 0.167), $F_s(1,19)=10.715$, $p=.004$, partial $\eta^2=0.361$; $F_i(1,70)=79.80$, $p<.001$, partial $\eta^2=0.533$ (see *Figure 11* for onsets and *Figure 12* for codas).

Unlike Experiment 3, partially significant effects for place of articulation were found in Experiment 4. For onsets, the mean difference between words with coronal initial consonant and words with noncoronal initial consonants was 0.029 (95% CI: 0.001, 0.057), $F_s(1, 27) = 5.055$, $p = .033$, partial $\eta^2=0.158$; $F_i(1, 70) = 4.204$, $p = 0.044$, partial $\eta^2=0.057$. For codas, the mean difference between words with coronal codas and words with noncoronal codas was 0.051 (95% CI: 0.009, 0.093), $F_s(1, 19) = 4.151$, $p = .056$, partial $\eta^2=0.179$; $F_i(1, 70) = 5.952$, $p = .017$, partial $\eta^2=0.078$. These weak effects were not confirmed in the Bayesian analyses presented below.

Again, the critical question was whether adults were asymmetrically sensitive to different directions of familiar word mispronunciations, as would be indicated by significant pronunciation by place of articulation interactions. As in the infant data, contrary to predictions from underspecification, no significant interactions were found for onsets $F_s(1, 19) = 0.155$, $p = 0.697$, partial $\eta^2=0.006$; $F_i(1, 70) = 0.043$, $p = .836$, partial $\eta^2=0.001$, or for codas $F_s(1, 19) = 0.116$, $p = .738$, partial $\eta^2=0.006$; $F_i(1, 70) = 0.164$, $p = .686$, partial $\eta^2=0.002$. Like 19-month-olds, adults were equally sensitive to both directions of familiar word mispronunciations.

As in Experiment 3, Bayesian statistical analyses were conducted to evaluate whether the underspecification hypothesis or the equivalent specified hypothesis better accounts for the adult data. First, an analysis was conducted in which priors for the main

effects of (mis)pronunciation and place of articulation and the interaction between the two factors were all set to zero. This is similar to null hypothesis significance testing with 2x2 repeated-measures ANOVAs. JAGS code for each of these analyses is given in Appendix J. Posterior distributions for comparisons testing the two main effects and the interaction are shown for Experiments 4a and 4b in Figures 13 and 14, respectively. For both experiments, the HDIs of the posterior distributions for the main effect of (mis)pronunciation exclude zero; in both instances, this main effect was credibly non-zero. By contrast, zero was not excluded from the HDIs of the posterior distributions for the (mis)pronunciation by place of articulation interaction for either experiment: these effects are not credibly different from zero.

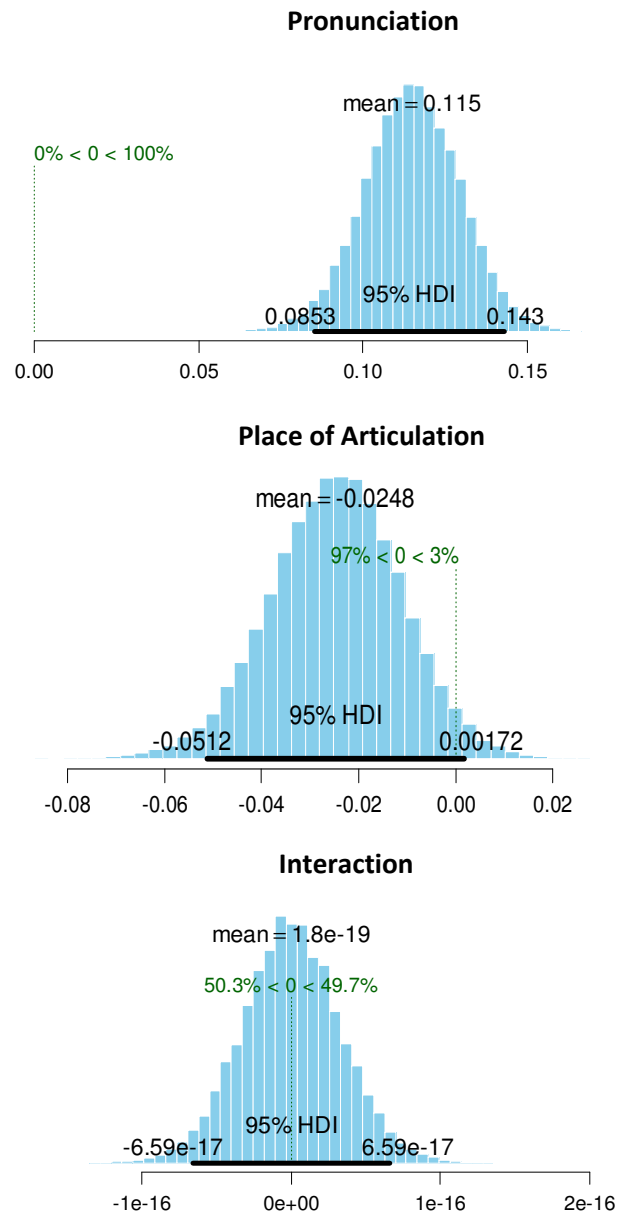


Figure 13. Planned contrasts for adults' proportional looking time in Experiment 4a.

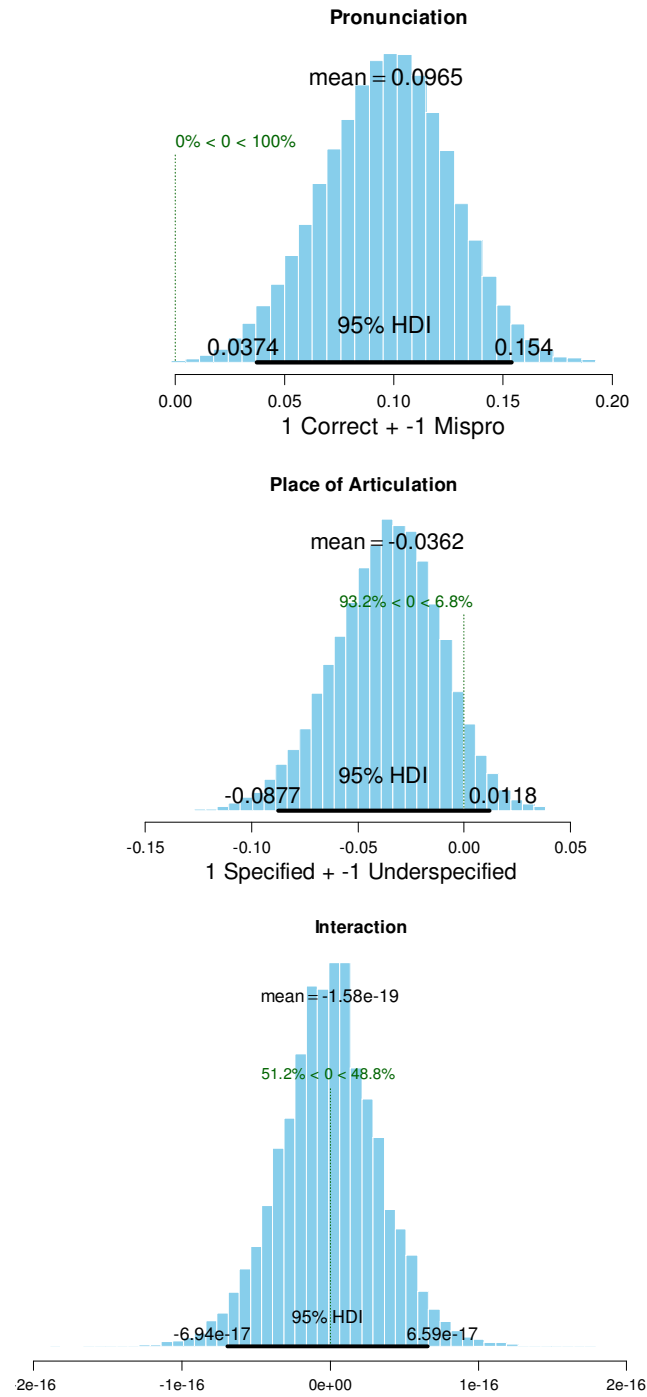


Figure 14. Planned contrasts for adults' proportional looking time in Experiment 4b.

To compare the underspecification and equivalent-representation accounts, informed sets of priors were established according to the same set of principles as in Experiment 3. Data from Tin, White, & Morgan (2014) was used to establish the priors for both onset and coda mispronunciations in Experiments 4a and 4b. Means were z-standardized, and Maximum Likelihood Estimation (MLE) was used to estimate regression coefficients. Specific values are given in Table 7.

Table 7

Priors for the standardized coefficients used in Bayesian model comparisons

	Effects	Adult Onset	Adult Coda
Model 1	Pronunciation	0.46	0.46
	Place of Articulation	0	0
	Interaction	0	0
Model 2	Pronunciation	0.46	0.46
	Place of Articulation	0.46	0.46
	Interaction	0.46	0.46

Given the onset data in Experiment 4a, the posterior likelihood of the model based on parameters derived from the underspecification account ($P(M_U|D)$) is 0.0002, whereas the posterior likelihood of the model based on parameters derived from the equivalent representation account ($P(M_{ER}|D)$) is 0.9998 according to the MCMC estimation. As in Experiment 3, the prior probability of the two sets of parameters was assumed to be equal. Thus, based on Bayes' rule, and using the same derivation procedure as in Experiment 3a, the Bayes factor reduces to the ratio of the two posteriors, hence:

$$K = \frac{P(D|M1)}{P(D|M2)} = \frac{P(M_1|D)}{P(M_2|D)} * \frac{P(M_2)}{P(M_1)} = \frac{0.9998}{0.0002} = 4999$$

A Bayes factor of this magnitude is typically interpreted as “decisive” (Jeffreys, 1961) or “very strong” (Kass & Raftery, 1995).

Given the coda data in Experiment 4b, the posterior likelihood of the model based on parameters derived from the underspecification account ($P(M_U|D)$) is 0.002, whereas the posterior likelihood of the model based on parameters derived from the equivalent representation account ($P(M_{ER}|D)$) is 0.998 according to the MCMC estimation. Thus,

$$K = \frac{P(D|M_1)}{P(D|M_2)} = \frac{P(M_1|D)}{P(M_2|D)} * \frac{P(M_2)}{P(M_1)} = \frac{0.998}{0.002} = 499$$

A Bayes factor of this magnitude is typically interpreted as “decisive” (Jeffreys, 1961) or “very strong” (Kass & Raftery, 1995). Therefore, for the data in both Experiments 4a and 4b, the equivalent representation parameters provide credible accounts, whereas the underspecification parameters do not.

Discussion

Like 19-month-olds, when assessed in a task that measured online processing, adults were sensitive to 1-feature mispronunciations, but they were NOT sensitive to the direction in which those mispronunciations took place. These findings suggest that for purposes of immediate processing, coronal and noncoronal onsets and codas are equally represented with phonological details at the lexical level for adults, contrary to predictions from the underspecification hypothesis.

However, does the visual world procedure used in Experiments 3 and 4 tap into underlying lexical representations? In the procedure used for Experiment 4, subjects saw images of possible referents in the salience phase before they heard any label in the test phase; both toddlers (Mani & Plunkett, 2010) and adults (Meyer, Belke, Telling, & Humphreys, 2007) can generate implicit labels upon seeing referents with known names.

The task in Experiments 3 and 4 could thus be conceived as recruiting “a phono-lexical matching process between the heard label and the [implicit] label for the perceived object [that determines] the goodness-of-fit” (Mayor & Plunkett, 2014, p. 93). This suggests an alternative explanation for the failure to find asymmetries: if the implicit labels are fully encoded surface forms with all values filled in for any features that are unspecified in underlying representations, the task amounts to asking subjects to compare one surface form to another, and no asymmetries would be predicted on any account.

This concern applies even more strongly to Mitterer’s (2011) studies, which used printed words as referents rather than images. Although printed referents circumvent limitations of picturability and allow a wider choice of stimulus words (McQueen & Viebahn, 2007), use of such referents introduces additional interpretive difficulties. First, orthographic similarities may influence competitor activations (Salverda & Tanenhaus, 2010). In Mitterer’s experiments, competitors differed from targets in the initial letter (and onset phone), but then shared several subsequent letters. Thus competition may have occurred at an orthographic level in addition to or even in place of phonological competition.

Salverda and Tanenhaus (2010) argue that “in the absence of sufficient preview time to read the printed words, orthographic rather than phonological similarity will mediate the mapping between an unfolding spoken word and potential printed referents” (p. 1109). To counteract unwanted orthographic effects, Mitterer allowed subjects additional time to read the printed words before hearing the spoken stimuli (personal communication, 7/26/2014). However, while increased preview time may have possibly eliminated orthographic effects, it also provides another alternative account for Mitterer’s

findings: with additional preview time, subjects were more likely to have generated representations of productive pronunciations either through explicit subvocal rehearsal (Baddeley & Hitch, 1975) or via implicit access to articulatory codes for the printed stimuli (Wheat, Cornelissen, Frost, & Hansen, 2010; Klein et al., 2014). Again, if such productive phonological representations are fully specified surface forms, there will be no cause for asymmetric effects on any account.

To resolve this problem, a new version of the adult onset study was conducted in Experiment 5. The stimuli and design were the same as in Experiment 4a, except that the salience phase was eliminated so that the pictures appeared synchronously with the onset of the label at the start of each trial. In this fashion, subjects were not afforded opportunities to generate implicit pronunciations but rather could only compare the heard label with existing lexical representations to determine goodness-of-fit.

EXPERIMENT 5

Methods

Participants

Twenty-four mono-lingual English-speaking adults aged from 19 to 57 were recruited from students and staff members at Brown University.

Stimuli

The same stimuli were used as in Experiment 4a.

Procedure

The salience phase used in earlier experiments was eliminated; only the test phase as described earlier was used. Participants' eye-movements were recorded by an SMI

eye-tracker. Each trial began with a fixation point in the center of the screen. Once the participant fixated the fixation point for 500 msec, the trial began automatically: the label was played and the two pictures appeared simultaneously.

Results

The dependent measure was participants' looking proportion to the target object. The data are shown in Figure 15. It is obvious from inspection of the Figure that the underspecification account was again not well supported.

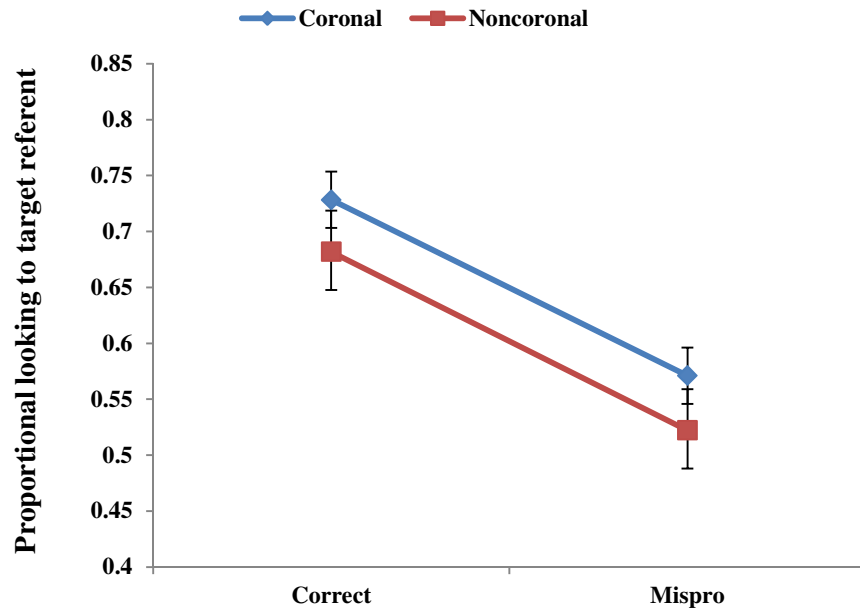


Figure 15. Experiment 5: Adults' sensitivities to different directions of onset mispronunciations when the baseline phase is removed from the task.

Like Experiment 4, mean proportional looking times did not differ across the two counterbalancing groups either by subject or by item, as shown by one-way ANOVAs. For onsets, the mean difference between the two counterbalancing groups was 0.082 (95% CI: 0.024, 0.188), $F_s(1, 23) = 2.284$, $p = 0.146$, partial $\eta^2 = 0.103$; $F_i(1, 71) = 1.175$,

$p = 0.282$, partial $\eta^2 = 0.016$. Thus we collapsed the mean proportional looking times from the two counterbalancing groups into a single group for further statistical analysis.

The same ANOVAs as those used in Experiment 4a were applied to the data from Experiment 5. Significant effects of pronunciation were found both by subjects and by items. The mean difference between correctly pronounced words and their mispronounced counterparts was 0.123 (95% CI: 0.108, 0.127), $F_s(1,23)=23.804$, $p<.001$, partial $\eta^2=0.509$, $F_i(1,70)=131.673$, $p<.001$, partial $\eta^2=0.653$.

As in Experiment 4, significant effects were found for place of articulation. The mean difference between words with coronal initial consonant and words with noncoronal initial consonants was 0.045 (95% CI: 0.018, 0.072), $F_s(1, 23) = 32.398$, $p<.001$, partial $\eta^2=0.585$; $F_i(1, 70) = 10.807$, $p = 0.002$, partial $\eta^2=0.134$. A main effect of place of articulation was also evident in the Bayesian analysis reported below.

Again, whether adults were asymmetrically sensitive to different directions of familiar word mispronunciations would be indicated by significant pronunciation by place of articulation interactions. As in the infant data and Experiment 4, contrary to predictions from underspecification, no significant interactions were found $F_s(1, 23) = 0.018$, $p = 0.894$, partial $\eta^2=0.001$; $F_i(1, 70) = 0.006$, $p = .940$, partial $\eta^2=0.000$. Adults were equally sensitive to both directions of familiar word mispronunciations even when the salient phase was removed.

As in Experiment 4, series of Bayesian statistical analyses were conducted. First, an analysis was conducted in which priors for the main effects of (mis)pronunciation and place of articulation and the interaction between the two factors were all set to zero. Second, priors were specified that were appropriate for the underspecification and

equivalent representation accounts to provide principled quantitative estimates of the relative credibility of the two accounts. JAGS code for each of these analyses is given in Appendix G.

Posterior distributions for comparisons testing the two main effects and the interaction are shown in Figure 16. The HDI of the posterior distribution for the main effect of (mis)pronunciation excluded zero; whereas the HDI of the posterior distribution for the (mis)pronunciation by place of articulation interaction did not exclude zero. As in earlier experiments, the main effect of (mis)pronunciation was credibly non-zero but the interaction was not credibly different from zero.

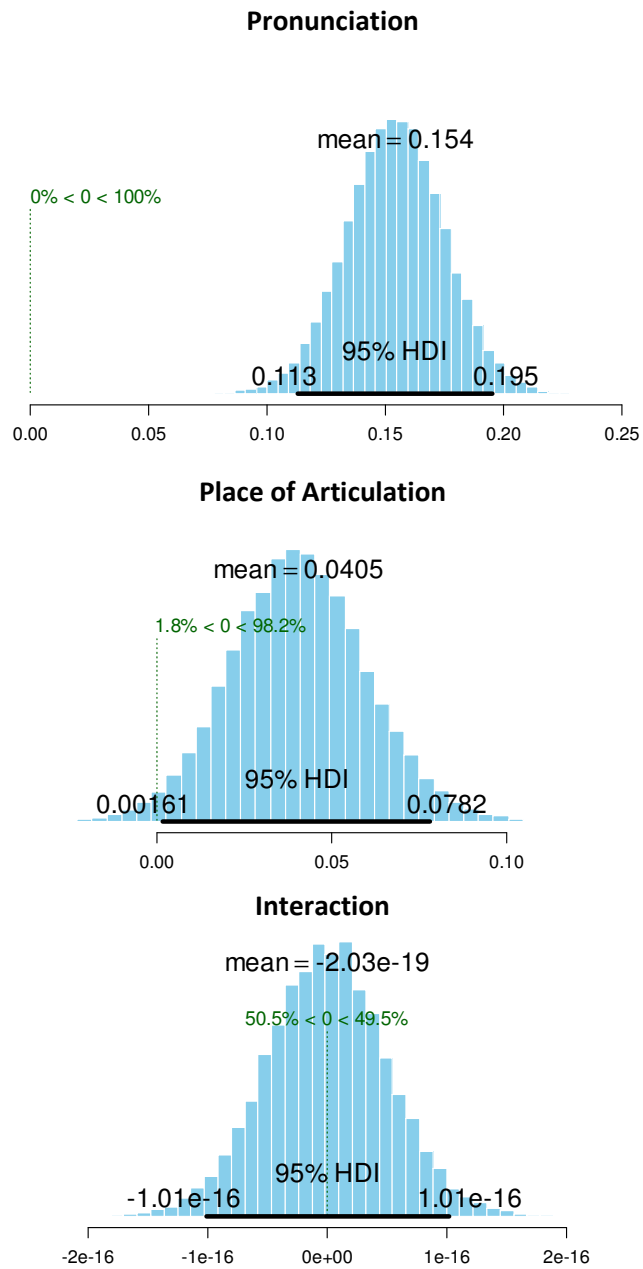


Figure 16. Planned contrasts for adults' proportional looking time in Experiment 5.

To compare the underspecification and equivalent-representation accounts, informed sets of priors were established according to the same set of principles as in Experiment 3. As in Experiment 4, data from Tin, White, & Morgan (2014) was used to establish priors for onset and coda mispronunciations. Means were z-standardized, and Maximum Likelihood Estimation (MLE) was used to estimate regression coefficients.

The posterior likelihood of the model based on parameters derived from the underspecification account ($P(M_U|D)$) is 0.04, whereas the posterior likelihood of the model based on parameters derived from the equivalent representation account ($P(M_{ER}|D)$) is 0.96 according to the MCMC estimation. As in Experiment 4, the prior probability of the two sets of parameters was assumed to be equal. Thus, the Bayes factor reduces to the ratio of the two posteriors, hence:

$$K = \frac{P(D|M_1)}{P(D|M_2)} = \frac{P(M_1|D)}{P(M_2|D)} * \frac{P(M_2)}{P(M_1)} = \frac{0.96}{0.04} = 24$$

A Bayes factor of this magnitude is typically interpreted as “very strong” (Jeffreys, 1961) or “strong” (Kass & Raftery, 1995), favoring the equivalent specification hypothesis.

Discussion

When assessed in a task that measured online processing but did not allow for implicit generation of labels for the images, adults, as in Experiment 4, were sensitive to 1-feature mispronunciations but were insensitive to the direction in which those mispronunciations took place. These results cannot be ascribed to artifactual comparisons of surface forms but rather indicate that coronal and noncoronal segments are equally specified in underlying lexical representations.

As the first attempt in investigating whether certain phonetic features are consistently (under)specified across development, Chapter 4 and Chapter 5 reported a

series of studies, two with toddlers and three with adults, using a simplified visual-world procedure that directly taps into lexical activations. The results of these experiments are uniform: for both toddlers and adults, for both onsets and codas, with and without prior exposure to images, looking to the referents of familiar words was decreased when those words were mispronounced. In no case, however, did the specification nature make any difference: whether coronals were mispronounced as noncoronals or noncoronals were mispronounced as coronals, looking time to the familiar referent decreased in much the same proportion. These results suggest that coronals and noncoronals are equally specified in both 19-month-olds' and adults' lexical representations.

These findings are clearly inconsistent with predictions from underspecification. The adult studies in the current chapter mirror those of Mitterer (2011), who tested Dutch-speaking adults. As in Experiment 4, Mitterer consistently failed to find evidence for asymmetries predicted by underspecification: looking to printed-word onset competitors differing from targets by one or two features was the same regardless of feature relations. Thus, across different languages and ages, converging results from online experiments seem to indicate that there are no differences in lexical representations of coronal versus noncoronal segments.

As noted earlier, however, a number of studies have found evidence for the asymmetries predicted by underspecification. For example, in a gating task, Bengali speakers gave few CVN responses to early gates of CVN stimuli, consistent with the claim that coronal nasal consonants are underspecified for place of articulation (Lahiri & Marslen-Wilson, 1991). Lahiri & van Coillie (1999) found that the German word for “railway”, *bahn* could still prime the word for “train” when it was mispronounced as

bahm. However, the word for “tree”, *baum*, did NOT prime the word for “bush” when it was mispronounced as *baun*. In several follow-up studies, Lahiri and colleagues replicated the same pattern of asymmetry for word onsets (Friedrich, Eulitz & Lahiri, 2008), medial positions (Friedrich, Eulitz & Lahiri, 2006), and vowels (Obleser, Lahiri & Eulitz, 2004; Lahiri & Reetz, 2010) using priming or lexical decision tasks. In addition, using the oddball paradigm, Cornell, Lahiri & Eulitz (2011, 2013) and Eulitz & Lahiri (2004) found similar patterns of asymmetry at an early stage of speech perception (around 250 milliseconds) for English consonant onsets and vowels and German vowels respectively.

Moreover, as reviewed in previous chapters, van der Feest & Fikkert (2015), using an Intermodal Preferential Looking Procedure similar to the one used in Chapter 4, showed that 20- and 24-month-old Dutch-learning toddlers respond asymmetrically to place-of-articulation mispronunciations of coronal- and labial-initial familiar words.

How can these contradictory findings be reconciled? The key, as suggested at the beginning of this chapter, may be in considering the nature of the representations engaged in the various procedures. The visual world/mispronunciation task used here exploits relations between word-forms and referents – relations that are fundamentally lexical. Below I will argue that some of the tasks that have yielded evidence in support of underspecification may engage either only pre-lexical (phonetic) representations, whereas other tasks recruit, in addition, metalinguistic representations.

As suggested by Mitterer, findings from studies that measure early stages of processing may reflect pre-lexical processing. Accordingly, stimuli might be processed at a phonetic, rather than lexical level. Recent EEG studies examining underspecification

have mostly used the oddball paradigm in which participants passively listen to stimuli changing from standard to nonstandard forms (Cornell, Lahiri & Eulitz, 2011& 2013; Eulitz & Lahiri, 2004). In these studies, participants' asymmetrical sensitivities to different directions of sound changes have usually been found at temporally very early stages. Moreover, no lexical activation is required for this sort of task: stimuli may comprise nonce items as well as existing words. Moreover, according to the PRIMR model (see Werker & Curtin, 2005, page 218), in tasks that are most like straightforward auditory processing, any discriminable differences at the General Perceptual plane will be detected. Therefore, it is very likely that tasks in such oddball paradigms have tapped onto a different level of representation than the lexicon.

Nor is it clear whether results from studies with young infants supporting underspecification can reasonably be construed as reflecting properties of *lexical* representations. Dijkstra & Fikkert (2011) tested six-month-old Dutch infants in a habituation paradigm on their sensitivities to different directions of changes between nonce items *paan* and *taan* and found discrimination of the items when infants were habituated to *paan* but not to *taan*. As discussed in Section 1.3, by six months, infants can form representations that have at least some lexical properties: they can use their own names and other highly familiar forms as anchors for segmentation (Bortfeld, Morgan, Golinkoff, & Rathbun, 2005), and they can associate some form of reference with well-known words (Bergelson & Swingley, 2012). Moreover, by eight months, infants can use syllable combinations extracted from statistical learning tasks in word-like fashions (Erickson, Thiessen, & Graf Estes, 2014; Saffran, 2001). However, there is no evidence that infants can do any of these things with the repeated, isolated forms characteristic of

habituation stimuli, or that they assign them any sort of lexical representation. The conservative interpretation, thus, is that infants treat such stimuli as phonetic sequences.

If subjects are processing stimuli solely at a phonetic level in these procedures, it is possible that the asymmetries found in these studies may reflect statistical or distributional differences between speech categories in competition. In particular, coronal sounds are statistically more frequent and more variable as a class than are noncoronal sounds. In the next chapter, I will present a mathematical model which shows that either of these properties is sufficient for producing asymmetrical judgments.

Other tasks that have been used to support underspecification, such as lexical decision and semantic priming, however, clearly do engage lexical representations. Here the critical difference between studies that have found evidence consistent with predictions from underspecification and those that have not appears to involve the number of times items presented for judgments. In both our studies and those of Mitterer (2011), items were presented for judgment one time only; the asymmetries characteristic of underspecification were not observed. Item repetition is central to some of the tasks that have supported underspecification. In ERP studies, some of which have also incorporated behavioral lexical decision tasks, items are repeated multiple times. In gating studies, auditory stimuli are presented repeatedly, with progressively less truncation, while subjects are asked to guess which word they have heard. Lahiri & Marslen-Wilson (1991) report data on 11 gates per item. Priming and lexical decision procedures do not require item repetition, but in studies using these techniques that have supported underspecification, item repetition has been the rule. For example, in Roberts, Wetterlin & Lahiri (2013) in which a delayed semantic priming task was used, each

testing item was repeated 4 times. In Friedrich, Lahiri and Eulitz (2008), Experiment 2, which used auditory priming, each testing item was also repeated 4 times. Similarly, in a study with toddlers, somewhat similar to our own, van der Feest (2007) presented target items 14 times each. Most tellingly, lexical decision tasks have consistently failed to find evidence for underspecification when there was no item repetition (e.g. Friedrich, Lahiri, & Eulitz, 2006; Experiment 1 in Friedrich, Lahiri, & Eulitz, 2008).⁶

A voluminous literature has considered effects of item repetition in memory and lexical decision tasks (e.g., Britt, Mirman, Kornilov, & Magnuson, 2014; Dutilh, Krypotos, & Wagenmakers, 2011; Forbach, Stanners, & Hochhaus, 1974; Rugg, 1990); a comprehensive review of this literature is beyond the scope of this chapter. However, item repetitions affect behavior in a variety of ways: latencies become shorter, speed of information processing increases, and responses biases may be adjusted. Judging the same items repeatedly also gives subjects additional opportunities to invoke strategies which may fundamentally change the perceived nature of the task. For example, whereas hearing *guck* once may interrogate the degree to which it activates *duck* - a perceptual task - hearing these two pronunciations multiple times might be construed as asking whether *guck* is a plausible mispronunciation of *duck*. This latter is a metalinguistic task; subjects may use their (implicit) knowledge that coronals frequently assimilate in place to preceding or following segments whereas noncoronal segments do so rarely if at all to judge that *guck* may be a plausible mispronunciation of *duck*, but *tat* is not a plausible

⁶ One possible exception might be Lahiri & Reetz (2010), who used a delayed priming task in which items were presented one time only. The authors report that mispronunciations of unspecified [+ coronal] vowels did not inhibit priming (priming = 15.37ms, $t = 1.99$; $p = .047$) whereas mispronunciations of specified vowels did inhibit priming (priming = 9.27ms, $t = 1.05$; $p > .05$). However, the authors failed to test the difference between these conditions directly, and it is doubtful whether the 6 millisecond difference was significant.

mispronunciation of *cat*. Tasks in which responses are influenced by supra-lexical, metalinguistic knowledge cannot provide decisive evidence on lexical underspecification: underspecification predicts asymmetries in judgments, but asymmetries in judgments may not all be attributable to underspecification.

Nevertheless, regardless of whether underspecification of coronal place of articulation is the mechanism in which distinctive features are structured in lexical representations, consistent findings between 19-month-old toddlers and adults suggest developmental continuity. In an online word recognition task, the most natural experimental procedure that best resembles daily lexical processing, both 19-month-olds and adults are symmetrically sensitive to different directions of familiar word mispronunciations. To this end, we can conclude that different stages of development do share similarity in the functional architecture used for information processing. Evidence from lexical processing, as shown in Chapter 4 and 5, is a stunning display for this important aspect of continuity.

CHAPTER 6

INTERPRETING UNDERSPECIFICATION USING BAYESIAN INFERENCE

This chapter presents a computational model to provide an alternative explanation for infants' and adults' asymmetrical sensitivities to different directions of sound changes as predicted by underspecification. By conceiving speech perception as a signal driven-process, the computational model shows that heterogeneity in the statistical parameters of the distribution of different speech categories can introduce asymmetrical perception. As suggested in Chapter 5, findings from studies with adults (Cornell, Lahiri & Eulitz, 2011& 2013; Eulitz & Lahiri, 2004) that measure early stages of processing and with infants (Tuiji, Mazuka, Cristia & Fikkert, 2015; Dijkstra & Fikkert, 2011) that measure responses to nonwords using a purely speech perception procedure may both reflect pre-lexical processing. If subjects are processing stimuli solely at a phonetic level in these procedures, asymmetries found in these studies may simply reflect statistical or distributional differences between speech categories under concern.

6.1 The special status of coronal sounds: more frequent and more variable

Coronal sounds are more frequent and more variable as a class than are noncoronal sounds in many languages; here I argue that either of these properties is sufficient for producing asymmetrical perception. Table 8 shows the frequency of occurrence of relevant classes of sounds in the CALLHOME database for several languages in diverse language families, including Arabic (Kilany, Gadalla, Yacoub, & McLemore, 2002), Japanese (Kobayashi, Crist, Kaneko, & McLemore, 1996), and Spanish (Garrett, Morton, & McLemore, 1996). In each of these languages, coronal segments as a class are at least twice as frequent as either labial or velar segments. The same pattern generally holds

with respect to stop consonants as a class and with respect to minimally contrastive pairs of segments. Japanese is exceptional: velar stops (in particular, /k/ and /kʲ/) occur more often than coronal stops (in particular, /t/). Thus, Japanese can provide an interesting test-case for teasing apart effects of processing at featural and segmental levels.

Table 8
Frequencies of segments in CALLHOME transcripts by place of articulation

Language	Segments	Labial	Coronal	Velar
Arabic	All consonants	91,409	222,774	94,624
	Stops	25,592	54,279	38,544
Japanese	All consonants	57,513	236,813	99,760
	Stops	15,854	62,241	79,117
Spanish	All Consonants	101,717	320,167	53,483
	Stops	44,366	62,961	43,005

The second statistical difference between coronal sounds and noncoronal sounds is that the former are more variable as a class than the latter. This is not so much due to their greater susceptibility to effects of coarticulation as to the facts that coronals appear in wider ranges of contexts and are more likely to be subject to phonological alternations than are noncoronals. When coarticulatory contexts are controlled, coronals may display less variability in place of articulation than do labials or velars (Sussman, McCaffrey, & Matthews, 1991; Sussman, Hoemeke, & McCaffrey, 1992). However, coronals typically appear in a much wider range of coarticulatory contexts than do noncoronals. In English, for example, of 45 possible onset clusters, only four (e.g., [kw] in *quick*) do not include a coronal segment, while of 73 possible monomorphemic coda clusters, only three (e.g. [nk] in *stink*) do not include a coronal segment. As we noted earlier, in Catalan, coronal nasals can assimilate in place to following stops, while velar or palatal nasals remain unchanged (Mascaro, 1976). Similarly, in English, coronal word-final obstruents may assimilate in place to following consonants, whereas noncoronal obstruents do not

(Marslen-Wilson, Nix, & Gaskell, 1995). Thus, whereas *sweet* may appear in assimilated forms (*sweep boy*, *sweek girl*), *deep* and *steak* cannot (**deet tank*, **steat dinner*).

As a first approximation, segment variability can be estimated by computing the likelihood that a given underlying segment will surface as a different segment. Cohen Priva (2012) used a modified Levenshtein distance metric to align underlying (dictionary) forms and phonetic realizations in the Buckeye Natural Speech Corpus (Pitt, Johnson, Raymond, Hume & Fosler-Lussier, 2007) and created an articulatory confusion matrix for English segments. Of the 43,915 underlying coronal stop tokens, 21,576 (49%) were pronounced as some other segment, whereas of the 64,288 underlying noncoronal stop tokens, only 2,997 (5%) were pronounced as an alternative segment, a significant difference, $t(4) = 248.83$, $p < .001$.⁷ Moreover, coronal stops have larger numbers of alternative manifestations (24 for /d/ and 19 for /t/) than do noncoronal stops (/b/, /g/, /p/, and /k/ each have 7 deviant forms). A chi-square goodness-of-fit test comparing against a theoretical expectation of 12.67 for the three voiced cells and 11 for the three unvoiced cells, i.e., testing against the hypothesis that the number of variants does not vary across different places of articulation, yields a $\chi^2(2)$ of 23.94, $p < .001$.

6.2 Interpreting perceptual asymmetry as Bayesian inference

Conceiving speech perception as a signal driven-process, Bayesian inference can be used to reconstruct the input stimulus from noise. In particular, the speaker articulates a target sound t from category c with a distribution $p(t|c)$. Then the listener hears a sound s with a probability $p(s|t)$ which is the target t passing to the listener's ears through noise. Assume

⁷ Cohen Priva also found that coronals (9% of all segments; 20% of stops) were more likely to be deleted than noncoronals (5% of all segments, 4% of stops).

that $p(t|c)$ is normally distributed with mean μ and variance σ^2 , and $p(s|t)$ is also generated from a normal distribution centered at the target with noise variance σ_s^2 .

Given a set of speech categories c and an observed sound s , the job of speech perception is to infer c from s , assuming that there are two categories under competition: one is the coronal sound c_1 and the other is the noncoronal sound c_2 ; then, according to Bayes' rule, we have the posterior probability of the observed sound s being perceived as the coronal sound c_1

$$p(c_1|s) = \frac{p(s|c_1)p(c_1)}{p(s|c_1)p(c_1) + p(s|c_2)p(c_2)} \quad (1)$$

In (1), $P(c_1)$ is the prior probability of c_1 and $P(c_2)$ is the prior probability of c_2 . The prior probabilities of the two sounds are determined by the category frequencies of these two sounds. $P(s|c_1)$ is the likelihood of observing stimulus s given it was generated by a coronal sound c_1 . Similarly, $P(s|c_2)$ is the likelihood of observing stimulus s given it was generated by a noncoronal sound c_2 . Now we rewrite - Bayes' Rule such that

$$p(c_1|s) = \frac{1}{1 + \exp\left(\log\left(\frac{p(s|c_2)}{p(s|c_1)} * \frac{P(c_2)}{P(c_1)}\right)\right)} \quad (2)$$

Let us call the prior ratio $\frac{P(c_2)}{P(c_1)}$ the *frequency bias factor*. We can easily see that the higher the value of $p(c_1)$, the more like sound s is perceived as the coronal sound c_1 . In this way, the difference in category frequency accounts for the perceptual asymmetry as predicted by underspecification. In particular, when the two categories under competition are unequal in their category frequencies of occurrence, the speech sound is more likely to be perceived as an exemplar of the category with the higher frequency. This prediction is

illustrated graphically in Figure 17 given the Gaussian assumptions made for the likelihood functions $p(s|t)$ and $p(t|c)$:

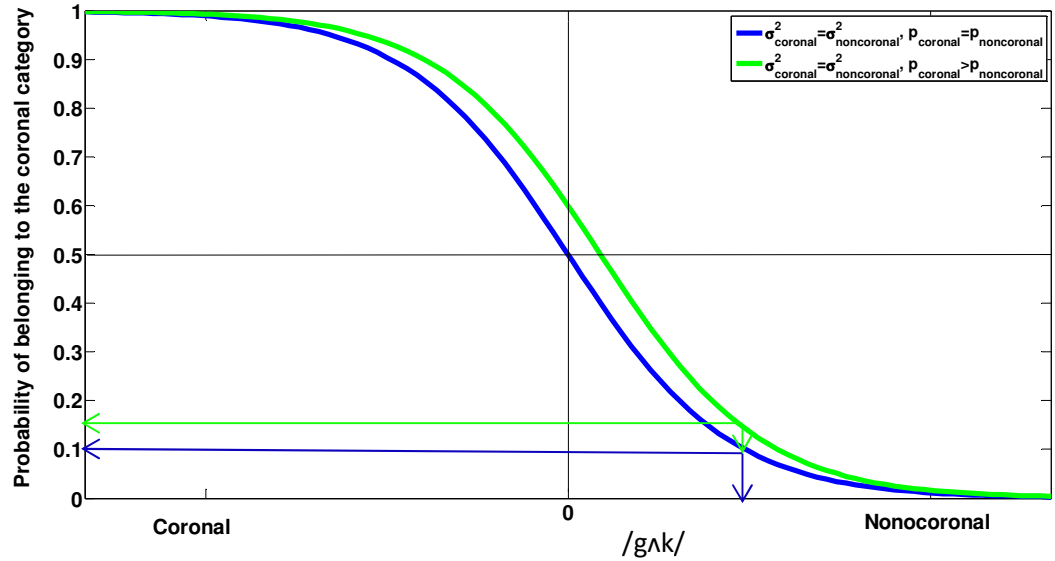


Figure 17a: Posterior Probability of the Coronal Category

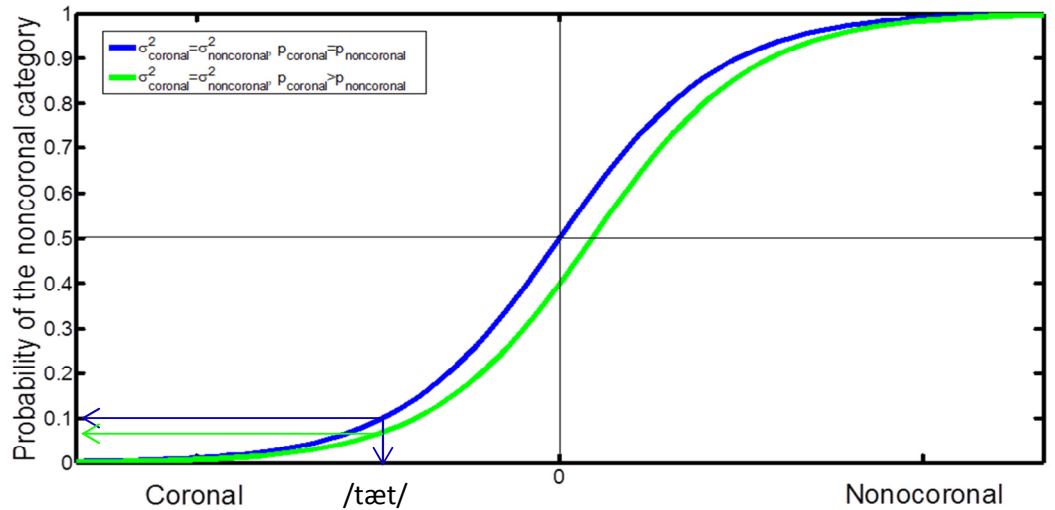


Figure 17b: Posterior Probability of the Noncoronal Category

Figure 17: Coronal category boundary shift caused by unequal category frequency.

In Figures 17a and 17b, the category boundary, as shown by the blue line, divides the perceptual space into two equal sized halves if the coronal category and the noncoronal category have equal frequencies of occurrence. However, in reality, the coronal category

is more frequent than the noncoronal category. Therefore, the category boundary shifted towards the noncoronal category, as shown by the green line, leaving more space for a sound s to be perceived as a coronal sound. For example, in Figure 17a, suppose the listener hears a sound $/g\Lambda k/$ and suppose the posterior probability of $/g\Lambda k/$ to be perceived as an exemplar of the word *duck* is 0.1 if we assume that the noncoronal phoneme $/g/$ and the coronal phoneme $/d/$ are equal in frequency of occurrence (see the blue arrows in Figure 17a). When the boundary shifts towards the noncoronal category $/g/$, the posterior probability of $/g\Lambda k/$ to be perceived as an exemplar of the word *duck* increases to around 0.15 (see the green arrows in Figure 17a). Similarly, if we assume that the noncoronal phoneme $/t/$ and the coronal phoneme $/k/$ are equal in frequency of occurrence, the listener hears a sound $/t\Lambda t/$ with the posterior probability of $/t\Lambda t/$ to be perceived as an exemplar of the word *cat* is 0.1 (see the blue arrows in Figure 17b). When the boundary shifts towards the noncoronal category $/k/$, the posterior probability of $/t\Lambda t/$ to be perceived as an exemplar of the word *cat* decreases to a score close to zero (see the green arrows in Figure 17b). In this way, by considering the frequency of occurrence for each category under competition, the model explains why a $/t\Lambda t/$ cannot be a *cat* but a $/g\Lambda k/$ can be a *duck*, the asymmetry predicted by underspecification.

The model also accounts for perceptual asymmetries that are resulted from differences in category variance. Following the Gaussian assumptions, in the model, target exemplars from each category are normally distributed with mean μ_i and variance σ_i^2

$$t|c_i \sim N(\mu_i, \sigma_i^2)$$

and the noisy stimulus s is generated from a normal distribution centered at the target with noise variance σ_s^2

$$s|t \sim N(T, \sigma_s^2)$$

Assuming that category variance is independent of the noise; then the distribution of the input sound s given the category is

$$s|c_i \sim N(\mu_i, \sigma_i^2 + \sigma_s^2) \quad (3)$$

Thus, plugging in the Gaussian density function, the likelihood function is given by (4):

$$p(s|c_i) = \frac{1}{2\pi\sqrt{\sigma_i^2 + \sigma_s^2}} \exp\left\{-\frac{(s - \mu_i)^2}{2(\sigma_i^2 + \sigma_s^2)}\right\} \quad (4)$$

(4) denotes the probability of a certain sound s being perceived given that the speaker wants to utter a sound from category c_i . Suppose there are two categories under competition: one is the coronal sound c_1 and the other is the noncoronal sound c_2 ; then the likelihood ratio of the two categories is given by (5).

$$\frac{p(s|c_2)}{p(s|c_1)} = \frac{\sqrt{\sigma_1^2 + \sigma_s^2}}{\sqrt{\sigma_2^2 + \sigma_s^2}} \exp\left\{-\frac{(s - \mu_2)^2}{2(\sigma_2^2 + \sigma_s^2)} + \frac{(s - \mu_1)^2}{2(\sigma_1^2 + \sigma_s^2)}\right\}, \quad (5)$$

which is equal to

$$\frac{\sqrt{\sigma_1^2 + \sigma_s^2}}{\sqrt{\sigma_2^2 + \sigma_s^2}} \exp\left\{-\frac{(\sigma_2^2 - \sigma_1^2)s^2 - 2s(\mu_1(\sigma_2^2 + \sigma_s^2) - \mu_2(\sigma_1^2 + \sigma_s^2)) + \mu_1^2(\sigma_2^2 + \sigma_s^2) - \mu_2^2(\sigma_1^2 + \sigma_s^2)}{2(\sigma_2^2 + \sigma_s^2) * (\sigma_1^2 + \sigma_s^2)}\right\}. \quad (6)$$

Let

$$g_1 = \frac{\sqrt{\sigma_1^2 + \sigma_s^2}}{\sqrt{\sigma_2^2 + \sigma_s^2}}$$

$$g_2 = \frac{(\sigma_2^2 - \sigma_1^2)}{2(\sigma_2^2 + \sigma_s^2) * (\sigma_1^2 + \sigma_s^2)}$$

$$g_3 = \frac{-2(\mu_1(\sigma_2^2 + \sigma_s^2) - \mu_2(\sigma_1^2 + \sigma_s^2))}{2(\sigma_2^2 + \sigma_s^2) * (\sigma_1^2 + \sigma_s^2)}$$

and

$$b = \mu_1^2 (\sigma_2^2 + \sigma_s^2) - \frac{\mu_2^2 (\sigma_1^2 + \sigma_s^2)}{2(\sigma_2^2 + \sigma_s^2)} * (\sigma_1^2 + \sigma_s^2)$$

Then

$$\frac{p(s|c_2)}{p(s|c_1)} = g_1 \exp\{g_2 s^2 + g_3 s + b\}, \quad (7)$$

Recall that the Bayes' rule was rewritten as follows in (2)

$$p(c_1|s) = \frac{1}{1 + \exp\left(\log\left(\frac{p(s|c_2)}{p(s|c_1)} * \frac{P(c_2)}{P(c_1)}\right)\right)} \quad (2)$$

Plugging (7) into (2), the posterior probability that a sound s is perceived as an exemplar of category c_1 is given by

$$p(c_1|s) = \frac{1}{1 + g_1 \exp\left\{g_2 s^2 + g_3 s + b + \log\left(\frac{p(c_2)}{p(c_1)}\right)\right\}} \quad (8)$$

When the two categories c_1 and c_2 are equal in their frequency of occurrence, the prior ratio $\frac{P(c_2)}{P(c_1)}$, i.e. the *frequency bias factor*, is reduced to 1 and the log prior ratio is zero. In this way, the posterior probability that a sound s is perceived as an exemplar of category c_1 is given by

$$p(c_1|s) = \frac{1}{1 + g_1 \exp\{g_2 s^2 + g_3 s + b\}} \quad (9)$$

When the variance term σ_1^2 of the coronal category c_1 is larger than the variance term σ_2^2 of the noncoronal category c_2 , then the category boundary is shifted away from the coronal category towards the noncoronal category, resulting in higher posterior probability for the coronal category, thus leaving more space for a sound s to be

perceived as a coronal sound. This category boundary shift is illustrated graphically in Figure 18.

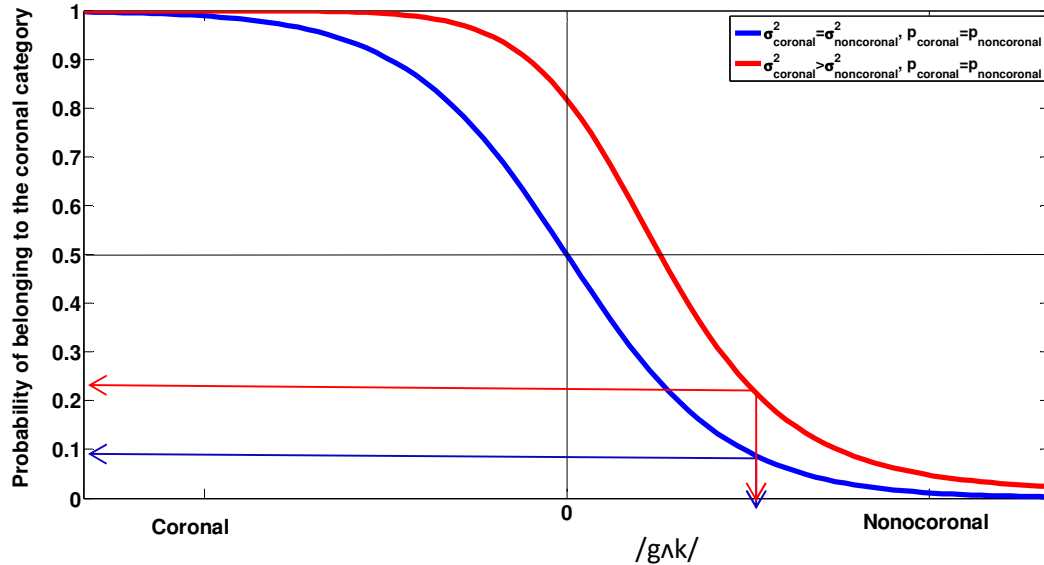


Figure 18a: Posterior Probability of the Coronal Category

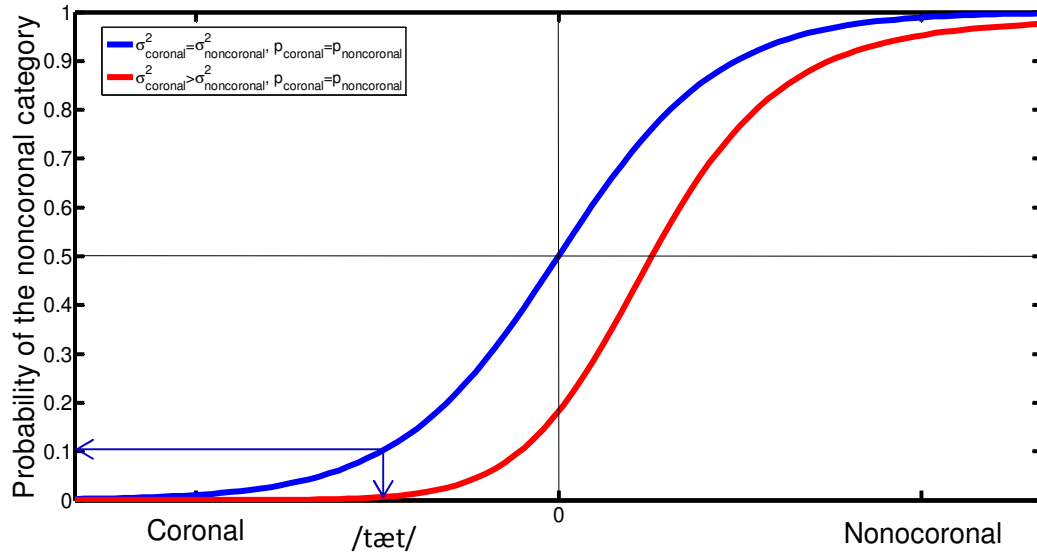


Figure 18b: Posterior Probability of the Noncoronal Category

Figure 18: Coronal category boundary stretch brought about by unequal category variance.

In Figure 18a, suppose the listener hears a sound /gʌk/ and suppose the posterior probability of /gʌk/ to be perceived as an exemplar of the word *duck* is 0.1 if we assume that the noncoronal phoneme /g/ and the coronal phoneme /d/ are equal in variance (see

the blue arrows in Figure 18a). When the boundary stretches towards the noncoronal category /g/, the posterior probability of /gʌk/ to be perceived as an exemplar of the word *duck* is increased to some score around 0.25 (see the red arrows in Figure 18a). Similarly, suppose the listener hears a sound /tæt/ and suppose the posterior probability of /tæt/ to be perceived as an exemplar of the word *cat* is 0.1 if we assume that the noncoronal phoneme /t/ and the coronal phoneme /k/ are equal in variance (see the blue arrows in Figure 18b). When the boundary stretches towards the noncoronal category /k/, the posterior probability of /tæt/ to be perceived as an exemplar of the word *cat* is decreased to some score close to zero. In this way, by considering the variability for each category under competition, the model also explains why a /tæt/ cannot be a *cat* but a /gʌk/ can be a *duck*, the asymmetry predicted by underspecification.

Now I consider the real case that the coronal category is higher both in frequency and variability than the noncoronal category. In this case, we cannot cancel out any part in equation (8).

$$p(c_1|s) = \frac{1}{1 + g_1 \exp \left\{ g_2 s^2 + g_3 s + b + \log \left(\frac{p(c_2)}{p(c_1)} \right) \right\}} \quad (8)$$

The effects of category frequency and category variability are additive, with the result that the category boundary both shifts and stretches towards the noncoronal category, leading to a higher posterior probability of the coronal category, as illustrated graphically in Figure 18.

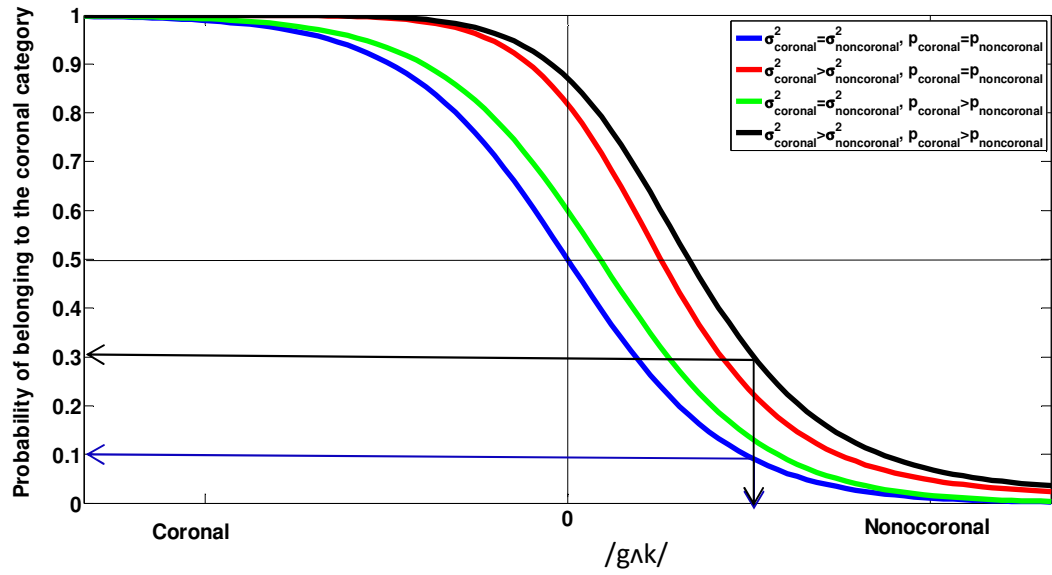


Figure 19a: Posterior Probability of the Coronal Category

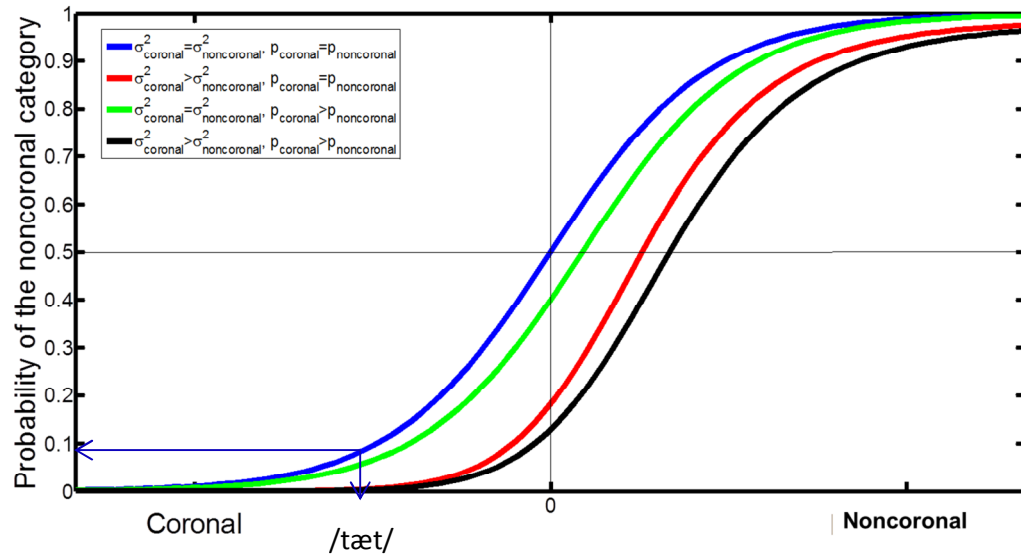


Figure 19b: Posterior Probability of the Noncoronal Category

Figure 19: Coronal category boundary stretch brought about by unequal category variance.

In Figure 19a, suppose the listener hears a sound /gΛk/ and suppose the posterior probability of /gΛk/ to be perceived as an exemplar of the word *duck* is 0.1 if we assume that the noncoronal phoneme /g/ and the coronal phoneme /d/ are equal in variance and in frequency of occurrence (see the blue arrows in Figure 19a). When the boundary stretches and shifts towards the noncoronal category /g/, the posterior probability of /gΛk/

to be perceived as an exemplar of the word *duck* is increased to a score around 0.3 (see the black arrows in Figure 19a). Similarly, suppose the listener hears a sound /tæt/ and suppose the posterior probability of /tæt/ to be perceived as an exemplar of the word *cat* is 0.1 if we assume that the noncoronal phoneme /t/ and the coronal phoneme /k/ are equal in variance and frequency of occurrence (see the blue arrows in Figure 19b). When the boundary stretches and shifts towards the noncoronal category /k/, the posterior probability of /tæt/ to be perceived as an exemplar of the word *cat* is decreased to almost zero. In this way, by considering the variability and frequency of occurrence for each category under competition, the model explains why a /tæt/ cannot be a *cat* but a /gʌk/ can be a *duck*, the asymmetry predicted by underspecification.

6.3 Discussion

In this chapter, I have argued that perceptual asymmetries found in speech perception and pre-lexical processing may be the consequence of differences in statistical properties – frequency and/or variance – of the categories under comparison. Asymmetries may thus arise from differences in the computational properties of the input speech categories.

Consequently, unlike the claim of underspecification where asymmetries are due to the unspecified coronal feature in the mental lexicon, the computational model proposed in the present chapter suggest that perceptual asymmetries can be the results of phonetic encoding rather than phonological (lexical) representations.

Also, in contrast to the domain-specific claim in the nativist tradition where linguistic principles, such as underspecification, belong to the language faculty which is proposed to be encapsulated and hence does not interact with other cognitive domains (Fodor, 1983), the Bayesian model proposed in this chapter is domain general. It applies

to any perceptual/conceptual asymmetries resembling those purportedly due to distribution factors such as those affecting underspecification. Quinn, Eimas & Rosencrantz (1993), for example, found that 4-month-olds habituated to pictures of cat faces could easily detect a change to a picture of a dog face, but infants habituated to dog faces failed to detect a change to a cat face. A series of follow-up studies investigating this asymmetry confirmed that the dog stimuli were more variable in appearance and that when variability was equated across categories, the asymmetry disappeared (Eimas, Quinn & Cowan, 1994; Mareschal, French & Quinn, 2000; Mareschal, & Quinn, 2001). Similarly, in music perception, Delbé, Bigand & French (2006) examined effects of variability by training non-musician adults with two distributions of pitch sequences differing in variability and then tested them on sensitivities to the two directions of changes. Their results indicated that changes from the less variable category to the more variable category are more detectable than vice versa. All of these findings can be explained using the current computational model.

As one of the markedness constraints, underspecification has been considered part of Universal Grammar (UG) according to the nativist approach of phonology (e.g. Prince & Smolensky, 1997). In this view, certain features such as the coronal place of articulation are underspecified because it is innately so. However, the computational model proposed in the present chapter suggests that asymmetries traditionally explained by the notion of underspecification/innateness can be **learned** from experience. In particular, exposure to native speech categories enables listeners to rely on the statistical properties of these categories, such as frequency of occurrence and variability. Therefore, rather than innate (nature), the model proposed in the present chapter suggest that certain

universal linguistic principles, such as underspecification, can be obtained by experience (nurture).

This model assumes that in pure speech perception tasks, infants and adults similarly rely on the statistical properties of the input signal in their speech categorization. Such a claim for the similarity between infants and adults' speech processing is subject to further testing. However, given the fact that infants and adults are similarly sensitive to the statistical properties of the speech input in various tasks (Saffran, Aslin & Newport, 1996 & 1998; Saffran, Johnson, Aslin & Newport, 1999), the model suggests a developmental continuity between infancy and adulthood in the functional architecture of the speech processing system.

Although statistical properties could be a source of perceptual asymmetries between coronal and noncoronal categories, recent studies in the literature have presented other alternative accounts to explain such asymmetry. The first alternative account is *phonotactic probability*. Bonte, Mitterer, Zellagui, Poelmans, and Blomert (2005) reported smaller ERPs in response to a coronal-to-labial change compared to the opposite direction, but only when the non-words containing labials had a higher phonotactic probability than those containing coronals. With opposite phonotactic probabilities, this asymmetry reversed. As discussed in Chapter 5, in a series of three eye-tracking experiments, Mitterer (2011) found no evidence for asymmetric perception consistent with underspecification. However, in their fourth experiment, the author did find an asymmetry again predicted by phonotactic probability, but not underspecification. The second alternative source could be *acoustic-phonetic properties*. To explain their findings that 4-month-olds are asymmetrically sensitive to different directions of changes between

[ɔmpa] and [ɔnta], Tsuji et. al. (2015) adopted such an idea of feature omission and reasoned that a similar process could explain asymmetries between /t/ and /p/: the burst of /t/ has a high frequency energy compared to /p/, which could lead to perceptual asymmetries. When listening to /t/ in succession to /p/, listeners get clear evidence for a change based on the added noisy /t/ burst, but when listening to /p/ after /t/, they might think that they missed the noisy burst in the second sound. Asymmetries based on such differences in acoustic characteristics have been demonstrated by Chang, Plauché, and Ohala (2001). More notably, asymmetries in infants' speech perception have played a major role in vowel discrimination research (see Polka & Bohn, 2011). Schwartz *et al.* (2005) and Polka & Bohn (2011) propose that formant convergence (or focalization) is the acoustic-phonetic pattern behind vowel perception biases. Thus, speech sounds that are more focal are *natural referent vowels* (NRV) due to their salient acoustic and articulatory properties which make them easier to detect and to hold in memory. Therefore, asymmetries in perception could be the natural sequences of an interaction of innate human biases as suggested by UG and Natural Reference Hypothesis (Polka & Bohn, 2003) and language exposure as suggested by the present model.

Recently, Nam & Polka (under review) extrapolated the NRV account to consonants. In that study, the authors examined in English-learning and French-learning 5-6-month-olds their asymmetrical sensitivities to changes between /b/ and /v/ sounds. Both sounds carry the labial feature in place of articulation but differ in manner (stop vs. fricative). In spoken English /b/ occurs more frequently than /v/ whereas in spoken French /v/ occurs more frequently than /b/. The author observed a directional asymmetry: infants noticed when /vas/ changed to /bas/ but not when /bas/ changed to /vas/. To

explain this finding, the authors suggest that this perceptual bias is a natural consequence of the acoustic-phonetic patterns generated by stop and fricative productions: The stops have a more rapid, abrupt amplitude shift, an acoustic property that the auditory nervous system. Therefore, the stops serve as a natural referent consonant in infant perception of consonant manner-of-articulation. Therefore, based on all the theoretical hypotheses reviewed above, asymmetries in perception could be the natural sequences of an interaction of innate human biases as suggested by UG and Natural Reference Hypothesis (Polka & Bohn, 2003) and language exposure as suggested by the present model.

Future studies need to test these hypotheses. For example, to test the frequency based account and the variability based account, we may need to analyze corpus evidence or test this domain general hypothesis by testing other cognitive stimuli such as faces. To test whether the NRV account applies to coronal and noncoronal asymmetry, further experimental studies are needed with both infants and adults. To further examine whether asymmetries are by nature or by nurture (or both), we need to test neonates. In conclusion, whereas underspecification may be a useful construct in some linguistic theories, the psycholinguistic status of this concept is, at best, uncertain.

GENERAL DISCUSSION

This thesis sought to answer two fundamental questions of developmental continuity: a) Does acquisition of native phonemes (i.e. speech categories) map onto word learning? This speaks to the question of continuity in early steps. b) Do infants and adults use similar architectures in lexical processing? This is the question of continuity in the long term. Using both experimental and computational methods, this thesis provides critical insight into these questions; with Part I and Part II examining the first and the second question, respectively.

In Part I, toddlers at 19 months were tested on their capacities to detect different severities of mispronunciation changes in consonant *codas* and lexical *tones*. This parallel investigation of distinctive feature representations in the early lexicon presented a unique opportunity to probe potential commonality and differences between segmental and suprasegmental dimensions in infants' early lexical representations and processing. Using such a novel approach, findings showed that at an early stage of lexical development, toddlers were able to perceive different degrees of mispronunciation changes for both consonants and tones in familiar word recognition. Therefore, it is evident that there is developmental continuity between acquisition of native speech categories during the first year of life and later word learning, and such developmental continuity occurs in both segment (consonant codas) and suprasegmental (tone) learning.

Given the central role of word learning in language acquisition, another critical question regarding developmental continuity is whether toddlers' early lexical representations resemble adults' mature lexical representations. Studies in recent years

have shown that early and mature lexical representations do include similar quantities of phonetic detail (Swingley & Aslin, 2000; White & Morgan, 2008). However, it has not previously been investigated whether the quality of phonetic details in toddlers' representations is also adult-like. As the first attempt to investigate whether certain phonetic features are consistently (under)specified across development, Part II of this dissertation has reported a series of studies, two with toddlers and three with adults, using a simplified visual-world procedure that directly and immediately taps into lexical activations. In particular, Chapter 4 and Chapter 5 tested 19-month-olds and adults on their sensitivities to different directions (coronal-to-noncoronal vs noncoronal-to-coronal) of mispronunciation changes in familiar word recognition tasks. The results of the studies reported in this part are uniform: for both toddlers and adults, for both onsets and codas, with and without prior exposure to images, looking to the referents of familiar words decreased when those words were mispronounced. In no case, however, did the nature of the mispronunciations make any difference: whether coronals were mispronounced as noncoronals or noncoronals were mispronounced as coronals, looking time to the familiar referent decreased in much the same proportion. These results suggest that coronals and noncoronals are equally specified in both 19-month-olds' and adults' lexical representations. Such findings are clearly inconsistent with predictions from underspecification-based accounts such as the Featurally Underspecified Lexicon model (Lahiri & Reetz, 2002, 2010). Most importantly, contrary to claims of phonological reorganization (Storkel, 2002; Walley, 2005), the nature of the represented detail is stable across development.

To further explore the origin of perceptual asymmetries as predicted by the notion of underspecification, Chapter 6 presented a computational model in the framework of Bayesian inference. Through investigation of the statistical parameters of the coronal category, the model revealed that asymmetries in perception can be due to statistical properties, namely the high frequency of occurrence as well as the great variability of the coronal speech category. Given the fact that both infants and adults are very sensitive to the distributional properties of speech input, it is not unreasonable that they perceive coronal categories and noncoronal categories asymmetrically in certain tasks that tap onto the level of representations where the distributional properties are analyzed. In this sense, the model predicts developmental continuity between infant speech perception and adult speech perception. Moreover, rather than putatively assuming innateness, this model challenges the fundamental principle of the nativist approach and suggests that universal linguistic constraints, such as underspecification, can be a product of learning/nurture from the speech input. In this sense, findings in this part of the thesis have not only provided critical insights into the question of developmental continuity in infants' and adults' mechanisms of lexical processing and speech perception, but also contributed to the Nature vs Nurture debate and most importantly how the two forces work together to make language acquisition and human learning possible.

Taking together all the findings, this thesis found developmental continuity for both short and long term development in both speech perception and lexical processing. Implications and questions raised by the findings of the previous chapters will be expanded upon below.

Developmental Continuity

The assessment of cognitive competence early in life and the prediction of mature cognitive stature from early performance are compelling and abiding topics of developmental studies. As reviewed in earlier chapters, classic opinions of human development have tended to favor discontinuity from infancy (Piaget, 1967). In this view, cognitive development comprises different hypothetical stages of competence; and the transition between these stages is assumed to be sudden and disconnected. Therefore, theories of discontinuity (references needed) posited that there is little or no association between infancy and adulthood in cognitive performance. However, since the 1980s, the foundations on which such discontinuity theories were originally established have gradually come under question with the advent of new assessment procedures (see Bornstein & Stigman, 1986 for a review). While the debate of (dis)continuity was about to disappear from the scene in the studies of many cognitive domains, it became a focus of debate in the study of language acquisition. In particular, as reviewed in Chapter 1, Stager & Werker (1997), for example, found that 14-month-olds could only discriminate different native phonemes (/bɪh/ vs. /dɪh/) in pure speech perception tasks which did not involve sound-meaning association. In a novel word learning task, 14-month-olds failed to differentiate /bɪh/ vs. /dɪh/ as referring to two different objects. This finding provoked a fierce dispute of developmental continuity (e.g. Swingle & Aslin, 2000; White & Morgan, 2008) vs. discontinuity (e.g. Hallé & de Boysson-Bardies, 1996) between infant speech perception and infants' early lexical representations which has lasted until today.

Findings in this thesis have significantly challenged discontinuity claims (reference). Beyond that, this thesis has further substantiated models of cognitive continuity from several important perspectives. Bornstein & Sigman (1986) proposed

three models to account for continuity in cognitive development, namely Continuity of Identical Behavior; Continuity of Underlying Process, and Continuity of Developmental Status.

Let us first consider Continuity of Identical Behavior, which states that “performance on behavior X at time 1 ought to correlate with performance on behavior X at time 2 if the identical behavior is being assessed ...using the same measurement”. The parallel results in Chapter 4 and Chapter 5 have best demonstrated this concept: both 19-month-olds and adults are symmetrically sensitive to different directions of mispronunciation changes in familiar word recognition.

Continuity of Underlying Process argues that “performance on X at time 1 ought to correlate with performance on X' at time 2 if X and X', though different behaviors, reflect the same underlying process that is continuous.” Studies reviewed in Chapter 1 and experimental findings in Chapter 2 have optimally exemplified such continuity: Although speech perception in infancy and word recognition in toddlerhood are not identical cognitive processes, infants are sensitive to their native features in a similar way as toddlers' are sensitive to different severities of mispronunciation change. Therefore, these two stages of language development share the same underlying processes, and are thus continuous. Interestingly, the speculations on the possible origins of perceptual asymmetry that have been offered in Chapter 6 may also provide a good example of Continuity of Underlying Process. In particular, although tested on different experimental tasks and assessed using different behaviors (e.g. looking time for infants vs. ERPs for adults), similar perceptual biases were found for both infants and adults (Dijkstra & Fikkert, 2011; Roberts, Wetterlin & Lahiri, 2013). Therefore, perceptual asymmetry may

reflect an age-invariant domain-general statistical learning mechanism in the human mind as suggested in the Bayesian model, or the acoustics related auditory bias as proposed by Tsuji et.al. (2015), in order to explain asymmetries in consonant discrimination found with 4-month-olds, following the framework of Gilmore, Hersh, Caramazza, & Griffin (1979).

Along these lines, this thesis has supported a significant proportion of the theory of developmental continuity. Through the investigation of early language acquisition, we obtained a broader vision of human learning and growth.

Lexical Normalization

The experimental evidence in the previous chapters has suggested that toddlers and adults process features symmetrically in *word recognition* tasks. However, studies in the literature suggested that infants and adults process features asymmetrically in *speech perception* tasks.

Similar discrepancies have also been found in other studies examining stop vs. fricative asymmetries. In a habituation (switch) task, Altvater-MacKensen & Fikkert (2010) tested Dutch-learning 14-month-olds on their discrimination between the non-word minimal pair /paap/ and /faap/. Infant subjects noticed the switch when they were habituated with /faap/, but did not notice the switch when they were habituated with /paap/. However, in a later study, Altvater-MacKensen et al. (2014) examined Dutch-learning 25-month-olds on their detection of familiar word mispronunciations using the intermodal preferential looking paradigm similar to the one used in this thesis. Their findings suggested that the 25-month-olds successfully detected both types of

mispronunciation. Thus, the ability to track this phonetic contrast appears to be fully developed by 25 months of age when recognizing familiar words. Such discrepancy between speech perception and lexical processing leads to the question: why is phonological representation symmetrical at the lexical level but asymmetrical at the phonetic level?

One possible explanation may lie in lexical normalization. As discussed in Chapter 3, traditional models of the functional architecture of the mental lexicon assume that all surface details that are not lexically relevant are stripped away via normalization processes (Blumstein & Stevens, 1979; Gerstman, 1968; Shankweiler, Strange & Verbrugge, 1977; Stevens, 1972). It is possible that surface acoustic properties of speech sounds such as abrupt/gradual shifts in amplitude and distributional properties of individual speech categories (frequency and variability) are stripped away and are not a part of the lexical representation of words. This process leaves information that is most relevant to features in the lexical representations that are used to identify lexical items. Such a pressure from the lexicon can come from knowledge of word minimal pairs. In particular, in a language such as English, there are many minimal pairs that involve coronal vs. noncoronal contrasts (e.g., pop, top, and cop). In a recent computational analysis of the CMU Pronouncing Dictionary, Cohen-Priva (personal communication, Feb 2015) suggested that out of 1507 CVC minimal pairs in English, there are 822 CVC minimal pairs in which coronal vs. non-coronal stop consonants contrast, and 748 such pairs in which two non-coronal stop consonants contrast. Overall, coronal-vs.-noncoronal minimal pairs occur more often than would be expected by chance. Facing a lexicon with so many minimal pairs involving coronal vs. noncoronal contrasts, it is possible that

listeners cannot afford to neglect the information conveyed by the coronal place of articulation in order to make word recognition successful.

Such hypotheses of lexical normalization are consistent with proposals from the PRIMR model of infant language acquisition (Werker & Curtin, 2006). According to this model, lexical representation of native speech categories (i.e. phonemes) “summarize across the context-sensitive variations that are represented as language-specific phonetic categories at the General Perceptual” level. This happens “once a criterial number of Word Forms and meaning linkages is present (Beckman & Edwards, 2000)” (see Werker & Curtin (2006), page 217-218).

As suggested in Chapter 4 and Chapter 5, both 19-month-olds and adults are equally sensitive to different directions of mispronunciation changes in word recognition tasks. Therefore, it is likely that starting from a very early age, infants may utilize this hypothesized mechanism of lexical normalization.

Future studies may investigate this lexical normalization hypothesis by testing infant and adult listeners who are learning/speaking a language in which there are significantly less coronal-vs.-non-coronal minimal pairs than noncoronal-vs.-noncoronal minimal pairs. In such languages, coronal feature would no longer be the most relevant feature in the lexical representations that are used to identify lexical items. Therefore, listeners of such languages do not have to differentiate between so many words that are coronal-vs.-non-coronal minimal pairs. Without such a pressure from the lexicon, it is likely that surface acoustic properties and distributional properties of coronal categories (frequency and variability) are preserved in the lexical representation of words. Therefore,

in word recognition tasks, we should expect that listeners of such languages show asymmetrical sensitivities to different directions of mispronunciations as predicted by underspecification.

On Nature and Nurture

The question “Where does human knowledge come from?” is not only the central issue of developmental continuity but also the fundamental problem of the *Nature vs Nurture* Debate. An answer to this question partially lies in the understanding of whether experience matters and how much it matters.

This dissertation primarily focuses on the *nurture* part of learning. In particular, it asks how infants learn native speech categories through speech input in the target language and how knowledge of native speech categories facilitates word learning. As discussed in the previous chapters, by the end of the first year of life, infants have acquired almost all native speech categories through exposure to their native language(s). Therefore, the speech architecture of the infant’s native language plays the essential role in determining how certain speech sounds are ultimately represented in the infant’s mind. In this sense, knowledge of native speech and lexical structure is obtained by nurture.

Nurture/learning is a powerful tool in obtaining knowledge. It is astounding that via learning native speech categories in the first year of life, infants can have very detailed lexical representations in terms of features such as place of articulation, manner, voicing and pitch (tone). It is also stunning that infants at 19 months have very delicate word recognition processes that could be utilized to differentiate various severities of sound deviations.

Moreover, as discussed in previous chapters, phonological theories claim that universal principles such as underspecification are the essential part of the universal grammar (see Prince & Smolensky, 1997). Therefore, based on such a claim, underspecifications are innate. However, as suggested in the computational model presented in this thesis, markedness constrains such as underspecification are not necessarily innate, instead they could also be a consequence of learning.

Given the power of nurture, the findings in the present thesis lead to the following question: what is the role of nature in the early stage of language acquisition? Although this question is obviously beyond the scope of the current thesis, it is clear that infants have an incredible predisposition to learn once they are exposed to natural language. Therefore, in future research, not only should we understand what is innate and what is learned, but most importantly know how they work jointly to shape language acquisition; because the interaction of the two forces is the real motive that drives every step of language acquisition to occur from infancy to adulthood and from features to words.

REFERENCE

- Alcina, J., & Blecua, J. M. (1975). Frecuencia de fonemas en español. *Gramática española*, 430-435.
- Altwater-Mackensen, N., van der Feest, S. V., & Fikkert, P. (2014). Asymmetries in early word recognition: the case of stops and fricatives. *Language Learning and Development*, 10(2), 149-178.
- Altwater-Mackensen, N., & Fikkert, P. (2010). The acquisition of the stop-fricative contrast in perception and production. *Lingua*, 120(8), 1898-1909.
- Anderson, J. L., Morgan, J. L., & White, K. S. (2003). A statistical basis for speech sound discrimination. *Language and Speech*, 46, 155–182.
- Archangeli, D. (1984). *Underspecification in Yawelmani Phonology and Morphology*. Ph.D Dissertation, MIT.
- Archangeli, D. (1986). The OCP and Nyangumarda buffer vowels. In *Proceedings of the North East Linguistic Society*, 16, 34-46.
- Archangeli, D. (1988). Aspects of underspecification theory. *Phonology*, 5, 183–207.
- Avery, P. & Rice, K. (1989). Segment structure and coronal underspecification. *Phonology*, 6, 179-200.
- Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 47-89.). New York: Academic Press.
- Bailey T. M., & Plunkett K. (2002). Phonological specificity in early words. *Cognitive Development*, 17, 1265–1282.
- Barton, D. (1976). Phonemic discrimination and the knowledge of words in children under three years. *Child Language Development*, 11, 61-68.
- Barton, D. (1980). Phonemic perception in children. In G. Yeni-Komshian, J. Kavanaugh, & C. Ferguson (Eds.), *Child phonology: Perception* (Vol. 2, pp. 97–116). New York: Academic Press.
- Ballem, K. D., & Plunkett, K. (2005). Phonological specificity in children at 1;2. *Journal of Child Language*, 32, 159–173.
- Bergelson, E., & Swingle, D. (2012). At 6 to 9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences of the USA*, 109, 3253-3258.

- Best, C. T., McRoberts G. W., and 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(3), 345-360.
- Bentin S., Mouchetant-Rostaing Y., Giard M.H., Echallier J.F., Pernier J. (1999). ERP manifestations of processing printed words at different psycholinguistic levels: time course and scalp distribution *Journal of Cognitive Neuroscience*, 11, 235–260.
- Blumstein S. and Stevens K. (1979). Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. *The Journal of the Acoustical Society of America*, 66(4), 1001-1017.
- Bohn, O. S., & Polka, L. (2001). Target spectral, dynamic spectral, and duration cues in infant perception of German vowels. *The Journal of the Acoustical Society of America*, 110(1), 504-515.
- Bonte, M. L., Mitterer, H., Zellagui, N., Poelmans, H., & Blomert, L. (2005). Auditory cortical tuning to statistical regularities in phonology. *Clinical Neurophysiology*, 116(12), 2765-2774.
- Bornstein, M. H., & Sigman, M. D. (1986). Continuity in mental development from infancy. *Child Development*, 251-274.
- Bortfeld, H., Morgan, J., Golinkoff, R., & Rathbun, K. (2005). Mommy and me: Familiar names help launch babies into speech stream segmentation. *Psychological Science*, 16, 298-304.
- Bosch, L., & Sebastián-Gallés, N. (2003). Language experience and the perception of a voicing contrast in fricatives: Infant and adult data. In *International congress of phonetic sciences* (eds MJ Sole, D. Recasens & J. Romero) (pp. 1987-1990).
- Bosch, L., & Sebastián-Gallés, N. (2003). Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. *Language and Speech*, 46(2-3), 217-243.
- Braun, B. & Johnson, E. K. (2011). Question or tone 2? How language experience and linguistic function guide pitch processing. *Journal of Phonetics*, 39, 585-594.
- Britt, A. E., Mirman, D., Kornilov, S. A., and Magnuson, J. S. (2014). Effect of repetition proportion on language-driven anticipatory eye movements. *Acta Psychologica*, 145, 128-138.
- Chandrasekaran, B., Krishnan. A., & Gandour J. T. (2007). Mismatch negativity to pitch contours is influenced by language experience. *Brain Research*, 1128, 148-156.

- Chang, S., Plauché, M. C., & Ohala, J. J. (2001). Markedness and consonant confusion asymmetries. *The Role of Speech Perception in Phonology*, 79-101.
- Charles-Luce, J. & Luce, P. A. (1990). Similarity neighbourhoods of words in young children's lexicons. *Journal of Child Language*, 17, 205–15.
- Charles-Luce, J. & Luce, P. A. (1995). An examination of similarity neighbourhoods in young children's receptive vocabularies. *Journal of Child Language*, 22, 727–35.
- Chao, Y. R. (1968). A grammar of spoken Chinese. Berkeley, CA: University of California Press.
- Chen, M. Y. (2000). *Tone sandhi: Patterns across Chinese dialects* (Vol. 92). Cambridge University Press.
- Chen, A., Liu, L., & Kager, R. (2015). Cross-linguistic perception of Mandarin tone sandhi. *Language Sciences*, 48, 62-69.
- Chomsky & Halle (1968) *The Sound Pattern Of English*. Harper & Row.
- Chomsky (1987) *On Power and Ideology*, The Managua Lectures. Boston: South End Press and Montreal and New York: Black Rose Books, 1987.
- Clements, G. N. (1985). The geometry of phonological features. *Phonology*, 2(01), 225-252.
- Clements, G. N. (1987). Phonological feature representation and the description of intrusive stops. *Chicago Linguistic Society*, 23(2), 29-50.
- Cohen Priva, U. (2012). *Sign and signal: Deriving linguistic generalizations from information utility*. Unpublished doctoral dissertation, Stanford University.
- Connine C. M., Titone D. Deelman, T. & Blasko D. (1997). Similarity mapping in spoken word recognition. *Journal of Memory and Language*, 37, 463–480.
- Corneille, O., Goldstone, R. L., Queller, S., & Potter, T. (2006). Asymmetries in categorization, perceptual discrimination, and visual search for reference and nonreference exemplars. *Memory & Cognition*, 34, 556–67.
- Cornell, S. A., Lahiri, A., & Eulitz, C. (2013). Inequality across consonantal contrasts in speech perception: Evidence from mismatch negativity. *Journal of Experimental Psychology: Human Perception and Performance*, 39, 757-772.
- Cornell, S. A., Lahiri, A., & Eulitz, C. (2011). “What you encode is not necessarily what you store”: Evidence for sparse feature representations from mismatch negativity. *Brain Research*, 1394, 79-89.

- Crystal, D. (1970). Prosodic systems and language acquisition. *Prosodic feature analysis*, 77-90.
- Curtin, S. (2009). Young infants encode lexical stress in newly encountered words. *Journal of Experimental Child Psychology*. 105, 367-385.
- Curtin, S., Mintz, T.H., & Christiansen, M.H. (2005). Stress changes the representational landscape: evidence from word segmentation. *Cognition* 96, 233–262.
- Curtin S., (2010) Young infants encode lexical stress in newly encountered words. *Journal of Experimental Child Psychology*. 105, 376–385.
- Cutler, C. M. & Clifton, C.E. Jr. (1985). The use of prosodic information in word recognition. In H. Bouma and D.G. Bouwhuis (eds.), *Attention and Performance X* (pp. 183-196). Hillsdale, NJ: Erlbaum.
- Cutler, A. (1986). Forbear is a homophone: Lexical Prosody does not constrain lexical access. *Language and Speech*, 29, 201-221.
- Cutler, A., & Chen, H. C. (1997). Lexical tone in Cantonese spoken-word processing. *Perception and Psychophysics*, 59, 165-179.
- Cutler A., & Otake, 1999 Pitch accent in spoken-word recognition in Japanese. *The Journal of the Acoustical Society of America*. 105, 1877-1888.
- Dahan, D., & Tanenhaus, M. K. (2005). Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. *Psychonomic Bulletin & Review*, 12, 453-459.
- Dale P. S., & Fenson L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments, & Computers*, 28, 125-127.
- Davies, M. (2011). The Corpus of Contemporary American English as the first reliable monitor corpus of English. *Literary & Linguistic Computing*, 25, 447-464.
- De Bree, E., Van Alphen, P. M., Fikkert, P., & Wijnen, F. (2008). Metrical stress in comprehension and production of Dutch children at risk of dyslexia. *32nd Annual Boston University Conference on Language Development* (60-71). Cascadilla Press.
- Delbé, C., Bigand, E., French, R. M. (2006). Asymmetric categorization in the sequential auditory domain. *Proceedings of the 28th Annual Cognitive Science Society Conference*.

- 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(41), 16027-16031.
- Dijkstra N. & Fikkert P. (2011). Universal constraints on the discrimination of place of articulation? Asymmetries in the discrimination of paan and taan by 6-month-old Dutch infants. In N. Danis, K. Mesh, & H. Sung (Eds.), *Proceedings of the 35th Annual Boston University Conference on Language Development* (pp. 172-180). Somerville, MA: Cascadilla Press.
- Duanmu, S. (2007). *The Phonology of Standard Chinese*. Second Edition, Oxford University Press, New York.
- Dutilh, G., Wagenmakers, E.-J., Visser, I., & van der Maas, H. L. J. (2011). A phase transition model for the speed-accuracy trade-off in response time experiments. *Cognitive Science*, 35, 211-250.
- Echols, C. H., Crowhurst, M. J., & Childers, J. B. (1997). The perception of rhythmic units in speech by infants and adults. *Journal of memory and language*, 36(2), 202-225.
- Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, 171(3968), 303-306.
- Eimas, P., Quinn, P., & Cowan, P. (1994). Development of exclusivity in perceptually-based categories of young infants. *Journal of Experimental Child Psychology*, 58, 418-431.
- Erickson, L. C., Thiessen, E. D., & Graf Estes, K. (2014). Statistically coherent labels facilitate categorization in 8-month-olds. *Journal of Memory and Language*, 72, 49-58.
- Eilers, R. E. & Oller, D. K. (1976). The role of speech discrimination in developmental sound substitutions. *Journal of Child Language*, 3, 319-329.
- Eulitz C., & Lahiri A.(2004). Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *Journal of Cognitive Neuroscience*, 16, 577-583.
- Fear, B. D. & Cuttler, A. & Butterfield (1995). The strong/weak syllable distinction in English. *Journal of Acoustic Society of America*, 97, 1983-1904.
- Feldman, N. H., Griffiths, T. L., & Morgan, J. L. (2009). The influence of categories on perception: explaining the perceptual magnet effect as optimal statistical inference. *Psychological Review*, 116, 752-82.

- 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.
- Fennell C. T. (2007). Asymmetries in infants' ability to notice mispronunciations. Poster presented at *The Conference of the Canadian Society for Brain, Behaviour, and Cognitive Science*. June 2007, Victoria, Canada.
- Fennell C. T., van der Feest S. V. H. & Spring M. (2010). Perceptual asymmetries of consonants at 14-months: Implications for phonological acquisition. Presented at *International Conference on Infant Studies*, Baltimore, Maryland, March, 2010.
- Fennell, C. T., & Waxman, S. R. (2010). What paradox? Referential cues allow for infant use of phonetic detail in word learning. *Child development*, 81, 1376-1383.
- Fifer, W. P., and Moon C. (1988). Auditory experience in the fetus. In *Behavior of the Fetus*, ed. W. P. Smotherman & S. R. Robinson. Telford Press.
- Floccia, C., Nazzi, T., Austin, K., Arreckx, F., & Goslin, J. (2011). Lexical stress and phonetic processing in word learning in 20-to 24-month-old English-learning children. *Developmental Science*, 14(3), 602-613.
- Friedrich C. K., Eulitz C., & Lahiri A. (2006). Not every pseudo word disrupts word recognition: An ERP study. *Behavioral and Brain Functions*, 2, 36-45.
- Friedrich C. K., Lahiri A., & Eulitz C. (2008). Neurophysiological evidence for underspecified lexical representations: asymmetries with word initial variations. *Journal of Experimental Psychology: Human Perception and Performance*, 34, 1545-1559.
- Fodor, J. A. (1983). *The Modularity of Mind: An Essay on Faculty Psychology*. MIT press. Cambridge, Massachusetts.
- Forbach, G. B., Stanners, R. F., & Hochhaus, L. (1974). Repetition and practice effects in a lexical decision task. *Memory & Cognition*, 2, 337-339.
- Fu, Q. J., Zeng, F. G., Shannon, R. V., & Soli, S. D. (1998). Importance of tonal envelope cues in Chinese speech recognition. *Journal of the Acoustical Society of America*, 104, 505-510.
- Garnica, O. K. (1973). The development of phonemic speech perception. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 215-222). New York: Academic Press.
- Garrett, S. Morton, T. & McLemore C. (1996). *CALLHOME Spanish Lexicon*. Linguistic Data Consortium, Philadelphia.

- Gaskell, M. G., & Marslen-Wilson, W. D. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human perception and performance*, 22, 144-158.
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2003). *Bayesian data analysis*. Boca Raton, Florida: CRC press.
- Gerstman, L. J. (1968). Classification of self-normalized vowels. *Audio and Electroacoustics, IEEE Transactions on*, 16(1), 78-80.
- Hao, M., Shu, H., Xing, A. & Ping, L. (2008). Early vocabulary inventory for Mandarin Chinese. *Behavior Research Methods*, 40, 728-733.
- Gilmore, G. C., Hersh, H., Caramazza, A., & Griffin, J. (1979). Multidimensional letter similarity derived from recognition errors. *Perception & Psychophysics*, 25(5), 425-431.
- Goldstone, R. L. (1998). Perceptual learning. *Annual review of psychology*, 49, 585-612.
- Hallé, P. A., & de Boysson-Bardies, B. (1996). The format of representation of recognized words in infants' early receptive lexicon. *Infant Behavior and Development*, 19, 463-481.
- Harrison, P. (2000). Acquiring the phonology of lexical tone in infancy. *Lingua*, 110, 581-616.
- Herold, B., Höhle, B., Walch, E., Weber, T., & Obladen, M. (2008). Impaired word stress pattern discrimination in very-low-birthweight infants during the first 6 months of life. *Developmental Medicine & Child Neurology*, 50(9), 678-683.
- Höhle, B., Bijeljac-Babic, R., Herold, B., Weissenborn, J., & Nazzi, T. (2009). Language specific prosodic preferences during the first half year of life: Evidence from German and French infants. *Infant Behavior and Development*, 32, 262-274.
- Holt, L. L., & Lotto, A. J. (2008). Speech perception within an auditory cognitive science framework. *Current directions in psychological science*, 17, 42-46.
- Houston, D. M. & Jusczyk, P. W. (2000). The Role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human perception and performance*, 26, 1570-1582.
- Howie, J. (1976). *An acoustic study of Mandarin tones and vowels*. Cambridge, UK: Cambridge University Press.
- Hume, E. (1996). Coronal consonant, front vowel parallels in Maltese. *Natural Language & Linguistic Theory*, 14(1), 163-203.

- Hunter M. A., Ames E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli, in Rovee-Collier C., Lipsitt L. P., editors. *Advances in Infancy Research*, Vol. 5, eds. (Stamford, CT: Ablex;), 69-95.
- Ito, J. (1986). *Syllable theory in prosodic phonology*. Ph.D. dissertation, University of Massachusetts, Amherst.
- Jakobson, R. (1941). *Child Language, Aphasia and Phonological Universals*. The Hague: Mouton.
- Jeffreys, H. (1961), *Theory of Probability (3rd. ed.)*, Oxford, U.K.: Clarendon Press
- Jusczyk, P. W., & Thompson, E. (1978). Perception of a phonetic contrast in multisyllabic utterances by 2-month-old infants. *Perception & Psychophysics*, 23(2), 105-109.
- Jusczyk, P.W., Cutler, A., & Redanz, N. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development*, 64, 675–687.
- Kaplan, E., & Kaplan, G. (1971). The prelinguistic child. In J. Elliot (Ed.), *Human Development and Cognitive Processes* (pp. 359-381). New York: Holt, Rinehart & Winston.
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90, 773-795.
- Kaye, J., Lowenstamm, J., & Vergnaud, J. R. (1985). The internal structure of phonological elements: a theory of charm and government. *Phonology*, 2(01), 305-328.
- Kay-Raining Bird, E., & Chapman, R. S. (1998). Partial representations and phonological sensitivity in the comprehension of 12- to 16-month olds. *First Language*, 18, 105-107.
- Kean, M.-L.(1975). *The theory of markedness in generative grammar*. Ph.D. dissertation, MIT. Distributed by Indiana University Linguistics Club.
- Keating, P. A. (1990). Coronal places of articulation. *UCLA Working Paper in Phonetics*, 74, 35-60.
- Kilany, H. H., Gadalla, H. A., A. Yacoub, A. E.-H., & McLemore, C. (2002). *Egyptian Colloquial Arabic Lexicon*, Linguistic Data Consortium, Philadelphia.
- Klein, M., Grainger, J., Wheat, K. L., Millman, R. E., Simpson, M. I. G., Hansen, P. C., & Cornelissen, P. L. (2014). Early activity in Broca's area during reading reflects fast access to articulatory codes from print. *Cerebral Cortex*.

- Kobayashi, M., Crist, S., Kaneko, M., & McLemore, C. (1996). *CALLHOME Japanese Lexicon*. Linguistic Data Consortium, Philadelphia.
- Koster, M., & Cutler, A. (1997). Segmental and suprasegmental contributions to spoken-word recognition in dutch. In *EUROSPEECH*.
- Kruschke, J. (2010). *Doing Bayesian Data Analysis: A Tutorial Introduction with R*. Burlington: Academic Press.
- Kuhl, P. K. (1991). Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, 50(2), 93-107.
- Kuhl, P. K., Williams K. A., Lacerda F., Stevens K. N., Lindblom B. (1992). Linguistic experience alters phonetic perception in infants by 6 month of age. *Science*, 255, 606-608.
- Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews Neuroscience*, 5(11), 831-843.
- Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental science*, 9(2), F13-F21.
- Lahiri, A., & Reetz, H. (2002). Underspecified recognition. *Laboratory Phonology*, 7, 637-675.
- Lahiri, A., & Reetz, H. (2010). Distinctive features: Phonological underspecification in representation and processing. *Journal of Phonetics*, 38(1), 44-59.
- Lahiri A., & Marslen-Wilson W. D. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition*, 38, 245-294.
- Lahiri A., van Coillie S. (1999). *Non-mismatching features in language comprehension*. Unpublished manuscript, University of Konstanz.
- Lass, R. (1976). *English phonology and phonological theory*. Cambridge: Cambridge University Press.
- Levelt, C. (2009). An experimental approach to coda omissions in early child language, Poster presented at the 34th *Boston University Conference on Language Development*, Boston MA.
- Lewis, M. (1951). *Infant Speech, A Study of the Beginnings of Language*. Humanities Press. New York.

- Liu, L. & Kager, R. (2014). Perception of tones by infants learning a non-tone language. *Cognition*, 133, 385-394.
- 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 Acoustic Society of America*, 94, 1242-55.
- Lv, S. & Ding, S. (2005). *Dictionary of Contemporary Chinese*. Fourth Edition. Commercial Press. Beijing, China.
- Maddieson, I. (1984). *Patterns of sounds*. Cambridge: Cambridge University Press.
- Mani, N. & Plunkett, K. (2007). Phonological specificity of vowels and consonants in early lexical representations. *Journal of Memory and Language*, 57, 252-272.
- Mani, N., & Plunkett, K. (2010). In the infant's mind's ear evidence for implicit naming in 18-month-olds. *Psychological Science*, 21(7), 908-913.
- Mani, N., & Plunkett, K. (2011). Does size matter? Subsegmental cues to vowel mispronunciation detection. *Journal of Child Language*, 38, 606-627.
- Mareschal, D. French, R. M., & Quinn, P. C. (2000). A connectionist account of asymmetric category learning in early infancy. *Developmental Psychology*, 36, 635-645.
- Mareschal, D., & Quinn, P. C. (2001). Categorization in infancy. *Trends in cognitive sciences*, 5(10), 443-450.
- Mareschal, D., Quinn, P. C., & French, R. M. (2002). Asymmetric interference in 3- to 4-montholds' sequential category learning. *Cognitive Science*, 26, 377-389.
- Mascaro J. (1976). *Catalan phonology and the phonological cycle*. Doctoral Dissertation, MIT, Cambridge, MA.
- Marslen-Wilson W. D.(1975). The temporal structure of spoken language understanding. *Cognition*, 8, 1-71.
- Marslen-Wilson W. D., Nix A., & Gaskell M. G. (1995). Phonological variation in lexical access: Abstractness, inference and English place assimilation. *Language and Cognitive Processes*, 10, 285-308.
- Mattock, K., Molnar, M., Polka, L., & Burnham, D. (2008). The developmental course of lexical tone perception in the first year of life. *Cognition*, 106(3), 1367-1381.

- Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive psychology*, 38(4), 465-494.
- Mayor J. & Plunkett K. (2014). Infant word recognition: Insight from TRACE simulations. *Journal of Memory and Language*. 71, 89-123.
- Masuda, T., & Nisbett, R.E. (2001). Attending holistically vs. analytically: Comparing the context sensitivity of Japanese and Americans. *Journal of Personality and Social Psychology*, 81, 922-934.
- McQueen, J. M., & Viebahn, M. (2007). Tracking recognition of spoken words by tracking looks to printed words. *Quarterly Journal of Experimental Psychology*, 60, 661-671.
- Metsala, J. L., & Walley, A. C. (1998). Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability.
- Metasala, A. C. (2005). Speech perception in childhood. In D. B. Pisoni & R. E. Remez (Eds.), *Handbook of speech perception* (pp. 449-468). Oxford: Blackwell.
- Meyer, A. S., Belke, E., Telling, A. L., & Humphreys, G. W. (2007). Early activation of object names in visual search. *Psychonomic Bulletin & Review*, 14, 710-716.
- Milberg W., Blumstein S., Dworetzky B. (1988). Phonological factors in lexical access: Evidence from an auditory lexical decision task. *Bulletin of the Psychonomic Society*, 26, 305-308.
- Milliken, S. (1989). Why there is no third tone sandhi rule in standard Mandarin. *Papper presented at the Tianjin International Conference on Phonetics and Phonology*. Tianjin Normal University, China, 7-10 June.
- Minagawa-Kawai, Y., Mori, K., Naoi, N., & Kojima, S. (2007). Neural attunement processes in infants during the acquisition of a language-specific phonemic contrast. *The Journal of Neuroscience*, 27(2), 315-321.
- Mitterer, H. (2011). The mental lexicon is fully specified: evidence from eye-tracking. *Journal of Experimental Psychology: Human Perception and Performance*, 37, 496-513.
- Morgan, J. L., & Demuth, K. (2014). *Signal to Syntax: Bootstrapping From Speech to Grammar in Early Acquisition*. Psychology Press.

- Morse, P. A. (1972). The discrimination of speech and nonspeech stimuli in early infancy. *Journal of experimental child psychology*, 14, 477-492.
- Mugitani, R., Pons, F., Fais, L., Dietrich, C., Werker, J. F., & Amano, S. (2009). Perception of vowel length by Japanese-and English-learning infants. *Developmental psychology*, 45, 236-247.
- Mullennix, J. W., & Pisoni, D. B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception & Psychophysics*, 47, 379-390.
- Nam, Y. & Polka, L. (under review). Opening the speech perception door in infancy – it's easier with a stop consonant than a fricative. *Manuscript submitted to Cognition*.
- 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, 407-420.
- Nazzi, T., Floccia, C. & Bertoncini, J. (1998). Discrimination of pitch contours by neonates. *Infant Behavior & Development*, 21, 779-784.
- Nazzi, T. & Ramus, F. (2003). Perception and acquisition of linguistic rhythm by infants. *Speech Communication*, 41, 233-243.
- Nazzi T. (2005). Use of phonetic specificity during the acquisition of new words: Differences between consonants and vowels. *Cognition*, 98, 13–30.
- Nazzi, T., & Bertoncini, J. (2009). Phonetic specificity in early lexical acquisition: New evidence from consonants in coda positions. *Language and Speech*, 52, 463-480.
- Nisbett, R. E., & Miyamoto, Y. (2005). The influence of culture: holistic versus analytic perception. *Trends in cognitive sciences*, 9, 467-473.
- Obleser J., Lahiri A., & Eulitz C. (2004). Magnetic brain response mirrors extraction of phonological features from spoken vowels. *Journal of Cognitive Neuroscience*, 16, 31–39.
- Paradis, C., & Prunet, J. F. (1988). Locality in a theory of constraint violations. *Glow Newsletter*, 22, 46-48.
- Paradis, C., & Prunet, J. F. (1989a). Markedness and coronal structure. *Proceedings of North East Linguistic Society*, 19, 330-344.
- Paradis, C., & Prunet, J. F. (1989b). On coronal transparency. *Phonology*, 6(02), 317-348.
- Piaget, J. (1967) Cognition and conservations: Two views. *Contemporary Psychology*, 12, 530-533.

- Piggott, G. (1988). A parametric approach to nasal harmony. *In Features, segmental structure and harmony processes*. 1, 131-167.
- Pisoni, D.B., Lively, S.E., & Logan, J.S. (1994). Perceptual learning of non-native speech contrasts: Implications for theories of speech perception. In: Goodman, J.; Nusbaum, H.C., editors. *Development of speech perception: The transition from speech sounds to spoken words*. Cambridge, MA: MIT Press, 121-166.
- Pitt, M.A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., & Fosler-Lussier, E. (2007). *Buckeye Corpus of Conversational Speech (2nd release)*. Department of Psychology, Ohio State University.
- Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 421.
- Polka, L., & Bohn, O. S. (1996). A cross-language comparison of vowel perception in English-learning and German-learning infants. *The Journal of the Acoustical Society of America*, 100, 577-592.
- Polka, L., Colantonio, C., & Sundara, M. (2001). A cross-language comparison of /d/-/ð/perception: evidence for a new developmental pattern. *The Journal of the Acoustical Society of America*, 109, 2190-2201.
- Polka, L., & Bohn, O. S. (2003). Asymmetries in vowel perception. *Speech Communication*, 41, 221-231.
- Polka, L., Bohn, O. S., & Molnar, M. (2005). Natural referent vowels guide the development of vowel perception. *The Journal of the Acoustical Society of America*, 117, 2398-2398.
- Polka, L., & Bohn, O. S. (2011). Natural Referent Vowel (NRV) framework: An emerging view of early phonetic development. *Journal of Phonetics*, 39, 467-478.
- Polka, L., & Sundara, M. (2012). Word segmentation in monolingual infants acquiring Canadian English and Canadian French: native language, cross-dialect, and cross-language comparisons. *Infancy*, 17, 198-232.
- Pons, F., & Bosch, L. (2010). Stress pattern preference in Spanish-learning infants: The role of syllable weight. *Infancy*, 15, 223-245.
- Pons, F., Albareda - Castellot, B., & Sebastián - Gallés, N. (2012). The interplay between input and initial biases: asymmetries in vowel perception during the first year of life. *Child Development*, 83, 965-976.

- Posner M.I., Keele S.W. 1968. On the genesis of abstract ideas. *Journal of Experiment Psychology*, 77, 353–63.
- Prince, A., & Smolensky, P. (1997). Optimality: From neural networks to universal grammar. *Science*, 275, 1604-1610.
- Pulleyblank, D. (1986). *Tone in lexical phonology* (Vol. 4). Springer Science & Business Media.
- Quinn, P., Eimas, P., & Rosenkrantz, S. (1993). Evidence for representations of perceptually similar natural categories by 3- and 4-month-old infants. *Perception*, 22, 463-475.
- Quinn, P., Eimas, P. D., & Rosenkrantz, S. L. (1993). Evidence for representations of perceptually similar natural categories by 3-month-old and 4-month-old infants. *Perception-London*, 22, 463-463.
- Rafel, J. (1980). Dades sobre la freqüència de les unitats fonològiques del català. *Estudis Universitaris Catalans*, 24, 473–496.
- Ramon-Casas, Marta, Daniel Swingley, Núria Sebastián-Gallés & Laura Bosch (2009). Vowel categorization during word recognition in bilingual toddlers. *Cognitive Psychology*, 59, 96-121.
- Redford, M. A., & Diehl, R. L. (1999). The relative perceptual distinctiveness of initial and final consonants in CVC syllables. *The Journal of the Acoustical Society of America*, 106, 1555-1565.
- Ren, J., & Morgan, J.L. (2013). Segmental and suprasegmental details in early lexical representations. Paper presented at the *83rd Annual Meeting of the Linguistic Society of America*, Boston, MA.
- Ren, J., Austerweil, J. L & Morgan, J.L. (2014). Interpreting language universal constraints with Bayesian inference. Paper presented at the *84th Annual Meeting of the Linguistic Society of America*, Minneapolis, MN.
- Ren J. & Morgan J. L. (2011). Sublexical details in early lexical representations of word codas. *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS XVII)*, August, 2011, Hong Kong.
- Roberts, A. C., Wetterlin, A., & Lahiri, A. (2013). Aligning mispronounced words to meaning: Evidence from ERP and reaction time studies. *The Mental Lexicon*, 8, 140-163.

- Rugg, M. D. (1990). Event-related brain potentials dissociate repetition effects of high and low-frequency words. *Memory and Cognition*, 18, 367-379.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294), 1926-1928.
- Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, 70(1), 27-52.
- Saffran, J. R. (2001). Words in a sea of sounds: The output of statistical learning. *Cognition*, 81, 149-169.
- Saffran, J. R., Werker, J. F., & Werner, L. A. (2006). The infant's auditory world: Hearing, speech, and the beginnings of language. *Handbook of child psychology*.
- Salverda, A. P. & Tanenhaus, M. K. (2010). Tracking the time course of orthographic information in spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 1108-1117.
- Sansavini, A. (1994). Percezione della prosodia del linguaggio nei primi giorni di vita [Perception of the prosody of speech in the first days of life]. *Unpublished doctoral dissertation*, Bologna University, Bologna, Italy.
- Sansavini, A., Bertoncini, J., & Giovanelli, G. (1997). Newborns discriminate the rhythm of multisyllabic stressed words. *Developmental Psychology*, 33(1), 3-11.
- Sato, Y., Sogabe, Y., & Mazuka, R. (2007). Brain responses in the processing of lexical pitch-accent by Japanese speakers. *Neuroreport*, 18(18), 2001-2004.
- Sato, Y., Sogabe, Y., & Mazuka, R. (2010). Development of hemispheric specialization for lexical pitch-accent in Japanese infants. *Journal of Cognitive Neuroscience*, 22, 2503-2513.
- Sato, Y., Sogabe, Y., & Mazuka, R. (2010). Discrimination of phonemic vowel length by Japanese infants. *Developmental psychology*, 46, 106-119.
- Sato, Y., Kato, M., & Mazuka, R. (2012). Development of single/geminate obstruent discrimination by Japanese infants: early integration of durational and nondurational cues. *Developmental psychology*, 48, 18-34.
- Schafer, G., Plunkett, K., & Harris, P. L. (1999). What's in a name? Lexical knowledge drives infants' visual preferences in the absence of referential input. *Developmental Science*, 2, 187-194.

- Scharinger, M., Bendixen, A., Trujillo-Barreto, N. J., & Obleser, J. (2012). Scharinger M, Bendixen A, Trujillo-Barreto NJ, Obleser J (2012) A Sparse Neural Code for Some Speech Sounds but Not for Others. *PLoS ONE* 7(7): e40953.
- Schvachkin, N. K. (1973). The development of phonemic speech perception in early childhood. In C. Ferguson & D. Slobin (Eds.), *Studies of child language development* (pp. 91–127). New York: Holt, Rinehart, and Winston (originally published in 1948).
- Schwartz, J. L., Abry, C., Boë, L. J., Ménard, L., & Vallée, N. (2005). Asymmetries in vowel perception, in the context of the Dispersion–Focalisation theory. *Speech Communication*, 45, 425-434.
- Sebastián-Gallés, N. (2006). Native-language sensitivities: evolution in the first year of life. Trends in *Cognitive Sciences*, 10, 239-241.
- Segal, O., & Kishon-Rabin, L. (2012). Evidence for Language-specific influence on the preference of stress patterns in infants learning an Iambic language (Hebrew). *Journal of Speech, Language, and Hearing Research*, 55, 1329-1341.
- Sereno, J.A. and Lee, H. (2014). The contribution of segmental and tonal information in Mandarin spoken word processing. *Language and Speech*, 2014, 1-21.
- Shankweiler, D., Strange, W., & Verbrugge, R. R. (1977). Speech and the problem of perceptual constancy. Perceiving, acting, and knowing: Toward an ecological psychology, 315-345.
- Shi, R., Gao, J. & Aijun, L. (2010). Categorization of lexical tones in Mandarin-learning infants. *Technical Report of Phonetics Research*. The Chinese Academy of Social Sciences. Beijing, China.
- Shi, R., & Werker, J. F. (2001). Six-month-old infants' preference for lexical words. *Psychological Science*, 12(1), 70-75.
- Singh, L., Morgan, J., White, K. (2004). Preference and processing: The role of speech affect in early spoken word recognition. *Journal of Memory and Language*, 51, 173-189.
- Singh, L., White, K. S., & Morgan, J. L. (2008). Building a word-form lexicon in the face of variable input: Influences of pitch and amplitude on early spoken word recognition. *Language Learning and Development*, 4, 157-178.
- Skoruppa, K., Pons, F., Christophe, A., Bosch, L., Dupoux, E., Sebastián-Gallés, N., & Peperkamp, S. (2009). Language-specific stress perception by 9-month-old French and Spanish infants. *Developmental Science*, 12, 914-919.

- Skoruppa, K., Cristia, A., Peperkamp, S., & Seidl, A. (2011). English-learning infants' perception of word stress patterns. *The Journal of the Acoustical Society of America*, 130, EL50-EL55.
- Skoruppa, K., Pons, F., Bosch, L., Christophe, A. Cabrol. D. & PeperKamp, S. (2013). The development of word stress processing in French and Spanish Infants. *Language Learning and Development*, 9, 88-104.
- Slowiaczek, L. M. (1990). Effects of lexical stress in auditory word recognition. *Language and Speech*, 33, 47-68.
- Small, L. H., Simon, S. D., & Goldberg, J. S. (1988). Lexical stress and lexical access: Homographs versus nonhomographs. *Perception & Psychophysics*, 44, 272-280.
- Spring, D. R., & Dale, P. S. (1977). Discrimination of linguistic stress in early infancy. *Journal of Speech, Language, and Hearing Research*, 20(2), 224-232.
- 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.
- Steriade, D. (1987). Locality conditions and feature geometry. In *Proceedings of the North East Linguistic Society*, 17, 595-617.
- Stevens K. (1972). The quantal nature of speech: Evidence from articulatory-acoustic data. In E. E. David & P. B. Denes (Eds.), *Human Communication: A unified view*, 51-66. New York: McGraw-Hill.
- Storkel, H. L. (2002). Restructuring of similarity neighborhoods in the developing mental lexicon. *Journal of Child Language*, 29, 251-274.
- Sundara, M., Polka, L., & Genesee, F. (2006). Language-experience facilitates discrimination of /d-/in monolingual and bilingual acquisition of English. *Cognition*, 100, 369-388.
- Sussman, H. M., Hoemeke, K. A., & McCaffrey, H. A. (1992). Locus equations as an index of coarticulation for place of articulation distinctions in children. *Journal of Speech and Hearing Research*, 35, 769-781.
- Sussman, H. M., McCaffrey, H. A., & Matthews, S. A. (1991). An investigation of locus equations as a source of relational invariance for stop place categorization. *The Journal of the Acoustical Society of America*, 90, 1309-1325.
- Swingle, D. (2009). Onsets and codas in 1.5-year-olds' word recognition. *Journal of Memory and Language*, 60, 252-269.

- Swingley D., & Aslin R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science*, 13, 480–484.
- Swingley D., & Aslin R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, 76, 147-166.
- Tao, Y. & Xu. Q (2013). Phonological specificity of lexical tones in 12-month-old Chinese-speaking infants. *Acta Psychologica Sinica*. 45, 1111-1118.
- Tesar, B., & Smolensky, P. (1998). Learnability in optimality theory. *Linguistic Inquiry*, 29, 229-268.
- Tin, S. White, K. S. & Morgan, J. L. (2014). *Subsegmental detail in adult lexical representations*. Technical Report, Brown University.
- Tong, Y., Francis, A. L. & Gandour, J.T.(2008). Processing dependencies between segmental and suprasegmental features in Mandarin Chinese. *Language and Cognitive Processes*, 23, 689-708.
- Trehub, S. E. (1976). The discrimination of foreign speech contrasts by infants and adults. *Child development*, 466-472.
- Tsao, F. M., Liu, H. M., & Kuhl, P. K. (2006). Perception of native and non-native affricate-fricative contrasts: Cross-language tests on adults and infants. *The Journal of the Acoustical Society of America*, 120, 2285-2294.
- Tsao, F. M. (2006). The effect of acoustical similarity on lexical-tone perception of one-year-old Mandarin-learning infants. *Chinese Journal of Psychology*, 111-124.
- Tsao, W. (2010). Mandarin Tone Perception: A Report on Level F0 Contours. *Chinese Linguistics*. 6, 536-544.
- Tseng, C. Y. (1990). *An acoustic phonetic study on tones in Mandarin Chinese* (Vol. 94). Taipei: Institute of History & Philology, Academia Sinica.
- Tsuji, S., Mazuka, R., Cristia, A., & Fikkert, P. (2015). Even at 4 months, a labial is a good enough coronal, but not vice versa. *Cognition*, 134, 252-256.
- Van der Feest S. (2007). *Building a phonological lexicon: The acquisition of the Dutch voicing contrast in perception and production*. Ph.D. Dissertation, Radboud University, Nijmegen.
- van der Feest, S. V., & Fikkert, P. (2015). Building phonological lexical representations. *Phonology*, 32, 207-239.

- Vihman, M. M., Nakai, S., DePaolis, R. A., & Hallé, P. (2004). The role of accentual pattern in early lexical representation. *Journal of Memory and Language*, 50, 336-353.
- Whalen, D. H., & Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica*, 49, 25-47.
- Walley, A. C. (1993). The role of vocabulary development in children's spoken word recognition and segmentation ability. *Developmental Review*, 13, 286-350.
- Walley, AC (2005). Speech perception in childhood. In: Pisoni, DB.; Remez, RE. (Eds). *The handbook of speech perception*. Blackwell; Oxford. 449-468.
- 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., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7, 49-63.
- Wheeldon, L. & Waksler, R. (2004). Phonological underspecification and mapping mechanisms in the speech recognition lexicon. *Brain and Language*, 90, 401-412.
- Werker, J. F., & Lalonde, C. E. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology*, 24, 672-683.
- Werker, J. F., & Tees, R. C. (1983). Developmental changes across childhood in the perception of non-native speech sounds. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 37, 278-286.
- Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7, 49-63.
- Werker, J. F., Gilbert, J. V. H., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross language speech perception. *Child Development*, 52, 349-355.
- Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., & Stager, C. L. (1998). Acquisition of word-object associations by 14-month-old infants. *Developmental psychology*, 34, 1289-1309.
- Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development*, 1, 197-234.

- Wheat, K. L., Cornelissen, P. L., Frost, S. J., & Hansen, P. C. (2010). During visual word recognition, phonology is accessed within 100 ms and may be mediated by a speech production code: evidence from magnetoencephalography. *Journal of Neuroscience*, 30, 5229–5233.
- Wieman, L. A. (1976). Stress patterns of early child language. *Journal of Child Language*, 3, 283-286.
- White, K. S. & Morgan J. L. (2008). Sub-segmental detail in early lexical representations. *Journal of Memory and Language*, 59, 114-132.
- Xu, L., Tsai, Y., & Pfingst, B. E. (2002). Features of stimulation affecting tonal-speech perception: Implications for cochlear prostheses. *Journal of the Acoustical Society of America*, 112, 247-258.
- Ye, Y. & Connine, C. M. (1999). Processing Spoken Chinese: The role of tone information. *Language and Cognitive Processes*, 14, 609-630.
- Yeung, H.H., Chen, K.H., & Werker, J.F. (2013). When does native language input reorganize phonetic perception? The precocious case of lexical tone. *Journal of Memory and Language*, 68, 123-139.
- Yip, Moira (1980). *The tonal phonology of Chinese*. PhD dissertation, MIT.
- Yip, Moira (2002). *Tone*. Cambridge University Press. Cambridge, UK.
- Yoshida, K. A., Fennell, C. T., Swingle, D., & Werker, J. F. (2009). Fourteen-month-old infants learn similar-sounding words. *Developmental Science*, 12, 412-418.

APPENDICES

Appendix A. Assignment of stimulus pairs to conditions in Experiment 1

Display		Label				Condition
Familiar	Unfamiliar	Group1	Group2	Group3	Group4	
bath	drill	bath	baf	bav	<u>bab</u>	test stimuli
book	tent	boov	book	boop	boob	
nose	crown	nof	nok	nose	nov	
mouth	sword	mouf	mouv	moub	mouth	
bird	keg	bird	birb	birp	birf	
foot	gem	foov	foot	foop	foob	
bed	cage	bep	bef	bed	beb	
dog	vase	<u>dob</u>	<u>dop</u>	<u>dos</u>	<u>dog</u>	
brush	strut	brush	bruf	bruv	brub	
crib	stag	crish	crib	crig	crik	
bread	hut	break	breash	bread	breag	
duck	sloth	dut	dud	duv	duck	
car	jade	car	car	car	car	familiar filler
hand	wrench	hand	hand	hand	hand	
ball	hive	hive	hive	hive	hive	novel filler
girl	pump	pump	pump	pump	pump	
shoe	horn	horn	horn	horn	horn	
door	bass	bass	bass	bass	bass	

Appendix B. Results of vocabulary questionnaire for tonal stimulus selection

Pinyin	Gloss	Familiarity	Pinyin	Gloss	Familiarity
xie ²	shoe	2.13	ju ² zi	orange	1.74
yu ²	fish	2.13	xiao ³ yang ²	sheep	1.71
shao²zi	spoon	2.03	tao²zi	peach	1.68
qiu²	ball	2.03	hou²zi	monkey	1.68
chuang²	bed	2	pen ² zi	basin	1.68
bei ⁴ zi	quilt	2	xiao³xiong²	bear	1.65
bi²zi	nose	2	da ⁴ xiang ⁴	elephant	1.61
mao ⁴ zi	hat	1.97	bo ² zi	neck	1.52
xiang ¹ jiao ¹ (FF)	banana	1.97	she ² tou	tongue	1.52
tou ² fa	hair	1.94	fang ² zi	house	1.45
qi ⁴ che ¹ (FF)	car	1.94	ping ² zi	bottle	1.42
wa ⁴ zi	sock	1.94	shi ² tou ²	stone	1.26
tou²	head	1.9	xiao ³ lu ⁴	deer	1
xiao³niu²	cow	1.87	chuan ²	boat	0.97
xiao ³ niao ³	bird	1.87	long ² zi (UF)	cage	0.61
du ⁴ zi	belly	1.87	shao ⁴ zi	whistle	0.32
men²	door	1.87	bao ³ jian ⁴ (UF)	sword	0.13
tu ⁴ zi	rabbit	1.84	bao ³ shi ² (UF)	gem	0.06
ku ⁴ zi	pants	1.84	zhang ⁴ peng	tent	0
kou ³ zi	button	1.74	huang ² guan ⁴ (UF)	crown	0

Note: Items in bold were used as experimental stimuli. Items marked with FF and UF were used as familiar and unfamiliar fillers, respectively. Familiarity was computed on a scale of 0 to 3, as explained in the text.

Appendix C. Assignment of stimulus pairs to conditions in Experiment 2

Displayed Objects		Correct Pronunciation	Stimulus Tone			
Familiar	Unfamiliar		Group 1	Group 2	Group 3	Group 4
Test Stimuli						
sheep	salamander	xiaoyang ²	2	3	1	4
cow	mole	xiaoniu ²	4	2	3	1
spoon	drill	xiaoshaoer ²	1	4	2	3
bear	octopus	xiaoxiong ²	3	1	4	2
head	sloth	tou ²	2	3	1	4
door	reel	men ²	4	2	3	1
ball	vase	qiu ²	1	4	2	3
bed	wirenut	chuang ²	3	1	4	2
peach	squid	tao ² zi	2	3	1	4
orange	tool part	jv ² zi	4	2	3	1
nose	artifact	bi ² zi	1	4	2	3
monkey	tapir	hou ² zi	3	1	4	2
Familiar Fillers						
car	pump				qi ⁴ che ¹	
banana	lathe				xiang ¹ jiao ¹	
Unfamiliar Fillers						
horn	cage				long ² zi	
hand	sword				bao ³ jian ⁴	
shoe	gem				bao ³ shi ²	
bird	crown				huang ² guan ⁴	

Note: For experimental items, correct pronunciations were always tone 2, close mispronunciations were tone 3, farther mispronunciations were tone 1, and farthest mispronunciations were tone 4.

Appendix D. Items eligible for our study in the 16-month-old CDI of Hao et. al.

(2008)*

Pinyin	English	Familiarity
chuang ²	bed	91.89
men ²	door	81.08
tou ²	head	70.27
xie ²	shoe	67.57
shao ² zi	spoon	59.46
tang ² **+	candy	54.05
chong ² zi**+	bug	54.05

* This list is not used as the stimulus set of our study because it does not provide a balanced design.

**Words not included in our questionnaire of Appendix C

+ Words not used in our study.

Appendix E. Assignment of stimulus pairs to conditions in Experiment 3a

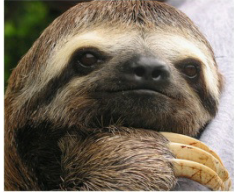
1. Experimental Design

Images		Label		
Familiar	Unfamiliar	Group1	Group2	Condition
dog	stag	gog	dog	coronal
door	drill	boor	door	coronal
duck	mole	buck	duck	coronal
table	vase	table	pable	coronal
toy	tent	toy	koy	coronal
TV	cage	teevee	keevee	coronal
baby	tapir	baby	daby	noncoronal
book	horn	book	dook	noncoronal
car	pump	tar	car	noncoronal
cat	squid	tat	cat	noncoronal
girl	sloth	girl	dirl	noncoronal
pig	mantis	tig	pig	noncoronal
ball	hive	hive	hive	novel filler
bottle	wrench	wrench	wrench	novel filler
foot	hut	hut	hut	novel filler
bath	reel	bath	bath	familiar filler
bed	punch	bed	bed	familiar filler
shoe	pulley	shoe	shoe	familiar filler

2. Pictures used for visual display

Target Name	Familiar	Unfamiliar
dog		
door		

duck		
table		
toy		
TV		
baby		
book		
car		
cat		

girl		
pig		
ball		
bottle		
foot		
bath		
bed		
shoe		

Appendix F. Assignment of stimulus pairs to conditions in Experiment 3b

1. Experimental Design

Images		Label		Condition
Familiar	Unfamiliar	Group1	Group2	
hat	vase	hat	hap	coronal
cat	stag	cat	cak	coronal
foot	gem	foot	fook	coronal
hand	drill	hang	hand	coronal
bed	hut	beb	bed	coronal
bird	mantis	birb	bird	coronal
cup	crown	cut	cup	noncoronal
sock	punch	sot	sock	noncoronal
duck	mole	dut	duck	noncoronal
dog	tapir	dog	dod	noncoronal
crib	cage	crib	crid	noncoronal
bib	strut	bib	bid	noncoronal
bath	pulley	bath	bath	familiar filler
book	reel	book	book	familiar filler
door	hive	door	door	familiar filler
cheese	wrench	wrench	wrench	novel filler
car	pump	pump	pump	novel filler
ball	horn	horn	horn	novel filler

2. Pictures used for display

Target Name	Familiar	Unfamiliar
hat		
cat		

foot		
hand		
bed		
bird		
cup		
sock		
duck		
dog		

crib		
bib		
bath		
book		
door		
cheese		
car		
ball		

Appendix G. JAGS code for Bayesian Model Comparison (two-prior approach)

```
modelstring = "  
model {  
  for ( i in 1:Ntotal ) {  
    y[i] ~ dnorm( a0 + a1[x1[i]] + a2[x2[i]] + a1a2[x1[i],x2[i]] + aS[S[i]] , pow( sigma,-2) )  
  }  
  # Prior distribution:  
  sigma ~ dunif(0,1)  
  #Hyperprior:  
  a0 ~ dnorm(m0[modelIndex], 0.001)  
  m0[1]<-0.079  
  m0[2]<-0.083  
  for ( j1 in 1:Nx1Lvl ) { a1[j1] ~ dnorm( m1[modelIndex] , pow( (abs( a1SDunabs ) + .000001) , -2 ) ) }  
  a1SDunabs ~ dt( 0 , 0.001, 26)  
  m1[1]<-0.3  
  m1[2]<-0  
  for ( j2 in 1:Nx2Lvl ) { a2[j2] ~ dnorm( m2[modelIndex] , pow( (abs( a2SDunabs ) + .000001) , -2 ) ) }  
  a2SDunabs ~ dt( 0 , 0.0001, 26)  
  m2[1]<-0.08  
  m2[2]<-0.08  
  for ( j1 in 1:Nx1Lvl ) { for ( j2 in 1:Nx2Lvl ) {  
    a1a2[j1,j2] ~ dnorm( mI[modelIndex] , pow( (abs( a1a2SDunabs ) + .000001) , -2 ) )  
  } }  
  a1a2SDunabs ~ dt( 0 , 0.001, 26 )  
  mI[1]<-0  
  mI[2]<-0.9  
  for ( jS in 1:NSLvl ) { aS[jS] ~ dnorm( mS[modelIndex] , pow( (abs( aSSDunabs ) + .000001) , -2 ) )  
  }  
  aSSDunabs ~ dt( 0 , 0.0001 , 26 )  
  mS[1]<-0.46  
  mS[2]<-0.46  
  # Hyperhyperprior:  
  modelIndex ~ dcat(modelProb[])  
  modelProb[1] <- .50  
  modelProb[2] <- .50  
  }  
  }  
"
```

Appendix H. Assignment of stimulus pairs to conditions in Experiment 4a

Experimental Design (see supplementary materials for pictures used for display)

Images		Label		
Familiar	Unfamiliar	Group1	Group2	Conditions
dove	1	dove	bove	coronal
donkey	2	donkey	bonkey	coronal
diver	3	diver	biver	coronal
dome	4	dome	bome	coronal
diamond	5	diamond	biamond	coronal
dolphin	6	dolphin	bophin	coronal
dipper	7	bipper	dipper	coronal
doctor	8	boctor	doctor	coronal
dish	9	bish	dish	coronal
dice	10	bice	dice	coronal
diary	11	biary	diary	coronal
desert	12	besert	desert	coronal
desk	13	desk	gesk	coronal
disk	14	disk	gisk	coronal
duck	15	duck	guck	coronal
diaper	16	giaper	diaper	coronal
daisy	17	gaisy	daisy	coronal
donut	18	gonut	donut	coronal
toe	19	toe	koe	coronal
TV	20	TV	KV	coronal
tent	21	tent	kent	coronal
tank	22	tank	kank	coronal
ticket	23	ticket	kicket	coronal
tissue	24	tissue	kissue	coronal
tower	25	kower	tower	coronal
tail	26	kail	tail	coronal
taxi	27	kaxi	taxi	coronal
toilet	28	koilet	toilet	coronal
tadpole	29	kadpole	tadpole	coronal
tie	30	kie	tie	coronal
turkey	31	turkey	purkey	coronal
tape	32	tape	pape	coronal
turtle	33	turtle	purtle	coronal
table	34	table	pable	coronal
tiger	35	tiger	piger	coronal
toad	36	toad	poad	coronal
bench	37	dench	bench	noncoronal

boat	38	doat	boat	noncoronal
bone	39	doan	bone	noncoronal
bowl	40	dowl	bowl	noncoronal
boy	41	doy	boy	noncoronal
book	42	dook	book	noncoronal
box	43	box	dox	noncoronal
bath	44	bath	dath	noncoronal
bird	45	bird	dird	noncoronal
bat	46	bat	dat	noncoronal
boot	47	boot	doot	noncoronal
bus	48	bus	dus	noncoronal
coach	49	toach	coach	noncoronal
cotton	50	totton	cotton	noncoronal
couch	51	tauch	couch	noncoronal
cage	52	tage	cage	noncoronal
cougar	53	tougar	cougar	noncoronal
keg	54	teg	keg	noncoronal
cushion	55	cushion	tushion	noncoronal
curtain	56	curtain	turtain	noncoronal
kettle	57	kettle	tettle	noncoronal
camel	58	camel	tamel	noncoronal
canoe	59	canoe	tanoe	noncoronal
candy	60	candy	tandy	noncoronal
pear	61	tare	pear	noncoronal
pasta	62	tasta	pasta	noncoronal
pizza	63	tizza	pizza	noncoronal
piano	64	piano	tiano	noncoronal
pencil	65	pencil	tencil	noncoronal
parrot	66	parrot	tarrot	noncoronal
girl	67	girl	dirl	noncoronal
gift	68	gift	dift	noncoronal
goose	69	goose	doose	noncoronal
garlic	70	darlic	garlic	noncoronal
ghost	71	dhost	ghost	noncoronal
garden	72	darden	garden	noncoronal
ear	73	ear	ear	familiar filler
fan	74	fan	fan	familiar filler
spoon	75	spoon	spoon	familiar filler
tree	76	tree	tree	familiar filler
hand	77	hand	hand	familiar filler
hat	78	hat	hat	familiar filler
bread	79	bread	bread	familiar filler
flag	80	flag	flag	familiar filler

foot	81	foot	foot	familiar filler
ladder	82	ladder	ladder	familiar filler
fox	83	fox	fox	familiar filler
guitar	84	guitar	guitar	familiar filler
comb	85	comb	comb	familiar filler
mango	86	mango	mango	familiar filler
pen	87	pen	pen	familiar filler
vase	88	vase	vase	familiar filler
wine	89	wine	wine	familiar filler
watch	90	watch	watch	familiar filler
bermate	91	bermate	bermate	novel filler
dalok	92	dalok	dalok	novel filler
dwatch	93	dwatch	dwatch	novel filler
golpsle	94	golpsle	golpsle	novel filler
gomtox	95	gomtox	gomtox	novel filler
boupet	96	boupet	boupet	novel filler
lepter	97	lepter	lepter	novel filler
nando	98	nando	nando	novel filler
nescot	99	nescot	nescot	novel filler
quish	100	quish	quish	novel filler
fansel	101	fansel	fansel	novel filler
fapton	102	fapton	fapton	novel filler
holek	103	holek	holek	novel filler
stipe	104	stipe	stipe	novel filler
swike	105	swike	swike	novel filler
topal	106	topal	topal	novel filler
twipter	107	twipter	twipter	novel filler
suple	108	suple	suple	novel filler

** The numbers in the column are numbered novel objects. Since there is no name for these objects, we used numbers to label them. See *supplementary material* for an example of one of these novel objects, and see supplementary materials for a complete list of the novel objects used in the current experiment.

Appendix I. Assignment of stimulus pairs to conditions in Experiment 4b

Experimental Design (see supplementary materials for pictures used for display)

Display		Label		
Familiar	Unfamiliar	Group1	Group2	Conditions
bed	1	bed	beb	coronal
bread	2	bread	breab	coronal
beard	3	beard	bearb	coronal
bird	4	birb	bird	coronal
food	5	foob	food	coronal
mold	6	molb	mold	coronal
head	7	head	heab	coronal
spade	8	spade	spabe	coronal
cloud	9	cloud	cloub	coronal
bud	10	bub	bud	coronal
blade	11	blabe	blade	coronal
mud	12	mub	mud	coronal
gold	13	gold	golg	coronal
blood	14	blood	bloog	coronal
slide	15	slide	slige	coronal
thread	16	threag	thread	coronal
sword	17	sworg	sword	coronal
crowd	18	crowg	crowd	coronal
cat	19	cat	cak	coronal
boot	20	boot	buk	coronal
foot	21	foot	fook	coronal
boat	22	boak	boat	coronal
goat	23	goak	goat	coronal
suit	24	suik	suit	coronal
jet	25	jet	jek	coronal
fruit	26	fruit	fruik	coronal
root	27	root	ruk	coronal
scout	28	scouk	scout	coronal
knight	29	knighk	knight	coronal
nut	30	nuk	nut	coronal
net	31	net	nep	coronal
flute	32	flute	flupe	coronal
light	33	light	lipe	coronal
gut	34	gup	gut	coronal
sprout	35	sproup	sprout	coronal
skate	36	skape	skate	coronal
crib	37	crib	crid	noncoronal

crab	38	crab	crad	noncoronal
shrub	39	shrub	shrud	noncoronal
orb	40	ord	orb	noncoronal
probe	41	prode	probe	noncoronal
club	42	clud	club	noncoronal
tube	43	tube	tude	noncoronal
tub	44	tub	tud	noncoronal
swab	45	swab	swad	noncoronal
flag	46	flad	flag	noncoronal
pig	47	pid	pig	noncoronal
plug	48	plud	plug	noncoronal
frog	49	frog	frod	noncoronal
dog	50	dog	dod	noncoronal
keg	51	keg	ked	noncoronal
fig	52	fid	fig	noncoronal
crag	53	crad	crag	noncoronal
drug	54	drud	drug	noncoronal
book	55	boot	boot	noncoronal
snake	56	snake	snate	noncoronal
truck	57	truck	trut	noncoronal
clerk	58	clert	clerk	noncoronal
stick	59	stit	stick	noncoronal
elk	60	elt	elk	noncoronal
stork	61	stork	stort	noncoronal
crock	62	crock	crot	noncoronal
yolk	63	yolk	yolt	noncoronal
soap	64	soat	soap	noncoronal
scope	65	scote	scope	noncoronal
tape	66	tate	tape	noncoronal
scalp	67	scalp	scalt	noncoronal
drop	68	drop	drot	noncoronal
jeep	69	jeep	jeet	noncoronal
strap	70	strat	strap	noncoronal
chop	71	chot	chop	noncoronal
pope	72	pote	pope	noncoronal
gun	73	gun	gun	familiar filler
hand	74	hand	hand	familiar filler
apple	75	apple	apple	familiar filler
ear	76	ear	ear	familiar filler
cheese	77	cheese	cheese	familiar filler
cream	78	cream	cream	familiar filler
door	79	door	door	familiar filler
pen	80	pen	pen	familiar filler

nose	81	nose	nose	familiar filler
mouse	82	mouse	mouse	familiar filler
wolf	83	wolf	wolf	familiar filler
hat	84	hat	hat	familiar filler
peach	85	peach	peach	familiar filler
clock	86	clock	clock	familiar filler
grape	87	grape	grape	familiar filler
mouth	88	mouth	mouth	familiar filler
fish	89	fish	fish	familiar filler
brush	90	brush	brush	familiar filler
shrok	91	shrok	shrok	novel filler
klup	92	klup	klup	novel filler
ralk	93	ralk	ralk	novel filler
blerm	94	blerm	blerm	novel filler
dalp	95	dalp	dalp	novel filler
dwatch	96	dwatch	dwatch	novel filler
slem	97	slem	slem	novel filler
prack	98	prack	prack	novel filler
pruse	99	pruse	pruse	novel filler
troge	100	troge	troge	novel filler
gompt	101	gompt	gompt	novel filler
plaft	102	plaft	plaft	novel filler
slopt	103	slopt	slopt	novel filler
laust	104	laust	laust	novel filler
drept	105	drept	drept	novel filler
monge	106	monge	monge	novel filler
nelp	107	nelp	nelp	novel filler
thraft	108	thraft	thraft	novel filler

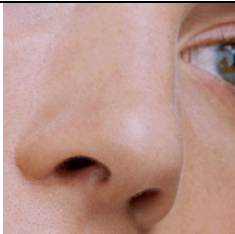
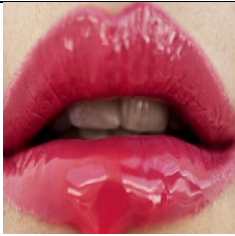
** The numbers in the column are numbered novel objects. Since there is no name for these objects, we used numbers to label them. See *Supplementary Material* for an example of one of these novel objects, and see supplementary materials for a complete list of the novel objects used in the current experiment.

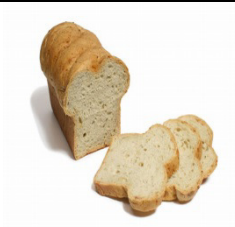
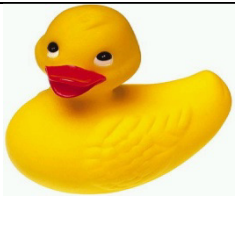
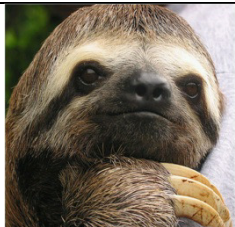
Appendix J. JAGS code for Bayesian parameter estimation (single-prior approach)

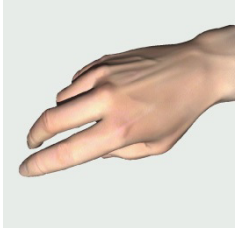
```
modelstring = "  
model {  
  for ( i in 1:Ntotal ) {  
    y[i] ~ dnorm( mu[i] , tau )  
    mu[i] <- a0 + a1[x1[i]] + a2[x2[i]] + a1a2[x1[i],x2[i]] + aS[S[i]]  
  }  
  # Priors  
  tau <- pow( sigma , -2 )  
  sigma ~ dunif(0,10) # y values are assumed to be standardized  
  # Hyperpriors  
  # for a0  
  a0 ~ dnorm(0,0.001) # y values are assumed to be standardized  
  # for a1  
  for ( j1 in 1:Nx1Lvl ) { a1[j1] ~ dnorm( 0.0 , a1tau ) }  
  a1tau <- 1 / pow( a1SD , 2 )  
  a1SD <- abs( a1SDunabs ) + .1  
  a1SDunabs ~ dgamma(0.2,0.2)  
  # for a2  
  for ( j2 in 1:Nx2Lvl ) { a2[j2] ~ dnorm( 0.0 , a2tau ) }  
  a2tau <- 1 / pow( a2SD , 2 )  
  a2SD <- abs( a2SDunabs ) + .1  
  a2SDunabs ~ dgamma(0.2,0.2)  
  # for a1a2  
  for ( j1 in 1:Nx1Lvl ) { for ( j2 in 1:Nx2Lvl ) {  
    a1a2[j1,j2] ~ dnorm( 0.0 , a1a2tau )  
  } }  
  a1a2tau <- 1 / pow( a1a2SD , 2 )  
  a1a2SD <- abs( a1a2SDunabs ) + .1  
  a1a2SDunabs ~ dgamma(0.2,0.2)  
  # for aS  
  for ( jS in 1:NSLvl ) { aS[jS] ~ dnorm( 0.0 , aStau ) }  
  aStau <- 1 / pow( aSSD , 2 )  
  aSSD <- abs( aSSDunabs ) + .1  
  aSSDunabs ~ dt( 0 , 0.001 , 2 )  
}  
"
```


SUPPLIMENTARY MATERIALS

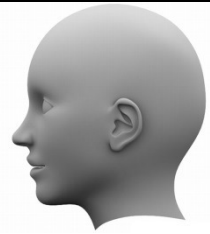
I. Visual stimulus pairings to in Experiment 1

Target Name	Familiar	Unfamiliar
bath		
book		
nose		
mouth		
bird		
foot		

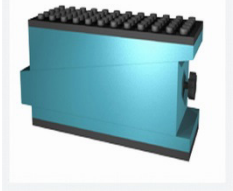
bed		
book		
brush		
crib		
bread		
duck		

car		
hand		
hive		
pump		
horn		
bass		

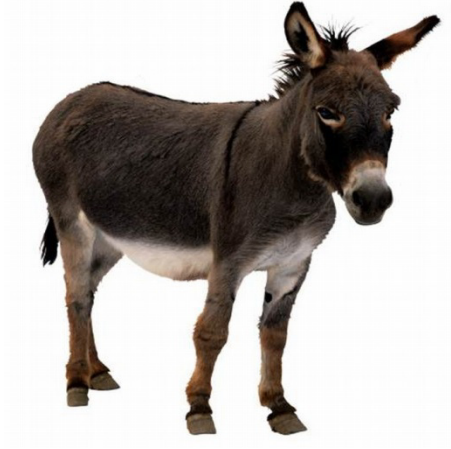
II. Visual stimulus pairings to in Experiment 2

Target Name	English Gloss	Familiar	Unfamiliar
tou ²	head		
men ²	door		
qiu ²	ball		
chuang ²	bed		
tao ² zi	peach		
ju ² zi	orange/ tangerine		

bi ² zi	nose		
hou ² zi	monkey		
xiaoyang ²	sheep		
xiaoniu ²	cow		
xiaoshao ²	spoon		
xiaoxiong ²	bear		

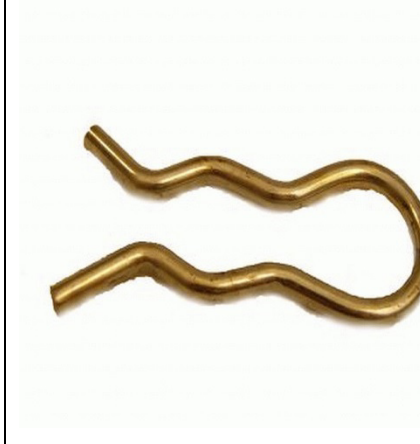
qi ⁴ che ¹	car		
xiang ¹ jiao ¹	banana		
long ² zi	cage		
bao ³ jian ⁴	sword		
bao ³ shi ²	Shoe		
huang ² guan ⁴	crown		

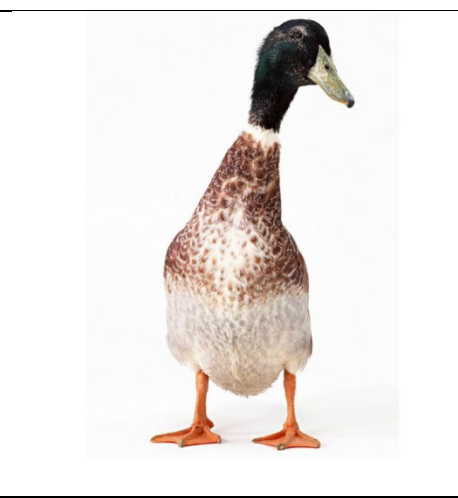
III. Visual stimulus pairings to in Experiment 4a & 5

Target Name	Familiar	Unfamiliar
dove		
donkey		
diver		

dome		
diamond		
dolphin		

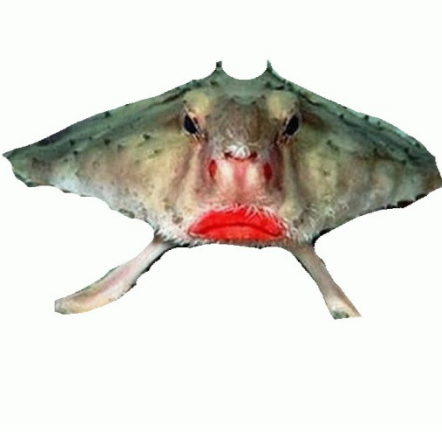
dipper		
doctor		
dish		

dice	 A single red die with white pips, showing the numbers one, two, and three.	 Two brass-colored metal rods with a wavy, zig-zag shape, one slightly longer than the other.
diary	 A black diary with a pattern of red and pink hearts. It has a silver-colored metal clasp.	 Two red plastic caps. One is a short, fluted cap with a threaded interior. The other is a taller, fluted cap with a textured top.
desert	 A photograph of a desert landscape with rolling sand dunes. A small group of camels and riders is visible in the distance.	 A dark green, textured rectangular object, possibly a piece of fabric or a small rug, with a fringed or frayed edge.

desk		
disk		
duck		

diaper		
daisy		
donut		

toe		
TV		
tent		

tank		
ticket		
tissue		

tower	 A photograph of the Eiffel Tower in Paris, France, showing its iconic lattice structure against a clear sky.	 A photograph of a traditional wooden crossbow, showing its stock, trigger mechanism, and the horizontal limbs.
tail	 A photograph of a long, dark, and bushy monkey tail, likely from a howler monkey, shown against a white background.	 A photograph of a cast iron pump housing or manifold, featuring multiple ports and a central circular opening.
taxi	 A photograph of a yellow taxi cab, specifically a Honda Civic, shown from a side profile.	 A photograph of a blue hydraulic pump or motor unit, featuring a large central port and various mounting points.

toilet		
tadpole		
tie		

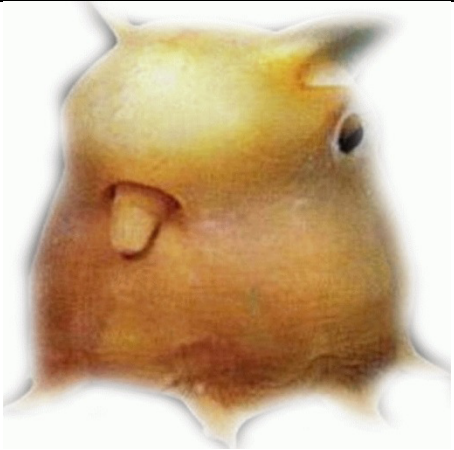
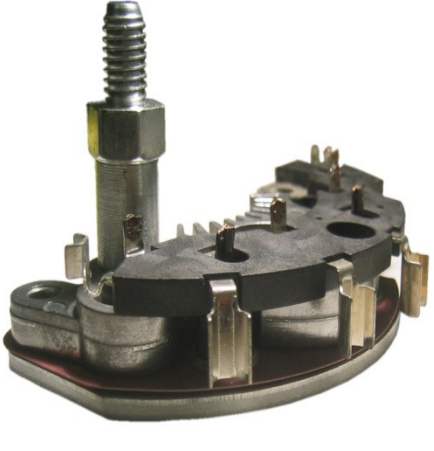
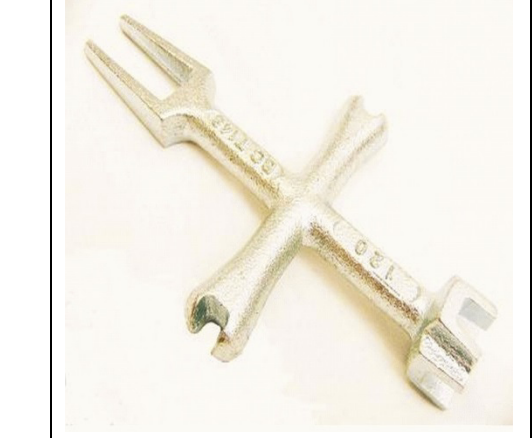
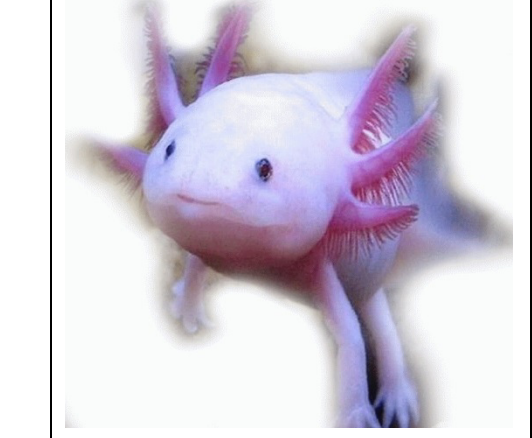
turkey		
tape		
turtle		

table	 A dark-stained wooden table with a square top and four turned legs. The table has a decorative carved apron and is shown against a plain white background.	 A complex mechanical assembly featuring a central threaded rod with a hex nut. The assembly includes various metal components, including a base plate with multiple mounting points and a series of small, rectangular metal pieces arranged in a row.
tiger	 A close-up photograph of a tiger's face, showing its distinctive orange fur with black stripes and its intense yellow eyes. The tiger is looking slightly to the left.	 A close-up photograph of a dark-colored bird's head, possibly a crow or raven, showing its long, pointed beak and dark feathers. The bird is looking to the left.
toad	 A photograph of a green and brown spotted toad, likely a common toad, shown from a side profile. The toad has a bumpy, textured skin and is sitting on a plain white background.	 A dark, rectangular object, possibly a piece of wood or metal, with a circular feature in the center. It has a rough, weathered appearance and is shown against a plain white background.

bench		
boat		
bone		

bowl		
boy		
book		

box		
bath		
bird		

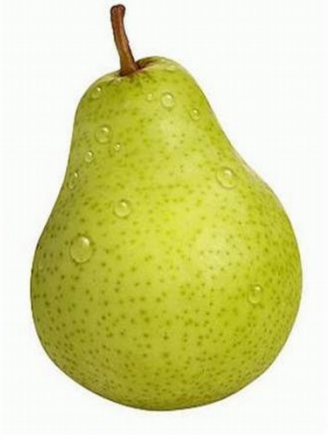
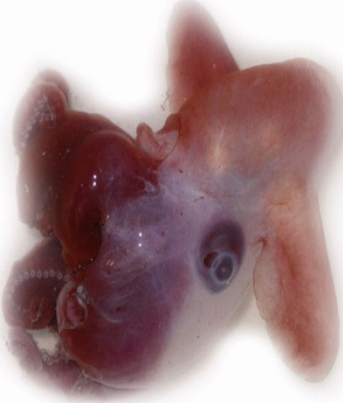
bat		
boot		
bus		

coach		
cotton		
couch		

cage		
cougar		
keg		

cushion		
Curtain		
kettle		

camel		
canoe		
candy		

pear		
pasta		
pizza		

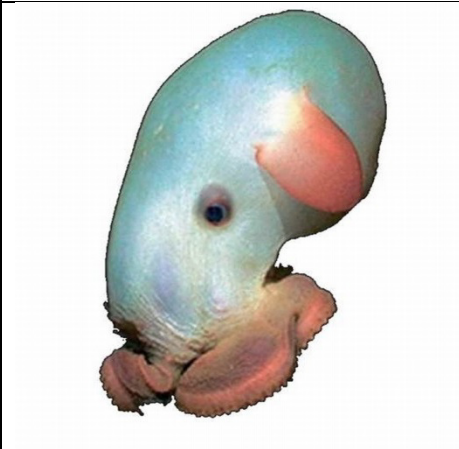
piano



pencil



parrot



girl

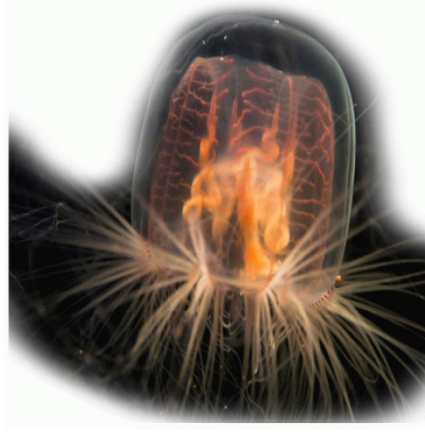


gift



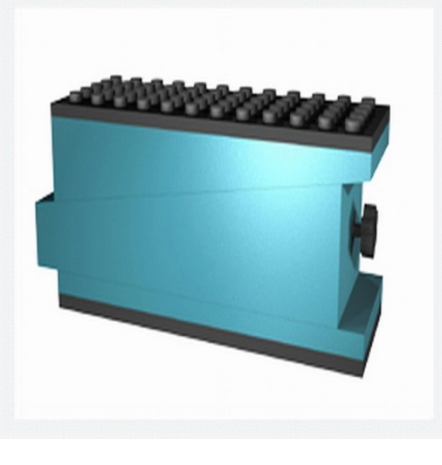
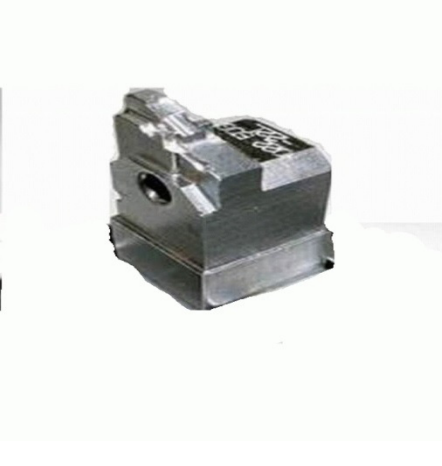
goose



garlic		
ghost		
garden		

ear		
fan		
spoon		

tree		
hand		
hat		

bread	 A photograph of a loaf of bread and several slices of bread, showing the texture of the crust and the soft interior.	 A 3D rendering of a blue mechanical component, possibly a part of a machine or a tool, with a textured surface and a central opening.
flag	 A photograph of the United States flag, showing the stars and stripes waving.	 A photograph of a circular metal component, possibly a part of a machine or a tool, with a central hole and a textured surface.
foot	 A photograph of a baby's foot, showing the toes and the sole of the foot.	 A photograph of a small metal component, possibly a part of a machine or a tool, with a textured surface and a central opening.

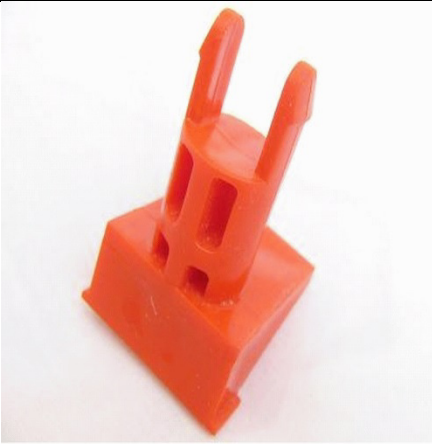
ladder		
fox		
guitar		

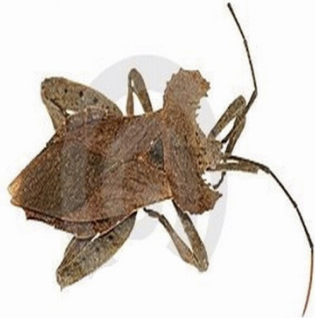
comb		
mango		
pen		

vase		
wine		
watch		

bermate		
dalok		
dwatch		

golpsle		
gomtox		
boupet		

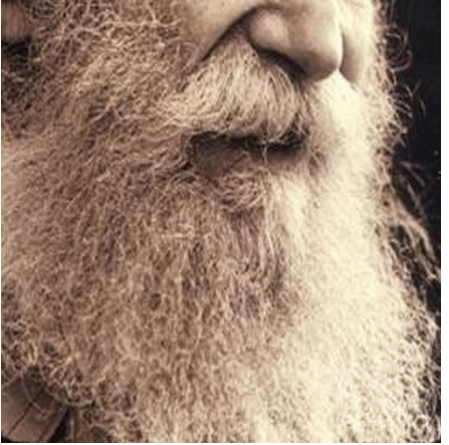
lepter	 A bright orange plastic object with a rectangular base and two vertical, slightly curved prongs extending upwards from the center.	 A dark green vintage pickup truck, likely from the 1930s, shown from a front-three-quarter view against a white background.
nando	 A bright yellow lizard, possibly a spiny-tailed lizard, perched on a cluster of green leaves.	 A close-up of a capuchin monkey's face, showing its white fur, dark face, and large eyes.
nescot	 A white, dome-shaped electronic device with a blue button on top, sitting on a circular base.	 A single, ripe red apple with a green leaf attached to its stem, set against a white background.

quish		
fanel		
fapton		

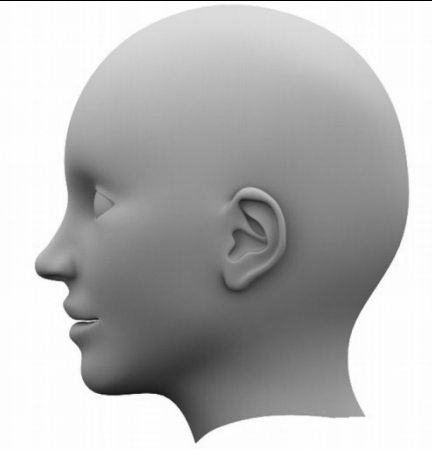
holek		
stipe		
swike		

topal		
twipter		
suple		

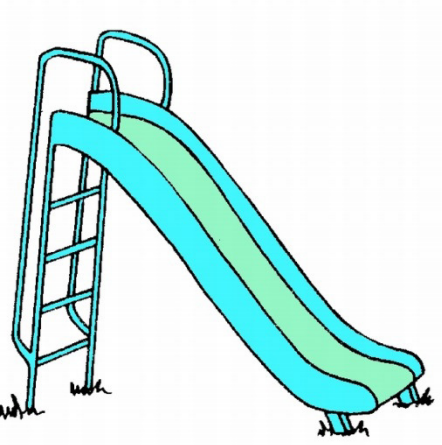
IV. Visual stimulus pairing to conditions in Experiment 4b

Target Names	Familiar	Unfamiliar
bed		
bread		
beard		

bird		
food		
mold		

head		
spade		
cloud		

bud		
blade		
mud		

gold		
blood		
slide		

thread		
sword		
crowd		

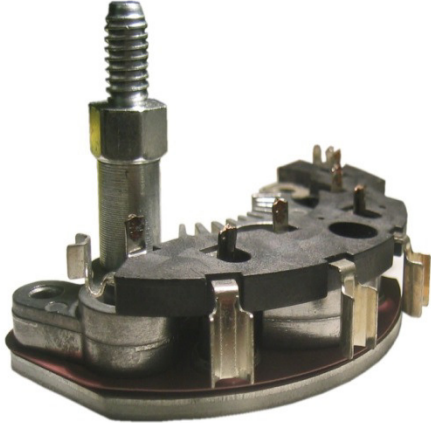
cat		
boot		
foot		

boat		
goat		
suit		

jet		
fruit		
root		

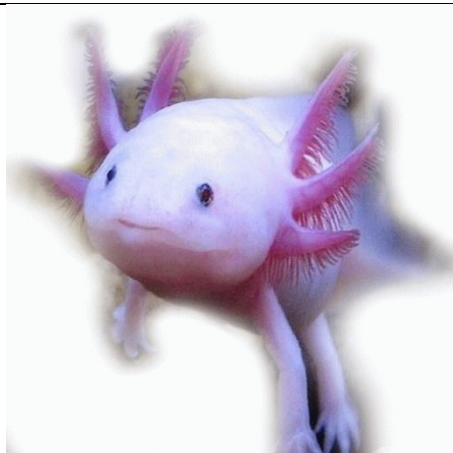
scout		
knight		
nut		

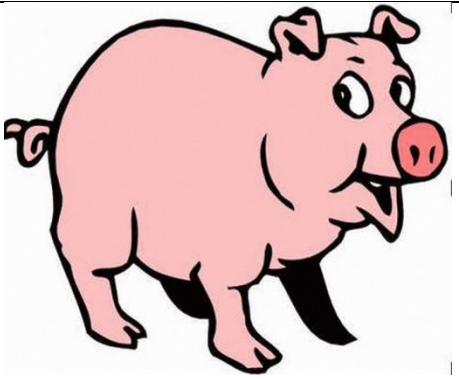
net		
flute		
light		

gut		
sprout		
skate		

crib		
crab		
shrub		

orb		
probe		
club		

tube		
tub		
swab		

flag		
pig		
plug		

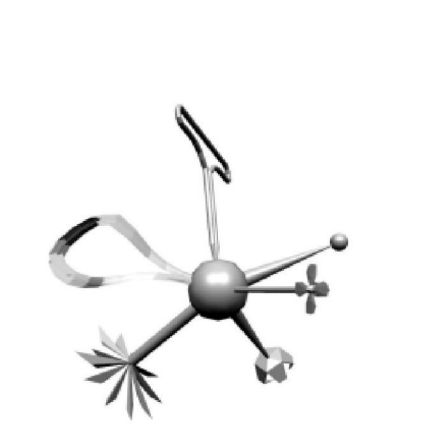
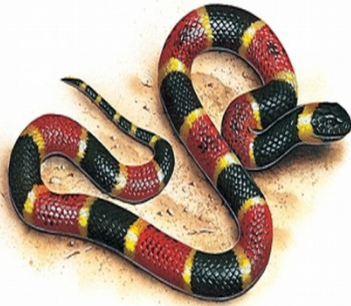
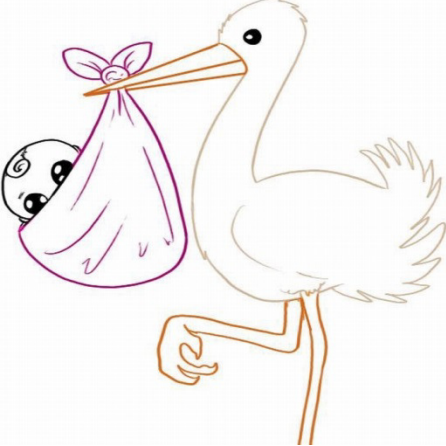
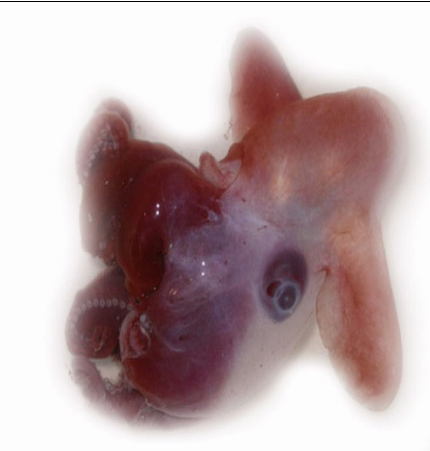
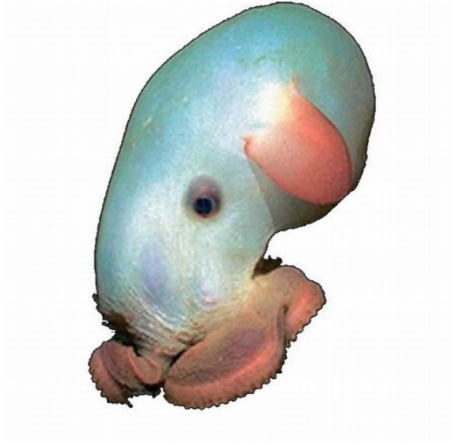
frog		
dog		
keg		

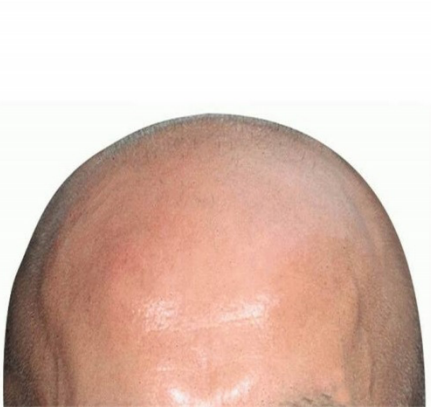
fig		
crag		
drug		

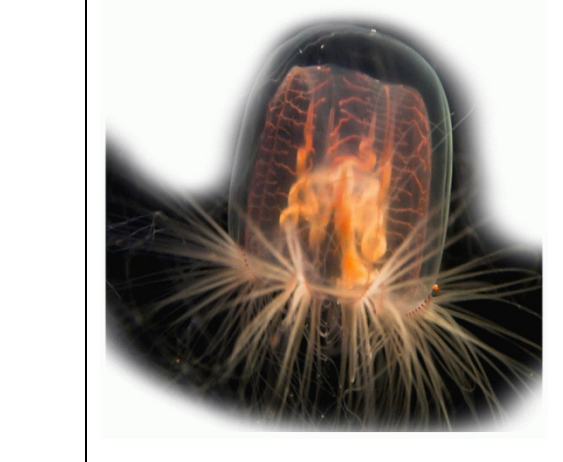
book		
snake		
truck		

clerk		
stick		
elk		

stork		
crook		
yolk		

soap		
scope		
tape		

scalp		
drop		
jeep		

strap		
chop		
pope		

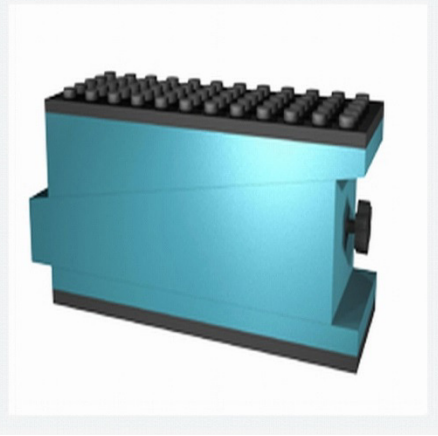
gun		
hand		
apple		

ear		
cheese		
cream		

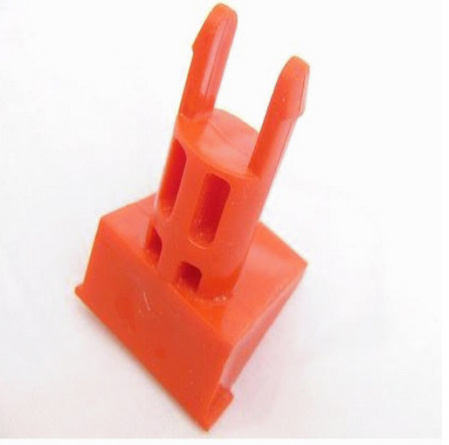
door		
pen		
nose		

mouse		
wolf		
hat		

peach		
clock		
grape		

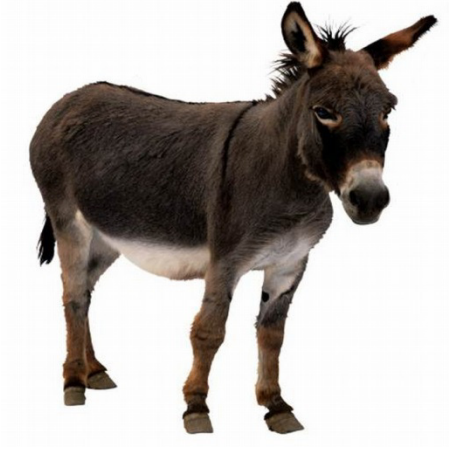
mouth		
fish		
brush		

shrok		
klup		
ralk		

blerm		
dalp		
dwatch		

slem		
prack		
pruse		

troge		
gompt		
plaft		

slopt		
laust		
drept		

monge		
nelp		
thraft		