Development of phonologically specified word forms

Karen Elizabeth Mulak

Thesis submitted to the University of Western Sydney
for the degree of Doctor of Philosophy

July 2012
To Mom and Dad, 
to Dziadziu, 
and to Damien.
Acknowledgements

This work could not have been completed without the guidance and expertise of my supervisory panel. I thank my principal supervisor Prof. Cathi Best for her support and encouragement. Her knowledge and experience have been invaluable at every stage of this project. I thank Dr. Michael Tyler for his assistance and advice in both practical and theoretical aspects of my research, and Dr. Christine Kitamura and Prof. Anne Cutler for their support and input. I also extend my thanks to Dr. Julia Irwin, for her early guidance and for encouraging me to do a PhD in this field.

It is difficult to imagine completing this project without such a supportive technical team. I thank lab manager Colin Schoknecht, Johnson Chen, Steven Fazio, and Matt Petoe, whose expertise and willingness to help at a moment's notice have turned potential Chernobyls into mere Three Mile Islands. Likewise, I thank Darlene Williams, Sonya O'Shanna, Karen McConachie, Gail Charlton, and Kim Buckley, who make up the fantastic administrative team at Marcs. Their
organization and attention to detail have ensured I have not missed a step as I've progressed through the PhD process.

Behind every successful infant study is a successful BabyLab, and I would like to thank the ladies behind Marcs BabyLab. Lab coordinator Rachel Lee has been happy to help in any way possible, and her recruitment efforts have proved very valuable. I also thank research assistants Michelle Pal, Anne Dwyer, and Carly Hayman, who gladly showed me the BabyLab ropes, assisted with scheduling and testing at times, and never turned down a request to do some coloring with a participant's older brother or sister.

It has been a pleasure doing a PhD at an institution that has had such a tight-knit student community in which students are willing to help, teach, and support one another whenever necessary. I thank my Marcs roommates, Tonya Agostini, Janise Shanley, Rikke Bundgaard-Nielsen, and Nathan Perry for their support, encouragement, and commiseration. I also thank other members of the Marcs community, Damien Smith, Mark Antoniou, Christa Lam-Cassettari, Christian Kroos, Bronson Harry, and Erin Cvejic for manuscript editing, software tutorials, stats lessons, and lunch table advice.

For stimulus recordings, I thank all the participants who lent their voices for a few hours. For the overseas recordings, I also thank Melanie Pinet and Dr. Bronwen Evans at University College London, Prof. Gerry Docherty at Newcastle University, and Prof. Paul Foulkes and Hazel Richards at the University of York, whose hospitality and assistance made the recording process and participant recruitment a breeze.
I thank my parents, Barbara and David, and my siblings, Brian and Lauren, who have always supported me and have made me feel that I am not half a world away. I also thank my friends and family who have shown support and encouragement through cards, postcards, online chats, and emails. As well, I thank Damien Smith, who has supported me in every step of this PhD through his unconditional love, tolerance, and spelling corrections.

Finally, infant research is 1% inspiration, and 99% participation, and thus I extend a huge thank you to my tiny participants, who did their best to listen to my funny sentences. I also thank all the wonderful mothers, fathers, and grandparents that I have met, who have gladly sacrificed their time to come in and participate, and without whom there would be no infant research.
The work presented in this thesis is, to the best of my knowledge and belief, original except as acknowledged in the text. I hereby declare that I have not submitted this material either in full or in part, for a degree at this or any other institution.
# Table of Contents

1 Overview...........................................................................................................................................................................1
  1.1 Motivation of the Current Research...........................................................................................................................1
  1.2 Structure of the Thesis......................................................................................................................................................4

2 The Development of Speech Perception.............................................................................................................................8
  2.1 Sensitivity to Native Meter..............................................................................................................................................8
  2.2 Attunement to Native Consonants and Vowels..................................................................................................................10
  2.3 Theories of the Development of Speech Perception.......................................................................................................13
    2.3.1 Exemplar accounts.....................................................................................................................................................13
    2.3.2 Perceptual Attunement Approach............................................................................................................................14

3 Development of Word Segmentation.................................................................................................................................17
  3.1 Word Segmentation as a Precursor to Word Learning....................................................................................................17
  3.2 Cues to Word Boundaries..............................................................................................................................................18
    3.2.1 Tracking transitional probabilities of syllables.........................................................................................................18
  3.3 Use of Segmental Features as Cues to Word Boundaries...............................................................................................19
    3.3.1 Phonotactics...............................................................................................................................................................19
    3.3.2 Allophones.................................................................................................................................................................20
    3.3.3 Coarticulation.............................................................................................................................................................20
  3.4 Use of Suprasegmental Features as Cues to Word Boundaries......................................................................................21
    3.4.1 Use of metrical patterns as cues for word boundaries...............................................................................................22
      3.4.1.1 Metrical classes of languages.................................................................................................................................22
      3.4.1.2 Sensitivity to native metrical patterning: The Metrical Segmentation Strategy (MSS).................................................................25
11.2.2 Differences in task demands in the listening preference task compared to the literature ................................................................. 192
11.2.3 Familiarity versus novelty preferences in listening preference task ................................................................. 194
11.2.4 Implications ................................................................................................................................. 197
  11.2.4.1 The importance of controlling task demands ................................................................. 197
  11.2.4.2 Fragility of listening preference tasks ........................................................................ 198
  11.2.4.3 Further directions (methodological) ............................................................................. 199
11.3 Development of Phonologically Specified Word Forms ................................................................. 200
  11.3.1 Emergence of phonological constancy .............................................................................. 200
  11.3.2 Role of vocabulary ......................................................................................................... 200
  11.3.3 Sensitivity to suprasegmental features ............................................................................. 201
  11.3.4 Sensitivity to segmental features .................................................................................... 203
  11.3.5 Category Goodness versus Category Shifting ................................................................ 204
11.4 Developmental Course of Word Segmentation and Phonologically Specified Word Forms: A Proposed Model ....................................................................................... 205
11.5 Conclusion ......................................................................................................................................... 211
12 References ............................................................................................................................................. 213
13 Appendix A: Additional Analyses for Experiment 1 .............................................................................. 243
14 Appendix B: Participant Information and Consent Forms ............................................................................. 253
List of Tables

Table 1. Target words and their phonetic realizations in the native (AusE) and non-native (JaME) regional accent......................................................... 98
Table 2. Target words and carrier sentences..........................................................130
Table 3. Target words for Experiment 3.............................................................156
Table 4. Carrier sentences used in Experiment 3...............................................157
Table 5. Target words for Experiment 4.............................................................172
Table 6. Carrier sentences used in Experiment 4...............................................173
Table A-1. Mean time of first-peak to the target image across 15- and 19-month-olds in the native and non-native regional accent conditions.........252
Table A-2. Mean time of first peak to the target image with a value of at least 420 pixels, across 15- and 19-month-olds in the native and non-native regional accent conditions........................................................252
Table A-3. Mean time of first peak to the target image with a value of at least 440 pixels, across 15- and 19-month-olds in the native and non-native regional accent conditions........................................................252
List of Figures

Figure 1. Placement of images and areas of interest...............................................101
Figure 2. Mean percentage of fixation time to the target image..............................107
Figure 3. Mean percentage of fixation time to the target image versus log-
transformed expressive vocabulary size at 15 months......................................108
Figure 4. Mean percentage of fixation time to the target image versus log-
transformed expressive vocabulary size at 19 months......................................109
Figure 5. Horizontal location of gaze for 15- and 19-month-olds in the native
regional accent condition...................................................................................111
Figure 6. Horizontal location of gaze for 15- and 19-month-olds in the non-
native regional accent condition.......................................................................112
Figure 7. Total looking time to sentences containing familiarized vs.
unfamiliarized words in the native (AusE) and non-native (FrE)
suprasegmentals conditions................................................................................137
Figure 8. Percentage total looking time for sentences containing familiarized vs.
unfamiliarized words in the native (AusE) and non-native (FrE)
suprasegmentals conditions................................................................................138
Figure 9. Total looking time to sentences containing familiarized vs.
unfamiliarized words in the native (AusE) non-native (TE) segments
conditions...........................................................................................................165
Figure 10. Difference in looking time to sentences containing familiarized vs.
unfamiliarized words in the native (AusE) and non-native (TE)
segments conditions...........................................................................................166
Figure 11. Total looking time in the non-native segments condition (TE) across sentences containing familiarized vs. unfamiliarized words that contain either a CG or CS vocalic variation

Figure 12. Difference in looking time in the non-native segments condition (TE) across sentences containing familiarized vs. unfamiliarized words that contain either a CG or CS vocalic variation

Figure 13. Total looking time to sentences containing familiarized vs. unfamiliarized words in the native (AusE) and non-native (seLE) segments conditions

Figure 14. Difference in looking time to sentences containing familiarized vs. unfamiliarized words in the native (AusE) and non-native (seLE) segments conditions

Figure 15. Total looking time in the non-native segments condition (seLE) across sentences containing familiarized vs. unfamiliarized words that contain either a CG or CS consonant variation

Figure 16. Difference in looking time in the non-native segments condition (seLE) across sentences containing familiarized vs. unfamiliarized words that contain either a CG or CS consonant variation

Figure A-1. Fifteen- and 19-month-olds' raw horizontal location of gaze during two target word repetitions in AusE, organized into 100 ms bins

Figure A-2. Fifteen- and 19-month-olds' raw horizontal location of gaze during two target word repetitions in JaME, organized into 100 ms bins

Figure A-3. Fifteen- and 19-month-olds' raw horizontal location of gaze during two target word repetitions in AusE, organized into 200 ms bins

Figure A-4. Fifteen- and 19-month-olds' raw horizontal location of gaze during two target word repetitions in JaME, organized into 200 ms bins

Figure A-5. Fifteen- and 19-month-olds' raw horizontal location of gaze during two target word repetitions in AusE, organized into 400 ms bins

Figure A-6. Fifteen- and 19-month-olds' raw horizontal location of gaze during two target word repetitions in JaME, organized into 400 ms bins

Figure A-7. Fifteen- and 19-month-olds' raw horizontal location of gaze during two target word repetitions in AusE, organized into 500 ms bins

Figure A-8. Fifteen- and 19-month-olds' raw horizontal location of gaze during two target word repetitions in JaME, organized into 500 ms bins

Figure B-1. Participant information form for Experiment 1

Figure B-2. Consent form for Experiment 1
Figure B-3. Participant information form for Experiments 2-4 (page 1 of 2)........256
Figure B-4. Participant information form for Experiments 2-4 (page 2 of 2)........257
Figure B-5. Consent form for Experiments 2-4...................................................... 258
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADS</td>
<td>Adult-directed Speech</td>
</tr>
<tr>
<td>AmE</td>
<td>American English</td>
</tr>
<tr>
<td>AOI</td>
<td>Area of Interest</td>
</tr>
<tr>
<td>AusE</td>
<td>Australian English</td>
</tr>
<tr>
<td>CanE</td>
<td>Canadian English</td>
</tr>
<tr>
<td>CG</td>
<td>Category Goodness assimilation</td>
</tr>
<tr>
<td>CS</td>
<td>Category Shifting assimilation</td>
</tr>
<tr>
<td>F0</td>
<td>Fundamental Frequency</td>
</tr>
<tr>
<td>FrE</td>
<td>French-accented English</td>
</tr>
<tr>
<td>IDS</td>
<td>Infant-directed Speech</td>
</tr>
<tr>
<td>JaME</td>
<td>Jamaican Mesolect English</td>
</tr>
<tr>
<td>MSS</td>
<td>Metrical Segmentation Strategy</td>
</tr>
<tr>
<td>OZI</td>
<td>Australian English Vocabulary Inventory</td>
</tr>
<tr>
<td>PAM</td>
<td>Perceptual Assimilation Model</td>
</tr>
<tr>
<td>PWC</td>
<td>Possible Word Constraint</td>
</tr>
<tr>
<td>seLE</td>
<td>Southeast London English</td>
</tr>
<tr>
<td>TE</td>
<td>Tyneside English</td>
</tr>
<tr>
<td>VOT</td>
<td>Voice Onset Time</td>
</tr>
</tbody>
</table>
Abstract

Word segmentation is qualitatively different between infants and adults, occurring in tandem with word recognition in adults, with words segmented as they are recognized from the speech stream. This requires a lexicon consisting of most commonly occurring words, which infants lack. Thus, infants are faced with the problem of needing to segment words to add them to the lexicon, but not having a lexicon large enough to segment words online as adults do.

Thus, early word segmentation occurs separately from word recognition. Infants segment words primarily by attending to the predominant metrical patterns in their language as cues to likely word boundaries, and can then learn the segmental patterns within those boundaries. In adults, attention to metrical patterns is a supplementary cue, and thus word segmentation by adults is robust to suprasegmental variation that can alter metrical patterns, such as that across foreign-accented speech.

Word recognition is also qualitatively different between young children and adults. Early word forms are comprised of strict phonetic patterns based on
experienced pronunciations that are intolerant to previously unexperienced variation. However, mature word forms are phonologically specified, comprised of the higher-order phonological form of the word above the level of phonemic variation. This allows a robustness to considerable segmental variation, even variation that violates native phonemic boundaries, as can occur across regional accents. Adults are generally able to recognize cross-accent pronunciations via their grasp of when variation does not change a word's underlying phonological form.

This thesis explores when these qualitative differences between immature and mature word segmentation and recognition begin to be resolved by young children. Experiment 1 examined when children can recognize words across both suprasegmental and segmental variation by testing 15- and 19-month-olds' ability to recognize words in their native regional accent (Australian English), and a non-native regional accent containing both suprasegmental and segmental variation (Jamaican Mesolect English). In a visual choice eye tracking task, 19-month-olds recognized words in both the native and the non-native regional accent, while 15-month-olds recognized words only in the native regional accent, suggesting that fully specified phonological word forms are present by 19 months, but not 15 months.

Experiment 2 addressed when children shed their primary reliance on attention to metrical patterns for word segmentation by examining their ability to segment words containing suprasegmental variation that altered the metrical patterning of words. In a serial preference task, Experiment 2 compared 12-month-olds' ability to discriminate words containing familiarized versus unfamiliarized words either in the native Australian English, or in a resynthesised accent that
combined native Australian English segmental details of the test sentences with prosody from French-accented productions of the same sentences. Twelve-month-olds preferred sentences containing unfamiliar words in the non-native accent, but not in the native accent, suggesting that the task was too difficult for 12-month-olds, causing them to fall back on their attention to native metrical patterns to discern word boundaries. While this directed attention away from segmental information in the native accent condition, in the non-native accent condition children were able to ignore the non-informative suprasegmental cues and access the segmental patterns of the familiarized words. This ability to segment words despite non-informative suprasegmental cues suggests a progression to a more mature word segmentation strategy in which words are segmented via recognition of the segmental patterns.

Experiments 3 and 4 examined when children develop phonologically specified word forms by comparing their ability to segment words from their native regional accent and a resynthesized accent in which the suprasegmental cues were from Australian English, but the segmental detail was from Tyneside English, which varies from Australian English primarily in its vowels (Experiment 3), or southeast London English, which varies primarily in its consonants (Experiment 4). In Experiment 3, which tested 12- and 15-month-olds, 12-month-olds did not discriminate in either condition, but 15-month-olds preferred sentences containing familiarized words in the non-native segments condition, in particular, words that had vowel variations that mapped to another vowel category in the native regional accent, suggesting a sensitivity to vowel changes in word forms by 15 months. In Experiment 4, which tested 12-, 15-, and 19-month-olds, 12-month-olds demonstrated a familiarity preference in the native regional accent, and 19-month-
olds demonstrated a familiarity preference in the non-native segments condition for words containing consonant variations that mapped to another consonant category in the native regional accent, suggesting a sensitivity to consonant changes in word forms by 19 months.

Taken together, these results suggest that attention to predominant metrical patterning is still available as primary segmentation tool at 12 months, but that this dependence is fading, as segmentation can occur when such cues are non-informative. Further, more fully specified phonological word forms appear by 19 months, but word forms may become phonologically specified for vowels earlier, at 15 months.
1 Overview

1.1 Motivation of the Current Research

Spoken word recognition and word segmentation from connected speech differ qualitatively between infants and adults. Adults segment words as they are recognized, hence word segmentation in this manner is referred to as *spoken word recognition*, or recognition of words in continuous speech (as opposed to recognition of isolated words), with words in the lexicon competing for recognition based on their match to the incoming segmental (consonant and vowel) information. Effective spoken word recognition thus requires a substantial and detailed lexicon. However, this is something that infants are lacking. Instead, in order to build up their lexicon, word segmentation by infants is based primarily on the metrical features of their native language, with the breaks suggested by the predominant metrical patterns used to determine likely locations for word boundaries – a cue that is supplementary in adults' spoken word recognition. Based on the evidence from
multiple strands of research, this thesis introduces the novel proposal that what is a single, integrated process in adults likely begins as two separate processes in young children, allowing word forms that are segmented by alternate strategies to be used to build a lexicon, which in turn can provide the foundation for adult-like segmentation. Once children have begun to build up their receptive lexicon, they can begin to combine the processes of word segmentation and word recognition in a way that closely resembles adult spoken word recognition, in which words are segmented as they are recognized.

While a substantial lexicon is a prerequisite for adult-like word segmentation, it is not sufficient on its own. Adult segmentation also differs from children’s in its robustness. Adults are regularly able to efficiently recognize words in continuous speech across degraded conditions (e.g., background noise or conversation), as well as across suprasegmental and segmental variation. It has been proposed that to accomplish this, adults must have abstract, phonologically specified word forms (e.g., Best, Tyler, Gooding, Orlando, & Quann, 2009). These word forms are able to accommodate phonetic variation that does not change a word’s underlying phonological structure (termed phonological constancy), but exclude phonetic variation that does change the word’s identity (phonological distinctiveness; Best et al., 2009). As well, accessing phonologically specified word forms is relatively cognitively untaxing, allowing for rapid spoken word recognition across otherwise demanding conditions.

Word forms do not start out this way. Early word forms are phonetically specified, containing much detail regarding the phonetic structure of the word. Recognizing such word forms is more cognitively demanding on the perceiver, due
to the entirety of detailed phonetic features that must be stored and accessed for recognition of a word. As word recognition occurs on the basis of these detailed phonetic features, word recognition using these phonetically specified word forms is intolerant of much variation, demonstrated by children's initial failure to recognize words across regional-accent variations from native pronunciations (e.g., Best, Tyler, Gooding, Orlando, & Quann, 2009).

Thus, the development of phonological word forms enables mature spoken word recognition. Phonetically specified word forms are too cognitively taxing and rigid to allow segmentation to occur primarily via the segmental structure of the word, and children must still rely heavily on more global cues used for word segmentation such as attention to meter (e.g., the stress foot in English). Phonological word forms, however, allow mature spoken word recognition in three ways: firstly, the reduced cognitive demand imposed by attention to higher-order phonological word forms rather than the larger multitude of structured phonetic features frees the cognitive resources that allow children (and adults) to segment words in a more cognitively demanding setting (e.g., \textit{in vivo} versus in a laboratory); secondly, phonological word forms allow for recognition of words across segmental variation, allowing for more robust word recognition; and thirdly, attunement to the higher-order forms of words allows for faster, more efficient word learning that strengthens spoken word recognition by increasing the likelihood that a spoken word will be in the perceiver’s lexicon, as well as increasing the perceiver's available contextual knowledge.

The aim of the research presented here is to explore the development in children of this robust, mature word segmentation. To address this, the thesis first
examines the course of development of fully specified phonological word forms. Previous research has examined the development of sensitivity to phonological distinctiveness extensively, but questions still remain as to when the complementary skill of phonological constancy emerges, which would clarify the overall understanding of when phonologically specified word forms emerge. Thus in Experiment 1 phonological constancy is explored holistically, examining when children can recognize words produced with both segmental and suprasegmental variation from their native regional accent. In Experiment 2, we examine when children are able to segment words via the segmental structure by examining when children are able to segment words that contain non-native suprasegmental variation that renders the metrical pattern non-informative. Early on children use attention to the metrical pattern of their language as a primary cue to possible word boundaries. Thus the ability to segment words in the face of non-informative metrical patterns would indicate a reduced reliance on this, and reveal the extent of ability to segment words via their segmental structure. Experiments 3 and 4 examine more specifically the development of phonological word forms by investigating children's ability to segment words from continuous speech in the face of segmental variation from their native regional accent with respect to vowels (Experiment 3), and consonants (Experiment 4), with the overall goal of answering the questions of when children are able to segment words with a robustness akin to that of adults.

1.2 Structure of the Thesis

Chapters 2 through 6 review the existing literature surrounding the research questions. Chapter 2 describes the development of speech perception in infants,
from their developing sensitivity to suprasegmental features of speech through to their language specific attunement to consonants and vowels. In particular, that chapter describes two accounts that attempt to explain the development of children's perception of segments: Exemplar-registration accounts and the perceptual attunement approach.

In this thesis the existing strands of research into infant word recognition and segmentation are reviewed separately, and are considered in the light of successful models of adult spoken word segmentation/recognition, with a focus on the progression to adult-like spoken word recognition. Accordingly, Chapter 3 describes the development of word segmentation in children. The chapter is broken down into a description of how children use statistical, segmental, and suprasegmental information to cue possible word boundaries, and how those features may interact with one another. The chapter concludes by describing how children begin to transition towards mature spoken word recognition.

Chapter 4 describes spoken word recognition in adults, including the robustness of spoken word recognition across segmental and suprasegmental variation. The latter half is devoted to describing past and current models that attempt to account for how spoken word recognition is achieved in adults.

Chapter 5 explores the differences between word segmentation in children and spoken word recognition in adults, focusing on the impact of children's limited lexicons, and on the fragility of child word segmentation in comparison to adults. The chapter ultimately points to the need for phonologically specified word forms in order to account for the robust nature of adult spoken word recognition.
Chapter 6 reviews the development of word forms in children, describing how initial word forms are phonetically specified and relatively intolerant of segmental and suprasegmental variation. The chapter then describes research to this point that attempts to outline children's developmental progression to more abstract phonologically specified word forms.

Chapter 7 describes the research questions of the present project in specific details, including prior results that need clarification, which will be addressed in Experiment 1. This chapter further relates the research questions in the context of the model of the development of word segmentation proposed here.

Chapter 8 details Experiment 1, which looks holistically at when children develop phonological word forms. It tests 15- and 19-month-olds' ability to identify illustrations of familiar words spoken in their native and a non-native regional accent. Together with past results this study completes the picture of when children develop phonologically specified word forms – that is, phonological constancy. Findings from this study suggest that children have developed a grasp of phonological constancy by 19 months.

Chapter 9 describes Experiment 2. To determine when children are able to shed their primary reliance on suprasegmental features for segmenting spoken words from connected speech, the study examines 12-month-olds' ability to recognize spoken words in a synthesized accent containing native segmental, but non-native suprasegmental information. Findings suggest that by 12 months, children are able to segment words in continuous speech across foreign-accented suprasegmental information via their ability to ignore the unreliable prosodic cues,
indicating that attention to suprasegmental features does not necessarily take on a primary role in spoken word recognition by that age.

Chapter 10 describes Experiments 3 and 4. These experiments together look at children's ability to use their phonological word forms to achieve robust spoken word recognition from connected speech, akin to that demonstrated by adults. Specifically, these studies look at 12-, 15-, and/or 19-month-olds' ability to recognize spoken words in continuous speech across a synthesized accent containing native suprasegmental information, but non-native segmental information that varies from the native regional accent. Children’s ability to overcome segmental variation to vowels (Experiment 3) and consonants (Experiment 4) are examined separately, and in particular they examine separately how children respond to phonetic vowel and consonant variation that results in phonemic changes in the listener's regional accent versus variation that does not. The results suggest a developmental progression in the specification of phonological word forms, with specification for vowels emerging by 15 months, and specification for consonants emerging by 19 months.

Chapter 11 is a general discussion of the results of experiments carried out in this thesis, including theoretical implications and avenues for further research. The results further inform the developmental progression of spoken word recognition as well as the development of phonologically specified word forms, suggesting a disparate developmental course for the phonological specification of vowels and consonants. The general discussion concludes with outline of a proposed developmental model of phonological word forms and word segmentation in children, and further theoretical directions.
2 The Development of Speech Perception

2.1 Sensitivity to Native Meter

Children's exposure to their language environment begins even before they are born. In utero, the fundamental frequencies (F0, or voice pitch) of maternal speech, and other speech spoken in proximity to the mother, pass through into the womb. From this input, children become familiar with the prosodic features of their language environment. For example, prenatal infants were read a passage by their mother every day between the 33rd and 37th week of gestation. When later exposed to a recording in utero of their mother reading the same passage, and a different one, the infants preferred listening to the familiarized passage over the other passage, indicated via a change in heart rate only when exposed to the familiar passage (DeCasper, Lecanuet, Busnel, Granier-Deferre, & Maugeais, 1994). When tested soon after birth, infants demonstrate a general familiarity with the metrical structure of their native maternal language. Broadly speaking, languages can be
categorized into three prosodic or metrical classes: syllable-timed, mora-timed, and stress-timed (see Ramus, Nespor, & Mehler, 1999, and Section 3.4.1.1 Metrical classes of languages, below, for a review). By 1-4 days old, infants discriminate passages in their native language from a language of a different metrical class (e.g., English-learning infants discriminate English, which is stress-timed, from Japanese, which is mora-timed; Christophe & Morton, 1998; and from Spanish, which is syllable-timed; Moon, Cooper, & Fifer, 1993). Moreover, newborns in a French (syllable-timed) language environment can discriminate a language from one non-native metrical class from a language from another non-native metrical class (English from Japanese), but cannot discriminate two languages belonging to the same non-native metrical class (English from Dutch; Nazzi, Bertoncini, & Mehler, 1998). In these cases, discrimination also occurs when the speech has been low-pass filtered so that only the prosodic information is available, indicating discrimination was based on children’s attention to prosodic details, and not as a result of attention to segmental details (Mehler et al., 1988; Nazzi et al., 1998). Children's awareness of the metrical properties of their native language becomes more fine-tuned, such that by 5 months infants discriminate their native language from other languages (or regional accents) even within the same metrical class (i.e., American-English-learning infants discriminate American-English from Dutch and from British-English, which are all stress-timed; Nazzi, Jusczyk, & Johnson, 2000). However, recent evidence suggests this within-metrical category discrimination between one's native language and regional accent variations diminishes by 9 months, as children come to recognize that the two accents both reflect their native language, despite phonetic dissimilarities (Kitamura, Panneton, & Best, resubmitted). Nonetheless,
these findings suggest that children are sensitive to the metrical features of their language from a very early age, which in turn suggests this aspect of spoken language is available to them as a tool for word segmentation.

2.2 Attunement to Native Consonants and Vowels

Children are born able to discriminate nearly all consonant and vowel contrasts, regardless of whether or not the contrasts are found in their native language environment (e.g., Aslin & Pisoni, 1980; for reviews, see Best, 1994a; Burnham, 1986; Werker & Tees, 1999). Infants brought up in an English-speaking environment can discriminate the phonetic contrast [pa] and [ta] (unaspirated, short-lag voice onset time [VOT] stops), which corresponds to the phonemic contrast /b/-/d/ in English, but can also discriminate the Hindi dental [ta] and retroflex [ʈa] stops despite the fact that English does not use this as a phonemic contrast (Werker, Gilbert, Humphrey, & Tees, 1981). Infants can also discriminate phonetic contrasts such as the ejectives [k’i] and [q’i] that are found in Salish, but neither of which are found in English (Werker & Tees, 1984), and can also discriminate prevoiced [b] and short-lag unaspirated [p], which can both appear in English, but which adults perceive as the single phoneme /b/ (Aslin, Pisoni, Hennessy, & Perey, 1981). This (near-) universal ability of young infants is shared across quail (Kluender, Diehl, & Killeen, 1987), chinchillas (Kuhl & Miller, 1975), macaques (Kuhl & Padden, 1982) and rhesus monkeys (Morse, Molfese, Laughlin, Linnville, & Wetzel, 1987), suggesting that rather than being specific to language processing, it more likely reflects a prelinguistic or nonlinguistic capacity grounded in more basic perceptual abilities.
However, within the first year of life, children become perceptually tuned to the contrasts found within their native language environment. Young infants brought up in an English language environment can discriminate the Hindi contrast [ṭa]-[ṭa] and Salish contrast [k’i]-[q’i], but by 10-12 months their ability to discriminate such consonant contrasts has declined, and continues to do so until, like English-speaking adults, they are no longer able to reliably discriminate many contrasts that are not present in their native language environment. By the same token, children brought up in Hindi or Salish language environments continue to discriminate the contrasts present in their native languages, as do Hindi-speaking and Salish-speaking adults, while losing the ability to discriminate most contrasts that are relevant only to the English language (Werker & Tees, 1983, 1984). Thus consonant perception develops to be categorical, meaning that clear perceptual boundaries exist between phones that are contrastive in the native language, such as between [p] and [b] in English. That is, when presented with a continuum of synthetic phones that change on a gradient of equal steps of acoustic variation from long-lag aspirated VOT [pʰ] to short-lag unaspirated VOT [p], there will be a definitive boundary at which phones preceding the boundary will be perceived as the phoneme /p/, but phones occurring after the boundary will be perceived as the phoneme /b/. Likewise, discrimination for non-contrastive (within-category) phones is unreliable, as in the case described above in which adults were unable to discriminate [b] from [p], instead perceiving them as two instances of /b/ (Aslin et al., 1981). Notably, this developmental perceptual pattern is not specific to speech perception, but is seen in other perceptual domains such as vision, for instance in face perception (e.g., Kotsoni, de Haan, & Johnson, 2001).
It should be noted though that some contrasts do not show an age-related decline, despite their absence in the native language. Best and McRoberts (2003) found that both 6- to 8-month-olds and 10- to 12-month-olds discriminated the Tigrinya ejective contrast [p’]-[t’], despite that it is a non-native contrast. It is believed that this is because 10- to 12-month-olds retained categorical discrimination of [tʰ]-[pʰ] in their native language, and that they perceptually mapped both elements of the Tigrinya contrast onto those contrasting native language categories (Best & McRoberts, 2003; see also Best, 1995).

For vowels, native perceptual attunement begins even earlier. As with consonant perception, young infants are able to discriminate vocalic contrasts that are linguistically relevant to their native environment, as well as those that are not. English-learning 4-month-olds can discriminate the native vowel contrast /i/-/a/ as well as the non-native German vowel contrasts back versus front rounded /ʊ/-/ʏ/ (lax/short vowels) and /u/-/y/ (tense/short vowels), but the ability to discriminate the German vowel contrasts declines by 6-8 months, and infants are no longer able to reliably discriminate these non-native vowel contrasts by 10-12 months (Polka & Werker, 1994). But unlike consonant perception, this attunement to linguistic relevance at 10-12 months is only temporary. Vowel perception by adults goes on to become less categorical and more continuous, with adults recovering the ability to discriminate non-native vowel contrasts (Beddor & Strange, 1982; Polka & Bohn, 1996; Stevens, Liberman, Studdert-Kennedy, & Öhman, 1969), and with identification of vowels on a continuum suggesting less sharp category boundaries than for consonant perception (Eimas, 1963; Fry, Abramson, Eimas, & Liberman, 1962).
2.3 Theories of the Development of Speech Perception

2.3.1 Exemplar accounts. Exemplar accounts propose that our general ability to learn via tracking statistical patterns in the input extends to language learning (e.g., K. Johnson, 1997; Pierrehumbert, 2003). By this view, phonemic categories emerge from infants' exposure to the multidimensional phonetic properties of the native speech signal. As exposure increases, certain regularities emerge that correspond to phonemes, or phonetic categories that sufficiently correspond to phonemes. Stronger representations emerge from more extensive and more recent exposure to members of that category, and a perceived phoneme will activate the categories it is perceptually nearer to. These categories compete via both strength of activation and mutual inhibition, which might explain the phenomenon of categorical perception. In this way, phonemes are abstracted from repeated exposure to various exemplars from that category. Initial phonemic development is bottom-up, and becomes increasingly language-specific, or even regional-accent-specific, as phonetic properties that are uninformative or random are eventually minimized in the collective memory trace of the exemplars the child has experienced.

Such approaches are supported by evidence in infants as well as adults. In a series of artificial language learning studies, 6- and 8-month-olds in an English language environment (Maye, Werker, & Gerken, 2002), and adult native English speakers (Maye & Gerken, 2000) were exposed either to a unimodal or a bimodal distribution of tokens along a continuum from prevoiced [da] to unaspirated [ta], which are not contrastive in English. In subsequent discrimination tasks, only those infants and adults who had been exposed to the bimodal distribution were able to
discriminate the contrast. The language-specificity of such an approach is indicated by cross-linguistic studies that can relate people's inability to distinguish a non-native contrast with their inability to attend to a natively uninformative phonetic property. For example, it may be that Japanese speakers' inability to reliably discriminate the English contrast /ɹ/ and /l/, which differ most reliably on F3, is due to their non-attendance to F3, as it does not distinguish any contrasts in Japanese (Strange & Dittmann, 1984; see also Iverson et al., 2003).

2.3.2 Perceptual Atunement Approach. The perceptual attunement approach is derived from the Perceptual Assimilation Model of language-specific effects on speech perception (PAM; Best, 1994b, 1995). It is a direct realist approach, modeled on the ecological theory of perception (Gibson, 1966), which posits that the objects of perception are external events, rather than internal representations. In this way speech events are detected directly, rather than resulting from the cognitive interpretation of raw acoustic data. This is analogous to the direct realist view of visual perception in which we do not perceive reflected light that is then interpreted as a given object, but rather we perceive the object as the source of that reflected light (Gibson, 1972). By this view, the objects of speech perception are not acoustic features or patterns, but are instead dynamic articulatory gestures arising from the speech articulators (e.g., vocal folds, velum, tongue, lips) that are responsible for having generated the multimodal speech signal. This view is supported by research on cross-modal speech perception. The McGurk effect (McGurk & MacDonald, 1976) occurs when a person is presented with an auditory syllable (e.g., bilabial stop, [pa]), synchronized with a visual syllable of the same manner, but a different place of articulation (velar stop, [ka]), and perceives an
intermediate place (alveolar stop, [tə] or interdental fricative, [ðə]) rather than perceiving the syllable presented in either modality. This suggests that rather than attending to acoustic features or visual speech, people instead perceive the articulatory gesture that (appears to have) created that speech, and occurs in clear and non-conflicting auditory and visual conditions. Even more strikingly, the McGurk effect also occurs when people are presented with an auditory syllable paired with a non-congruent syllable presented haptically (i.e., the perceiver feels the speaker’s lip movements with their fingertips; Fowler & Dekle, 1991). While adult perceivers have extensive experience with both auditory and visual speech, the same cannot be said for haptic speech perception. The successful integration of the haptic-auditory speech provides further support for the notion that speech events are perceived as articulatory gestures, rather than as acoustic patterns.

Developmentally, the perceptual attunement approach posits that speech perception follows a perceptual learning course (see also Gibson, 1966, 1979; Gibson & Gibson, 1955) in which infants' initial perception of these speech events is phonetically-based, accounting for their ability to discriminate most native as well as non-native contrasts (e.g., Aslin & Pisoni, 1980; Burnham, 1986; Werker & Tees, 1999). Infants undergo perceptual learning with experience, and soon discover the abstract, systematic features in the language input that give rise to the coordinated articulatory gestures within the phonological space of their language, which correspond functionally more or less to phonemes in their language. In tandem with this process infants learn to ignore irrelevant phonetic features for that language, enabling speech perception to become more efficient and optimized to their language environment.
This is distinct from exemplar accounts of speech learning in that children do not develop phonological knowledge via tracking the statistical distribution of instances of sounds, but instead via exploration of unimodal and multimodal speech input. Through this they discover the critical invariants of gestural patterns within their native language so that, for example, they are able to recognize that \( [p^b] \) as spoken by a man and a woman they've never met before are both the phoneme /p/ because both speakers produce the sound by a gesture of lip-closure and release, with laryngeal devoicing. They have accomplished the same articulatory gesture, despite the acoustic variance, and not simply because both events fall within the realm of previously experienced instances of /p/, as in exemplar accounts.

In both exemplar accounts and the perceptual attunement account, the findings on infant segmental speech perception demonstrate the development of a skill whereby infants learn to attend to features that distinguish the phonemes of their language, and to ignore features that are not relevant within their native language. Furthermore, this is accomplished in the first year of life, and this process of attunement may be reflected in later developments regarding word recognition. Most especially, it is reflected in the children's ability to segment words from connected speech, the issue addressed in the next chapter.
3 Development of Word Segmentation

3.1 Word Segmentation as a Precursor to Word Learning

In parallel with the development of early speech perception, in their first year infants must begin building a lexicon. However, in order to learn words in their language, children must first reliably separate words from running speech so that they can then learn the word forms and attach meanings to these words. This is a particularly difficult task when one is reminded that running speech does not have word boundaries marked by pauses or other acoustic cues (Cole & Jakimik, 1980; Libermann, F. S. Cooper, Shankweiler, & Studdert-Kennedy, 1967), and neither adult- nor infant-directed speech generally involves presenting words in isolation (Brent & Siskind, 2001; Woodward & Aslin, 1990). This task is also conceptually daunting when one is reminded that adults primarily accomplish this in tandem with word recognition, such that a word is segmented as it is recognized (e.g., McClelland & Elman, 1986; Norris, 1994; Norris & McQueen, 2008; see Chapter 4
Spoken Word Recognition in Adults). Thus children are faced with the paradoxical task of needing to build a lexicon so that they can segment the speech stream, but must first segment the speech stream to build their lexicon.

Instead, it appears infants attend to a variety of cues to discern likely word boundaries early on. These cues bootstrap the building of their lexicon, providing them with a means to separate out their first words and associate these words with concepts.

3.2 Cues to Word Boundaries

3.2.1 Tracking transitional probabilities of syllables. Transitional probabilities (TPs) are one such cue that can inform possible word boundaries. A high-frequency sequence of syllables is more likely to be within one word, while a sequence that does not often occur is more likely to occur across a word boundary. For example, in the phrase “pretty baby,” the transitional probability that “pre” precedes “ty” is higher than that of “ty” preceding “ba”. Thus, a dip in transitional probabilities between syllables could provide evidence of a word boundary (Saffran, Aslin, & Newport, 1996), and studies show that children are able to track and make use of these distributions (Friederici & Wessels, 1993; Goodsitt, Morgan, & Kuhl, 1993; Jusczyk, Luce, & Charles-Luce, 1994; Mattys & Jusczyk, 2001a; Mattys, Jusczyk, Luce, & Morgan, 1999; Saffran, Aslin, et al., 1996). In artificial language studies in which the only cues to word boundaries within a string of nonsense syllables presented during a familiarization period are the TPs across syllables, infants as young as 5.5 months are able to utilize TP information to segment the target words from the continuous speech stream (E. K. Johnson & Tyler, 2010),
demonstrated by their discrimination of statistically possible words (as determined by tracking TPs in the input) from statistically impossible words in a subsequent listening preference task. This ability extends into adulthood as well (Saffran, Newport, & Aslin, 1996). While statistical properties can be a strong cue to word boundaries, in addition to these features, cues to word boundary locations can come from the segments themselves.

3.3 Use of Segmental Features as Cues to Word Boundaries

In speech, segments are discrete units of speech, such as consonants and vowels. Combinations of segments (phonotactics and coarticulation) or features of segments (allophones) can serve as cues to possible word boundaries.

3.3.1 Phonotactics. Phonotactics are the rules that govern permissible sequences of phonemes and where they can appear in words in a given language. As with learning of transitions probabilities of syllables, infants acquire this knowledge via statistical learning, and can use it to help narrow down locations of possible word boundaries (Mattys & Jusczyk, 2001a, 2001b; Mattys et al., 1999). Sequences of phonemes that violate native language phonotactics appear only across syllable and/or word boundaries. For example, the sequences /dl/ or /tl/ cannot occur in syllable-initial position in English, but regularly occur across syllable boundaries, as in ANTLER or ATLANTIC. Similarly, the sequence /mdpr/ cannot occur within a syllable or across syllable boundaries, but can occur across word boundaries, for example in JAMMED PRINTER. By 9 months children are sensitive to the phonotactic rules of their native language, shown via their preference to listen to lists of phonotactically valid non-words over phonotactically
invalid non-words (Jusczyk, Luce, et al., 1994a). They have also been shown to
differentiate phonetic patterns found within words from phonetic patterns found
only across word boundaries (Mattys, Jusczyk, et al., 1999d), and can implement
that sensitivity to aid in word segmentation by recognizing phonetic patterns that
occur across words as possible word boundaries (Mattys & Jusczyk, 2001b).

3.3.2 Allophones. Allophones are different, often position-dependent
phonetic realizations of a given phoneme, and these variations can also help cue
word boundaries (Jusczyk, Hohne, & Bauman, 1999). Some allophones occur only
at certain lexical and/or syllabic positions. For example, in American English /k/ is
realized as the allophone [kʰ] word-initially, but as [k] following /s/ or /ʃ/ or in
stress- or word-final position (e.g., ACTIVE [ˈaktɪv] or TACK [tæk]). Sensitivity to
the systematic distribution of these allophonic variations thus uncovers possible
word boundaries. At 10.5 months, children are sensitive to allophonic variation, via
their ability to discriminate pairs such as NITRATE and NIGHT RATE in
continuous speech, in which the two pairs are identical save for the allophonic
variation between the instances of /t/ in each word, with the realization of /t/ in
nitrate having a longer burst and closure duration and presence of aspiration ([tʰ] as
compared to the unaspirated, unreleased /t/, [t̪]) in NIGHT of NIGHT RATE
(Jusczyk, Hohne, et al., 1999).

3.3.3 Coarticulation. Coarticulation occurs when the pronunciation of a
given segment in turn affects the pronunciation of an adjacent (either preceding or
following) segment(s). For example, in the word KEEP, /k/ is produced farther
forward on the palate compared to in the word COOP, in which /k/ is farther back
and thus truly velar, that is, at the soft palate (velum). In other words, the tongue
body closure for /k/ in KEEP shows anticipation of the tongue body’s position for the high front vowel, /i/, versus anticipation for /u/, a high back vowel in COOP. Coarticulation occurs primarily within words, and is less common across word and phrase boundaries (Fougeron & Keating, 1997). Adults are able to exploit this tendency to cue possible word boundaries (Mattys, 2004; Mattys, White, & Melhorn, 2005), and there is evidence that children are able to exploit this information as well. To simulate decreased coarticulation, E. K. Johnson and Jusczyk (2001) spliced artificial words into running nonsense speech and found that 8-month-olds were sensitive to those spliced boundaries and treated them like word boundaries, even though the boundaries conflicted with those suggested by tracking transitional probabilities of syllables.

3.4 Use of Suprasegmental Features as Cues to Word Boundaries

Infants also make use of suprasegmental cues in the speech stream to detect word boundaries. Suprasegmental features in speech are those that occur across segments, such as the meter, stress patterning, and intonation of an utterance. These properties are abundantly available to infants, as suprasegmental features are exaggerated in infant-directed speech (IDS). Compared to adult-directed speech (ADS), IDS has a higher average fundamental frequency (F0), wider F0 range, more variable pitch inflections (Fernald & Kuhl, 1987; Kuhl et al., 1997), and greater prosodic repetition (Fernald & Simon, 1984). Infants have been shown to attend to these cues, for example via their preference for prosodically exaggerated IDS over ADS (R. P. Cooper & Aslin, 1990; Fernald, 1985; Pegg, Werker, & McLeod, 1989), their ability to discriminate languages belonging to different
metrical classes (e.g., Mehler et al., 1988), their sensitivity to clause boundaries (e.g., Hirsh-Pasek et al., 1987) and phonological phrase boundaries (Gout, Christophe, & Morgan, 2004; Soderstrom, Seidl, Kemler Nelson, & Jusczyk, 2003), and their preference for the predominant metrical pattern within their native language (e.g., Houston, Jusczyk, Kuijpers, Coolen, & Cutler, 2000; Jusczyk, Cutler, & Redanz, 1993; Jusczyk, Houston, & Newsome, 1999). By using suprasegmental information as a cue to word boundaries, along with other IDS features such as simplified syntax and shorter words (Kuhl et al., 1997; Stern, Spieker, Barnett, & MacKain, 1983), infants are able to ease the complicated task of word segmentation by reducing the number of possible words into a much more manageable set of those fitting within the suprasegmental boundaries (e.g., Shukla, Nespor, & Mehler, 2007).

3.4.1 Use of metrical patterns as cues for word boundaries.

3.4.1.1 Metrical classes of languages. The metrical properties of a language provide the primary suprasegmental cues for informing word boundaries. Intuitively, one notices that different languages can have different qualitative meters compared to other languages. These metrical qualities of some languages have been likened to a “machine-gun meter” versus a “Morse code meter”, the first describing languages such as Cantonese and Italian, and the second describing languages such as English and Dutch (Lloyd James, 1940). Since then, those metrical classes have been referred to as syllable-timed and stress-timed, respectively. Stress-timed languages tend to be Germanic languages such as English, Dutch, and German, though not exclusively, as (for example) European Portuguese and some southern accents of Italian are also purported to be stress-timed. These languages can display
vowel reduction in unstressed syllables (e.g., replacing the vowel in a syllable with /ə/; Bertinetto, 1981; Dauer, 1983) and have a wide variety of syllable types (i.e., combinations of consonants and vowels; Dauer, 1983). Syllable-timed languages tend to encompass the Romance languages, such as Italian and French, have a more limited syllable type variety and typically lack vowel reduction. A third metrical class, mora-timed, includes Japanese, Gilbertese, and Luganda. A mora is a timing unit smaller than a syllable. It is possible for more than one mora to comprise a syllable, though definitions of what constitutes a mora can vary across languages. In Japanese, a mora is any consonant-vowel unit, or an individual vowel or nasal. For some languages there is debate as to which metrical class they belong. For example, Polish and Catalan have been classified at different times as stress-timed and syllable-timed (Hayes & Puppel, 1985; Mascaró, 1976; Rubach & Booij, 1985; Wheeler, 1979). This is because Polish, while typically regarded as stress-timed, is lacking in vowel reduction, which is a quality of many syllable-timed languages. Catalan is generally classified as syllable-timed, but does display vowel reduction as in stress-timed languages. This debate leaves open the possibility that additional metrical categories may exist (Levelt & van de Vijver, 1998), or that languages should be classified on a continuum rather than categorically (Dauer, 1987).

Originally, it was hypothesized that the different metrical qualities between languages in the various metrical classes were due to temporal constants that varied across metrical classes (e.g., Abercrombie, 1967). That is, in syllable-timed languages it was believed that each syllable carried roughly the same duration. In stress-timed languages, it was believed that every stress unit (a strong syllable
followed by any unstressed syllables) was roughly equal in duration, and likewise in mora-timed languages, it was believed that each mora carried roughly the same duration. However, subsequent empirical investigation was unable to substantiate this claim (e.g., Pamies Bertrán, 1999), and in fact ample evidence against this claim has been found. The durations of intervals between stress units in stress-timed languages, rather than being isochronous, are directly proportional to the number of syllables they contain (Bolinger, 1965; Lea, 1974; O’Connor, 1965; Shen & Peterson, 1962). Further, syllable duration has not been found to be isochronous in either Spanish (Borzone de Manrique & Signorini, 1983) or French (Wenk & Wiolland, 1982), with the variation of syllable durations being similar to that found in stress-timed languages (Roach, 1982).

Instead, the qualitative differences between languages from the three metrical classes may stem from differences in the proportion of vocalic intervals (a single vowel or string of consecutive vowels) to consonantal intervals (a single consonant or string of consecutive consonants) and variability in consonantal intervals. Acoustic analysis by Ramus, Nespor, and Mehler (1999) suggests that stress-timed languages tend to have a relatively low proportion of vocalic intervals, and greater variability in the consonantal intervals, while syllable-timed languages have a greater proportion of vocalic intervals, and less consonantal interval variability. Japanese was the only moraic language included in the analysis, but it was shown to have the greatest proportion of vocalic intervals, and the least variable consonantal intervals. It is reasonable that these variable segmental combinations result in production differences leading to the impression of qualitatively different metrical classes, corresponding to the naïve “machine-gun”
or “Morse code” descriptions. In addition, it is proposed that sensitivity to these variations may account for infants' early ability to discriminate across metrical classes (as discussed in Section 3.4.1.2 Sensitivity to native metrical patterning: The Metrical Segmentation Strategy (MSS); see Ramus et al., 1999), though it has not yet been posited that perceivers use such a sensitivity to segmental ratios as a cue to word boundaries. Still, as will be discussed below, the metrical properties of a language can serve as an informative cue for possible word boundaries, with attention focused on the boundaries between metrical units, and predominant metrical patterns.

3.4.1.2 Sensitivity to native metrical patterning: The Metrical Segmentation Strategy (MSS). The metrical structure of a language can aid word segmentation by narrowing the likely location of possible word boundaries down to those that fit within metrical units. For stress-timed languages, word boundaries naturally occur between stress units (a strong syllable followed by one or more weak syllables). In syllable-timed languages, word boundaries occur between syllables, and in mora-timed languages, word boundaries occur between morae. Adults are sensitive to the metrical boundaries of their native language (e.g., Cutler & Butterfield, 1992; Cutler, Mehler, Norris, & Segui, 1986; Cutler & Norris, 1988; Cutler & Otake, 1994; Mehler, Dommergues, Frauenfelder, & Segui, 1981; Otake, Hatano, Cutler, & Mehler, 1993; Otake, Yoneyama, Cutler, & van der Lugt, 1996), but are not sensitive to the metrical boundaries of languages belonging to another metrical class. Thus, when listening to non-native speech, adults will listen for the metrical units of their native language, even if the language they are listening to is of a different metrical class (Cutler et al., 1986; Cutler & Otake, 1994; Otake et al.,
1993, 1996). For example, French listeners are sensitive to the boundaries between syllables when listening to speech, as evidenced by their ability to be better primed in a word spotting task by the first syllable of a word, versus the first syllable plus or minus the adjacent phoneme (BA primes BA'LANCE better than BAL'CON, and vice versa for BAL; Mehler et al., 1981). French listeners also show this priming effect even for English words, but interestingly, English listeners do not, and are instead primed equally by the first syllable of a word or the first syllable plus or minus the adjacent phoneme. This is because English is not syllable-timed and therefore does not have as well defined syllable boundaries. In fact, in English, phonemes transcend syllable boundaries in some words (i.e, such phonemes are ambi-syllabic) – for example the /l/ in BALANCE belongs to both the first and second syllable. Thus, while French listeners attend to syllable boundaries when listening to any speech, as it is informative in their native language, English listeners are not sensitive to this relatively uninformative cue (Cutler et al., 1986).

The features that make up these metrical patterns can differ across languages and dialects within a given metrical class. For instance, Dutch and English are both classified as stress-timed, and compared to unstressed syllables, stressed syllables in both languages are longer in syllable duration, contain greater pitch movement, and may be louder (Cooper, Cutler, & Wales, 2002). Native Dutch listeners use these cues to discern stressed and unstressed syllables (and thus metrical patterns). In English, however, stressed syllables are also marked by full vowels, whereas unstressed syllables contain reduced vowels, adding a segmental cue of vowel quality to stress perception. The cues that native speakers of a given language actually attend to in order to perceive the metrical pattern can differ across
languages within the same metrical class. Native Dutch speakers use the
aforementioned suprasegmental cues to discern stressed and unstressed syllables,
whereas for English speakers, vowel quality is a more immediate cue to stress,
despite the available suprasegmental cues. Thus, when syllables are presented in
isolation, English-speaking adults are essentially perceptually deaf to the
suprasegmental cues of stress at the syllable level in an absolute sense, instead
attending to the segmental cues. However, English speakers are sensitive to the
relative change in suprasegmental information from one syllable to the next, so that
when presented with more than one syllable, English speakers are able to discern
stress based on suprasegmental information (N. Cooper et al., 2002).

Attention to these metrical structures to aid word segmentation in one's
native language is referred to as the Metrical Segmentation Strategy (MSS; Cutler
& Norris, 1988). It is supported by evidence that shows adults implement this
strategy to aid in word segmentation (Cutler & Butterfield, 1992; Cutler & Norris,
1988; McQueen, Norris, & Cutler, 1994; Vroomen & de Gelder, 1995). For
example, in a word spotting task, English listeners were faster at spotting a
monosyllabic real word (e.g., MINT, note that capital letters in this case denote
stressed syllables) embedded in a bisyllabic nonsense word when the added syllable
was unstressed, having a reduced vowel in the unstressed syllable (MINTef),
compared to when the added syllable was also stressed (MINTAYF). It was
reasoned that this was because the English listeners were sensitive to the boundaries
between stressed units as locations for word boundaries, and so had segmented
MINTAYF into two words, whereas they had kept MINTef as a single word.
Spotting MINT in MINTAYF therefore required recombining the word, which took
a longer processing time, accounting for the slower reaction time (Cutler & Norris, 1988).

Thus, the metrical pattern of language has correlates in speech that can be and are used by adult native listeners to aid in word segmentation. How one develops the use of these features is discussed below.

### 3.4.2 Children's use of the MSS

Infants begin to attend to their language-specific metrical cues at 6 months (Morgan, 1996), and by 7.5 months rely heavily on those cues for segmenting words from connected speech (Jusczyk, Houston, et al., 1999). The majority of polysyllabic words in English are disyllabic, and have a *trochaic* stress pattern (a stressed syllable followed by an unstressed syllable; Cutler & Carter, 1987). English-learning infants segment words based on a trochaic pattern, so much so that words not following this pattern are mis-segmented. When presented with the spoken phrase, “guiTAR is”, 7.5-month-olds incorrectly treated TARis as a word (Jusczyk, Houston, et al., 1999). This finding has been found across languages – Canadian-French-learning infants segment words based on *iambic* stress patterns (an unstressed syllable followed by a stressed syllable), the principal stress pattern in that language (Polka, Sundara, & Blue, 2002). While children continue to segment words via the predominant metrical pattern of their native language at 9 months (Jusczyk et al., 1993), as early as 10.5 months, infants begin to segment words that do not demonstrate the predominant metrical pattern in their language (E. K. Johnson, 2005; Jusczyk, Houston, et al., 1999), suggesting a progression toward a more adult-like word segmentation method.
3.5 Relationship of Statistical and Suprasegmental Cues to Word Boundaries

Evidence from studies that pit two types of word boundary cues against each other suggests that when children first begin to segment words from running speech, at around 7 months, this early word segmentation relies on distributional cues to clue likely word boundaries. But within the span of two months this changes to a favoring of the MSS. When 7- and 9-month-old English-learning infants were exposed to an artificial language in which the transitional probabilities and predominant metrical pattern each indicated different segmentations, 7-month-olds segmented the words based on differences in transitional probabilities across syllables, whereas the 9-month-olds segmented the words that adhered to the predominant metrical pattern of their language (Thiessen & Saffran, 2003). Similarly, while 9-month-olds are sensitive to phonotactic cues to possible word boundaries (that is, they can pick out clusters that are unlikely to occur within words as a possible boundary between words), they do not use these cues if they are in opposition to those suggested by the metrical properties of the language, instead segmenting words on the basis of meter (Mattys et al., 1999). Thus, while transitional probabilities across syllables and attention to segmental correlates of word boundaries may play a dominant role in the earliest beginnings of word segmentation, once the child is tuned to the metrical patterning of their native language at about 9 months, this becomes the most informative cue, with other cues likely taking on a supplementary role. However by 10.5 months, infants are able to segment words that do not demonstrate the predominant metrical pattern in their native language. After being familiarized to iambic words, such as guiTAR, 10.5-
month-old English-learning infants presented with the phrase “guiTAR is” embedded in continuous speech were able to segment the word guiTAR, where as younger children mis-segmented the word TARis due to their dependence on segmenting based on a trochaic stress pattern. That the 10.5-month-olds no longer rely solely on a trochaic stress pattern for segmentation indicates that they have reached a stage where they are able to integrate multiple cues to better detect probable word boundaries (Jusczyk, Houston, et al., 1999).

3.6 Possible Word Constraint

The Possible Word Constraint (Norris, McQueen, Cutler, & Butterfield, 1997) is a universal restriction in word segmentation in which candidate words in a continuous speech stream may only be considered as possible words if segmentation of those words does not leave a residue that itself cannot be a word. Any remnant must therefore consist of at least a valid syllable (or mora, in moraic languages) in that language, though it is not necessary that the syllable actually carry a lexical meaning. Typically, this means that a word cannot be segmented if segmentation of that word would leave consonants in isolation. For example, in the input WIND, the word WIN is found within the input, but cannot possibly be segmented, as it would leave D, which cannot be a word. This constraint limits the number of competitor words one must deal with when segmenting the speech stream. Importantly, this constraint allows reduction of competitor words independent of a person's lexicon, as it reduces words on the basis of universal impossibility, rather than on the basis of having no lexical entry for a possible word. Thus the possible word constraint is a valuable tool for infants undergoing
early word segmentation, as they have only a minimal lexicon but must segment words nonetheless in order to build their vocabulary. Twelve-month-olds have been shown to employ the possible word constraint in situations that require them to segment words from fluent speech (E. K. Johnson, Jusczyk, Cutler, & Norris, 2003). Here, infants familiarized to target words (e.g., RUSH) listened longer to lists of nonsense words in which segmentation of the target word did not violate the possible word constraint (NIPRUSH), versus lists of words in which it did (PRUSH).

3.7 Transition to Adult-like Word Segmentation

Eventually, children build a lexicon that is substantial enough so that word segmentation can occur primarily in tandem with word recognition, in which the number of unknown words in the speech signal is small enough such that those words can be segmented via the segmentation of surrounding known words. Thus word segmentation can become less reliant on the extralexical cues to word segmentation, instead focusing primarily on the relationship of the speech input to lexical entries. There is evidence that children can do this early on. From birth children receive significant exposure to their own name. By 6 months children are able to segment their own name from running speech (Mandel, Jusczyk, & Pisoni, 1995), and can use their ability to segment their name to cue the boundary of (and thus segment) an adjacent novel word (Bortfeld, Morgan, Golinkoff, & Rathbun, 2005). Despite this ability, children's reliance on primarily bottom-up cues to word segmentation persists. However, their ability to segment speech based on the possible word constraint appears to depend not on reaching a developmental
milestone, but a milestone in terms of the size of their lexicon. Furthermore, it should be kept in mind that while children may come to rely less on other cues to word segmentation, these cues may still be available to them and used when needed. Lastly, these results describe segmentation in a native speech environment, and do not consider segmentation when presented with grosser variations in segmental and suprasegmental information.
4 Spoken Word Recognition in Adults

While the previous chapter addressed what cues infants use to segment words from continuous speech, and how their use of those cues develops, the aim of the thesis is to examine when children transition from use of the immature strategies described in the previous chapter, to the mature strategies for spoken word recognition used by adults, described in this chapter. Adults are able to use many sources of information to segment words from the speech stream. Some sources, such as syllabic distributional information and the possible word constraint, overlap with those already used by infants in the first year of life. However, word segmentation by adults is clearly faster and more reliable than that of infants. What is it, then, that differentiates early from mature word segmentation? Before this question is more directly addressed in the following chapter, the characteristics of adult spoken word recognition are reviewed below to provide a basis for comparison between immature word segmentation and mature spoken word recognition.
4.1 Characteristics of Adult Spoken Word Recognition

4.1.1 Spoken word recognition is continuous. Spoken word recognition is a continuous process, as a result of the nature of the input. As the speech signal unfolds, possible words and segmentations are considered in parallel and continually updated. The continuous nature of speech recognition and segmentation is evidenced by studies showing that a segment substitution at the beginning of a word disrupts lexical access to that word more than a segment substitution at the end of a word. In a priming experiment, Dutch speakers exposed to the word HONING (honey) showed a faster lexical response to the semantically-related word, BIJ (bee), but did not show this priming for the segmentally-altered related word, WONING (dwelling). This indicates that HONING did not result in access of WONING, despite the segmental similarity between the two words (Marslen-Wilson & Zwitserlood, 1989). Further, in a gating task in which participants heard fragments of words and were asked to identify from four pictures which item was being named whilst their eye gaze was recorded, words that shared segments at their onsets but differed segmentally at the end of the word (e.g., BEETLE and BEAKER) showed equal lexical activation via equal looking to both items up until the point of disparity at which the correct item was identified (Allopenna, Magnuson, & Tanenhaus, 1998). This reflects the continuous temporal processing of speech in word segmentation and recognition, as the sooner there is a segmental mismatch between candidate words and the incoming input, the sooner a word is disadvantaged as a possible candidate (McQueen, 2007). Were it that spoken word processing was discontinuous, and instead words were perhaps segmented first and then recognized after the fact, then the location of segmental variation would be
irrelevant, as the presence of any segmental variation would immediately and categorically disadvantage a word as a possible candidate.

4.1.2 Primarily driven by segmental information. Segmental information is the predominant information source in adult spoken word recognition. Adults rely heavily on these cues – the alteration of a single segment can disrupt or even block recognition of a word. While any segmental alteration can disrupt lexical access of that word, the more deviant an alteration is to the intended word, the more disruption there will be to lexical access of that word (Connine, Blasko, & Titone, 1993; Connine, Titone, Deelman, & Blasko, 1997; Marslen-Wilson, Moss, & van Halen, 1996). Segmental alterations to monosyllabic words (Gow, 2001) are also more disruptive than to polysyllabic words (Connine et al., 1993), as a larger portion of the word remains intact in polysyllabic words. Alterations that result in another word (Gow, 2001) are also more disruptive than those that result in a non-word (Milberg, Blumstein, & Dworetzky, 1988).

This sensitivity to segmental information contrasts with the tolerance allotted to variations in suprasegmental information. Recognition of words persists across various suprasegmental contexts. From experience we know that adults can recognize words spoken in different affects and discourse properties, which alter the suprasegmental properties of a word. Further, as will be discussed in Section 4.1.6 Spoken word recognition is robust, adults can also segment accented speech, which in the case of foreign or second-language accented speech, can be produced with the suprasegmental properties of the speaker’s native language (Gut, 2003), rendering suprasegmental information relatively uninformative or even misleading to a native listener of the target language. That word recognition is more sensitive
to variation in segmental information as compared to suprasegmental information reflects its primary reliance on segmental information.

4.1.3 Words are considered and compete in parallel. At core, models of spoken word recognition in adults assume that incoming speech input is matched to existing lexical entries, with word recognition occurring once a suitable match is found (see Section 4.2 Models of Adult Spoken Word Recognition). Some early theories, such as the Autonomous Search Model (Forster, 1976), proposed that the incoming speech signal is assessed on its orthographic (if available), phonological, syntactic, and semantic properties, and a search is initiated for each domain to find a word that matches the assessments in each of those domains. Importantly, the search was purported to be serial, so that the compatibility of individual words for each feature set would be assessed sequentially, starting with high-frequency words, with the search terminating once a suitable match was found.

Another early model of spoken word recognition, the Logogen Model (Morton, 1969), proposed a parallel search. This model proposed that each lexical entry has a logogen that is used to determine the word's similarity to the speech input on the basis of several different features. Each logogen has an individual activation threshold based on the amount of shared similarities between the logogen's lexical entry and the speech input. Once the threshold for a logogen is reached, it is activated and its corresponding lexical entry is selected. Notably, this model allows for all lexical entries to be assessed on the basis of all available features in parallel, allowing for an enormous increase in speed, relative to the Autonomous Search Model.
Influenced by the Logogen Model, more recent models of adult spoken word recognition tend to propose a parallel rather than serial lexical search (e.g., McClelland & Elman, 1986; Norris, 1994; Norris & McQueen, 2008). The parallel nature of lexical search is supported by evidence showing the activation of multiple words at once (e.g., Connine, Blasko, & Wang, 1994; McQueen et al., 1994; Vroomen & de Gelder, 1995; Zwitserlood & Schriefers, 1995). For example, in a word identification task, participants were asked to indicate when they heard a real word in a list of real and non-words. When participants received previous exposure to the sequence CAPT, or /kӕpt/ – the onset of the words CAPTAIN and CAPTURE – they were faster to judge the semantically-related words SHIP (related to CAPTAIN) and GUARD (related to CAPTURE) as real words, indicating that CAPT activated both CAPTAIN and CAPTURE (Marslen-Wilson, 1987). Semantic priming also reveals that words embedded within larger words (Isel & Bacri, 1999; Luce & Cluff, 1998; Shillcock, 1990; Vroomen & de Gelder, 1995) or across word boundaries (Gow & Gordon, 1995; Tabossi, Burani, & Scott, 1995) are activated during spoken word recognition. In Dutch, the word BRANCARD (stretcher) contains the embedded word CAR (car). In an experiment, BRANCARD primed both the Dutch word for “hospital”, which is semantically related to the word for “stretcher”, and that for “bus”, which is semantically related to “car” (Isel & Bacri, 1999).

This parallel activation of words reflects a constant competition between the possible lexical candidates, with possible words constantly deemed more or less probable as the speech input progresses. In any type of competition, the more competitors there are, the longer it takes to find a champion. This too is the case for
competition among lexical competitors. Selection of the correct (or most probable) word from a lexical cohort takes longer when there are many similar-sounding words (Cluff & Luce, 1990; Luce & Large, 2001; Luce & Pisoni, 1998; Vitevitch, 2002; Vitevitch & Luce, 1998, 1999), as well as when a given word contains many embedded competitor words (Norris, McQueen, & Cutler, 1995; Vroomen & de Gelder, 1995).

**4.1.4 Suprasegmental information constrains the lexical cohort.** As discussed earlier in Section 3.4 Use of Suprasegmental Features as Cues to Word Boundaries, suprasegmental properties can signal information to the perceiver either on their own, or combined with other properties. While the segmental portion of the speech signal is the predominant source of speech information in spoken word recognition by adults, suprasegmental information nonetheless has a significant influence on spoken word recognition, overlapping with its role in developing word segmentation in children. This influence can be considered to occur in two ways.

**4.1.4.1 Suprasegmental information influences lexical access.**

Suprasegmental information can be tied to the lexical entry for the word. In English, a few minimal pairs exist in which two segmentally identical words are differentiated by stress, such as CONduct and conDUCT, and PERmit and perMIT. Pitch accent can also differentiate words in languages such as Japanese, Korean, Swedish, and Danish, and differences in pitch height and contour alter meaning at a lexical level in tonal languages such as Mandarin and Cantonese. When suprasegmental information is tied to the lexical entry for a word, this information can influence the consideration of a word as a possible parse in spoken word recognition. For example, in a lexical decision task, native Japanese perceivers’
recognition of spoken target words was primed when the priming word shared the same mora and pitch accent, but not when it shared only the same mora, indicating that the pitch accent as well as the segmental content of the mora was tied to the lexical entry of the target word (Cutler & Otake, 1999).

As mentioned previously, the influence suprasegmental information has on lexical selection is dependent on whether that information is tied to the lexical entry of a word. Suprasegmental features occur not only across segments, but also across word boundaries, for instance at the phrasal level. They can also signal the emotional affect of a word, or discourse properties (for example saying a word as a question or a statement), which may change the word's intention, but usually not its meaning. Suprasegmental information can also exist at the lexical level without being tied to the lexical entry of a word. For example, in French words carry fixed stress – that is, the stress is always in a certain location within the word, in this case on the final syllable. Because the stress is predictable and constant across all words, French listeners are not sensitive to it in terms of the lexical identity of the word (Dupoux, Pallier, Sebastian, & Meher, 1997).

### 4.1.4.2 Metrical patterning informs possible word boundaries

As seen in word segmentation by children (Section 3.4.1 *Use of metrical patterns as cues for word boundaries*), the other way in which suprasegmental information affects spoken word recognition is by narrowing the location of possible word boundaries down to those that fit within metrical units. These cues are specific to one's native language which can be seen in native listeners’ sensitivity to the relevant metrical pattern in their native language, and in their persistence in listening for this pattern even when perceiving a non-native language (for more, see Section 3.4.1.2. While
such cues are informative for cueing possible word boundaries, they can be considered supplementary to segmental cues for adults, as the metrical patterning alone provides little information for word identity. Rather, it is likely that in mature spoken word recognition the metrical patterning of the language may serve to disambiguate when multiple segmentations are possible. However, as will be discussed in Section 4.1.6.3 Spoken word recognition across foreign accents, misalignment of metrical patterning does not necessarily lead to discarding a possible word, but instead likely disadvantages it. Adults are still able to recognize words spoken with uninformative or conflicting meter, as can be the case in the perception of foreign accents, though recognition can be (at least initially) slower compared to native perception (e.g., Clarke & Garrett, 2004; Munro & Derwing, 1995a, 1995b; Rogers, Dalby, & Nishi, 2004; Schmid & Yeni-Komshian, 1999; van Wijngaarden, 2001; see also Gut, 2003).

4.1.5 Contextual information influences lexical selection. Sentential context also influences perceivers' selection of a word by biasing selection in favor or against a word, depending on possible candidates (Blutner & Sommer, 1988; Conrad, 1974; Lackner & Garrett, 1972; Lucas, 1987; Oden & Spira, 1983; Onifer & Swinney, 1981; Seidenberg, Tanenhaus, Leiman, & Bienkowski, 1982; Tanenhaus & Donnenwerth-Nolan, 1984; Tanenhaus, Leiman, & Seidenberg, 1979; Whitney, McKay, Kellas, & Emerson, 1985). In one study, perceivers were presented with words carrying an ambiguous meaning (e.g., BUG refers to an insect or a spying device) in sentences that pointed to one or the other meaning. While both meanings were initially activated as possible candidates (demonstrated by a priming of words semantically related to both interpretations in a subsequent lexical
decision task), after a short time only the contextually matching word remained activated (Swinney, 1979). Further, in a study in which native speakers were presented with speech in noise and asked to identify the final word in contextually informative (e.g., “The boat sailed across the bay.”) versus contextually uninformative sentences (e.g., “John was thinking about the bay.”), perceivers were able to tolerate higher noise levels for the contextually informative sentences than the uninformative ones (Mayo, Florentine, & Buus, 1997).

Aside from perceivers using top-down context to influence spoken word recognition, perceivers are sensitive to bottom-up contextual information as well. In two separate phoneme identification in noise tasks, native perceivers were better than non-native perceivers at identifying consonants that were preceded by low-level contextual cues of the preceding vocalic context and a constant-duration noise (Garcia Lecumberri & Cooke, 2006). This enhanced native ability was not found when the preceding vocalic cues were not available (Cutler, Weber, Smits, & Cooper, 2004; see Cutler, Garcia Lecumberri, & Cooke, 2008).

4.1.6 Spoken word recognition is robust.

4.1.6.1 Spoken word recognition in noise. From experience, we know that adults are able to recognize speech across conditions brought forth by the environment that impoverish the speech signal. We regularly are able to understand speech from multiple talkers, in windy environments, over a telephone, and across many combinations of those factors (albeit with varying difficulty). Recognizing speech in noise, for example at a noisy bar or party, is a common environmental factor regularly overcome by native speakers with considerable efficacy. Adults are able to identify segments amidst noisy conditions, with a mean correct identification
rate higher than 50% when segments were presented along with either competing
talkers speaking their own or a non-native language, speech-shaped noise (white
noise adjusted to match the amplitude envelope of speech), or eight-person babble,
all presented at a 0 dB signal-to-noise ratio (Garcia Lecumberri & Cooke, 2006).

The ability to identify individual segments in noise undoubtedly aids spoken
word recognition, reducing the number of spurious candidate words cued by
misperceived segments, and increasing the chance that the correct words are
included in the cohort of possible words. The next step is recognizing words in
noise, which adults are also able to do with considerable efficacy, being able to
identify words at the end of context-free sentences masked in babble played at over
70 dB (at that point reaching 50% accuracy; Mayo et al., 1997).

4.1.6.2 Spoken word recognition across non-native regional accents. Aside
from understanding speech in an impoverished environment, adults can also
understand speech when the speech signal contains variation from their native
pronunciations of words and sentences. Other regional accents of a language can
contain segmental (e.g., Adank, van Hout, & Van de Velde, 2007; Clopper, Pisoni,
& de Jong, 2005; Wells, 1982) as well as suprasegmental variation (e.g., Grabe,
2004; Wells, 1982) from the listener’s native accent. As regional accents are
variations upon the same native language, and are produced by native speakers of
the language, the variation across such accents is systematic and consistent with the
phonological structure of that language.1 Although adults are able to comprehend
language across regional accents, research has shown that the variation does slow
the processing of continuous speech. For example, participants were 30 ms faster to

---

1 Regional accents are defined here as variations of a given language that differ from one another
in the pronunciation of the same words. This contrasts with the term regional dialects, which also
subsumes substantial variations in vocabulary and grammar.
decide whether the final word in a sentence was a word or a nonword when the sentence was spoken in their native regional accent versus when it was spoken in a non-native regional accent (Floccia, Goslin, Girard, & Konopczynski, 2006). Similarly, speakers of Southern Standard British English, a regional accent in the Greater London area, were faster and more accurate at answering true/false questions when the questions were spoken in their native regional accent, versus a non-native regional accent, Glaswegian English, spoken in Glasgow, Scotland (Adank, Evans, Stuart-Smith, & Scott, 2009). However, the same study suggests that familiarity with a regional accent affects its processing time, as native speakers of Glaswegian English did not show any difference in processing time between sentences in their native regional accent and Southern Standard British English. Southern Standard British English is widely used in broadcasting in the United Kingdom, thus the Glaswegian English speakers were familiar with Southern Standard British English, but the reverse was not true (Adank et al., 2009).

Regional accents perturb processing of individual words as well. In an animacy judgment task, where participants had to judge whether a spoken item referred to a living or non-living thing, participants were faster for words presented in their native regional accent versus a non-native regional accent (Adank & McQueen, 2007). Sentences spoken in a non-native regional accent contain both segmental and suprasegmental variation from the listener's native accent, thus it is impossible based on this study to say whether one or the other, or both, are primarily responsible for the processing delay. However, the fact that regional accents perturb processing of individual words presented in isolation demonstrates that segmental variation is at least partly responsible (Adank & McQueen, 2007).
4.1.6.3 *Spoken word recognition across foreign accents.* Foreign accents, which occur in second language speech production, can also contain considerable variation from native pronunciations, and as with regional accents, a foreign accent slows perception of spoken words (e.g., Clarke & Garrett, 2004; Munro & Derwing, 1995a, 1995b; Rogers et al., 2004; Schmid & Yeni-Komshian, 1999; Snijders, Kooijman, Cutler, & Hagoort, 2007; van Wijngaarden, 2001). For example, in a cross-modal matching task in which participants had to decide whether a visually presented word matched the last spoken word in a sentence produced in the participants' native accent or in a foreign accent, Clarke & Garret (2004) found that participants' reaction times were 100-150 ms slower for foreign-accented speech than for speech in their native accent.

Results from studies that have compared perception in one’s native accent across perception in another regional accent or in a foreign accent suggest that the variation in foreign accents incurs a larger processing delay than that found in other regional accents. For example, in the Floccia et al. (2006) study described in the previous section, participants were 30 ms slower at deciding whether the final word in a sentence was a real word or nonword when the sentence was spoken in another regional accent compared to their native regional accent. But when spoken in a foreign accent, participants were 100-150 ms slower at the task. Likewise, native speakers of Southern Standard British English were faster and more correct at answering true/false questions in their native regional accent compared to a non-native regional accent (Glaswegian English), but were even slower when the questions were spoken by a native speaker of Spanish, in Spanish-accented English (Adank et al., 2009).
Clarke & Garrett (2004) proposed that the processing time required to comprehend an accent may depend on how acoustically distant the accent is from one's native accent. In this way, the different processing time between non-native regional accents and foreign accents may result from the tendency of regional accents to be less varied from one another compared with the variation found across a native accent and a foreign-accented speech. However, the difference may instead, or in conjunction, lie with the different nature of variation across other regional accents and foreign accent. Regional accents vary systematically from one another, as two regional accents are variations of a single language. In contrast, foreign accents do not contain systematic variation according to the target language. This is because the variation produced by a foreign or second language speaker will be systematic according to the speaker's own native-language phonology, and thus speech produced in a foreign accent may be altered in a way not predictable to the native perceiver of the target language. Additionally, while regional accents can vary from one another in suprasegmental features, this type of variation is generally greater across foreign accents. This is because regional accents will still generally share many suprasegmental features, particularly metrical features, as described in Section 3.4.1.1 *Metrical classes of languages*. In contrast, foreign accents can be pronounced according the to suprasegmental features of the speaker's native language (Gut, 2003), which can render the cues uninformative, or even misleading, to a native listener of the target language. Together, and along with variations in proficiency that can occur when speaking a non-native language, these differences in variations may account for the increased processing delays associated with perceiving foreign-accented speech compared to regionally accented speech.
**4.1.7 Relationship among segmental, suprasegmental, and contextual cues.**

This chapter has discussed how adults use segmental, suprasegmental, and contextual cues to inform which words are in the continuous speech signal. However, what remains to be discussed is how these cues are weighed by the perceiver. In a series of studies by Mattys et al. (2005) which pitted suprasegmental, segmental, and contextual/lexical cues together, the authors found that while all cues inform spoken word processing, the cues are weighted in a hierarchy, in which segmental cues preside over suprasegmental cues, and contextual cues preside over segmental cues. Thus spoken word recognition by adults appears to be dominated by top-down, knowledge-driven cues (when other cues are conflicting), with signal-based cues presiding when context is unavailable or degraded.

This helps clarify results where speech processing was delayed in accented speech compared to native pronunciations, as described in Sections 4.1.6.2 *Spoken word recognition across non-native regional accents* and 4.1.6.3 *Spoken word recognition across foreign accents*. The stimuli in those experiments were devoid of contextual information, and some even were devoid of lexical information, as one experiment explicitly allowed the possibility of non-words (Floccia et al., 2006). Thus participants had to rely on the segmental and suprasegmental cues alone, which were distorted according to native pronunciations, and thus led to delays in processing time. It is worth noting though that in real life situations, speech generally contains redundant contextual information. Thus the processing time associated with recognition of speech across other regional accents and foreign accents in real situations could perhaps be closer to native processing levels.
4.2 Models of Adult Spoken Word Recognition

Several attempts have been made to construct a model of word segmentation that can account for many of the behavioral results that have been presented here. The more prominent theories are detailed in the following section.

4.2.1 The Cohort model. Cohort is an early spoken word recognition model designed with the specific intention of modeling auditory speech processing, focusing on the temporal quality of the speech signal. It was based on shadowing research in which participants were asked to repeat the words they were presented as quickly as possible. It was found that participants often began repeating the word before it had ended, indicating that the word had been activated and recognized before the offset of the word (Marslen-Wilson, 1975; Marslen-Wilson & Welsh, 1978). From this finding, Marslen-Wilson and Welsh (1978) proposed their model in which as soon as the initial segments of a word were perceived in the incoming speech input, a list of all possible word candidates beginning with those segments were activated in a word-candidate “cohort”. As the speech signal unfolded, candidates that no longer matched the input were deactivated (or penalized, in later versions; Marslen-Wilson, 1987), with word recognition occurring when only one word remained in the cohort (at the word's uniqueness point).

While this model was original in its idea that spoken word recognition occurs online, in realtime, it did not reflect realistic spoken word recognition. Common criticisms (see McClelland & Elman, 1986) include that Cohort greatly overvalues the beginnings of words, as the list of candidate words is based on the words that share the initial consonant cluster and vowel. From there, as the speech signal unfolds, candidate words are eliminated once there is a segment mismatch.
between the candidate word and the input. Thus, successful word recognition requires the listener to have perceived the beginning of the word in order for a cohort to be built that accurately reflects the speech input. Additionally, it requires the speech signal contain only clear, intelligible phonemic information that matches the perceiver's phonemic inventory. That is, it is explicitly unable to handle a degraded speech signal (e.g., speech in noise), mispronunciations, or pronunciations from regional or foreign accents, as mismatched words are immediately eliminated from the Cohort without a method for recovery. As such, it cannot describe speech recognition as it occurs in the real, often noisy world.

4.2.2 The Trace model. Trace (McClelland & Elman, 1986) is a classic interactive-activation connectionist model in which the incoming phonetic information from the speech signal activates a level of processing units (nodes), which in turn influence nodes at the phonemic level, in turn influencing nodes at the lexical level. In this way, spoken word recognition is modeled on three hierarchical tiers, with inhibitory connections within a tier, and excitatory connections across tiers (i.e., activation of a phoneme will inhibit other phonemes, but will activate words containing that phoneme).

This approach means that Trace is more tolerant of variation in speech compared to Cohort. Words are not eliminated as a candidate words if there is a signal-to-word mismatch, but rather are just more weakly activated, and it is possible that the word will remain the most highly activated candidate despite variations to the speech signal. As well, Trace allows interaction and feedback across lexical analysis and phonemic processing, resulting in a system more forgiving of mismatches, for example, if forgiveness would result in a word where
otherwise there would be a non-word. The proposal that words are not eliminated as soon as a mismatch occurs is supported by eye tracking evidence in which participants were asked to point to a picture corresponding to an auditory target (e.g., BEAKER). While two words sharing an onset (BEETLE) both showed activation (inferred by looks to that image), rhymes of the target (SPEAKER) were likewise activated at some point after the onset (Alloppena et al., 1998).

While Trace proved popular due to its ability to simulate many aspects of spoken word recognition, and its explicit computational specification (as compared to the computationally vague Cohort model), it was still not without its critics (see Norris, 1994). Trace's architecture requires that it be set up at the beginning of each word – that is, it is designed to recognize only a single word left to right. In running speech, where a new word begins is often ambiguous, and getting around this requires many copies of Trace to occur at any given time, with the perceiver ignoring input from most of the instantiations. This is a very inefficient architecture, requiring a massive memory load, and it is unlikely that such massive inefficiency occurs in the habitually efficient architecture of the brain. Further, the architecture of Trace poses a problem for embedded words. If an embedded word is recognized – for example, if CAT and LOG are recognized in “catalogue”, there is no way to turn back and revise the decision to CATALOGUE. It is possible that such an architecture could be trained to hold off on a decision when it has perceived certain words that have previously been found embedded in others to determine whether or not they are embedded words. However, such exceptions would also have to be included for variants/mispronunciations of the embedded words (e.g., delay recognition if CAT is perceived to see whether it is CATALOGUE, but also if
CAD is perceived in case the word is CATALOGUE, but has been mispronounced as “cadalogue”), which would further add to the already improbable memory load.

Trace also supposes a continuous two-way flow of information between the phonemic and lexical tiers. In this way, lexical knowledge influences phonemic perception, which allows for handling of mispronunciations, degraded stimuli, or pronunciation differences. For example, if the word BLACELET is perceived phonemically, because BLACELET is not a word, but BRACELET is, that lexical knowledge can feed down and influence the comprehension of the second phoneme as /r/ instead of /l/, so that the word can be perceived despite the mispronunciation (McClelland & Elman, 1986). However, this feedback from the lexical to the phonemic level is redundant. If the information to accept a word despite a mispronunciation is available at the lexical level, then the decision can be made on that level without any need to feed that information down to the lower phonemic level (Norris, 1994). Finally, due to the nature of the top-down feedback, Trace also predicts that top-down effects too strongly influence perception of the final phonemes, as top-down lexical effects depend on activation, and phonemes perceived earlier (i.e. beginnings of words) should have activated words that alter perception of later phonemes. However, this is in direct contrast to behavioral data (McQueen, 1991).

4.2.3 The Shortlist A (Activation) model. Unlike Trace, Shortlist A (Norris, 1994; referred to as Shortlist prior to the introduction of Shortlist B; Norris & McQueen, 2008) models phonemic perception as being entirely bottom-up, rather than a combination of bottom-up and top-down influences. In this model, as the speech input comes in, there is an exhaustive search of the lexicon to create a set of
at most 30 (although this threshold is arbitrary) candidate words that have the highest bottom-up activation match to the input (a “short-list”). Candidate words then compete amongst each other via inhibition of overlapping candidates, with the most highly activated word ultimately recognized.

Thus this model assumes no top-down feedback from the lexical to phonemic level. Lexical candidates are selected via correspondence to bottom-up recognition of phonemes, and words compete amongst each other at the lexical level. Compared to Trace, this method is less computationally demanding, as it omits top-down feedback. As well, this model does not require multiple complete instantiations that duplicate the entire lexicon many times over. Instead, only a handful of words are activated at each phoneme, with the number of active short-lists at any time limited by the length of the longest word(s) in the lexicon. This greatly reduces the memory load required, and makes word recognition possible in face of the large vocabularies characteristic of adult perceivers. While Shortlist A was much more successful at describing phenomena that Trace could not predict, it also has its weaknesses, which are outlined in the following subsection.

4.2.4 The Shortlist B (Bayesian) model. While interactive-activation models have been very influential in the field of spoken word recognition, having considerable success in modeling large-scale and realistic speech input, there are still several issues that they have been unable to address in full. Firstly, these models rely on “activation” of possible words, with the most strongly activated word at the end of some period being segmented from the speech stream. However, the physiological and/or behavioral correlates of this “activation” remained unspecified (Norris & McQueen, 2008). Secondly, interactive-activation networks
invoke many different free parameters, all of which are assigned different and
generally arbitrary weights, and even small variations in the weightings can greatly
affect the efficacy and/or output of such models in simulations (Norris & McQueen,
2008). This is theoretically unsatisfying and does not match the robustness of adult
speech recognition.

While it maintains many assumptions from its predecessor, Shortlist A (see
Norris & McQueen, 2008), Shortlist B differs radically from corresponding
interactive-activation models in its replacement of these networks with a framework
based instead on Bayesian calculations. This greatly simplifies the computational
model compared to previous models, and replaces the vague “activation” output
with a direct output of posterior probabilities.

In Shortlist B, rather than a string of phonemes, the input is a string of
phoneme likelihoods based on phoneme confusability in the perceiver’s native
language. This was tested in Shortlist B using an exhaustive data set on diphone
confusion patterns from Dutch (Smits, Warner, McQueen, & Cutler, 2003). From
this input, word probabilities are generated, and these probabilities are recomputed
based on possible valid segmentations of the phoneme string. This approach is
powerful in its simplicity, and in its ability to overcome mispronunciations and
limited information. Further, it is able to account for effects of word frequency,
lexical status, and surrounding context (Norris & McQueen, 2008), and is presently
considered an especially viable model for adult spoken word recognition.

In summary, it appears that mature word segmentation and recognition is
best described as a process that is made robust in the face of degradation, variation,
and ambiguity in the speech signal by combining many different sources of
information, including knowledge about phoneme likelihoods, word frequencies, and various paralinguistic cues. While this process is powerfully efficient in adults, children must use other means in the absence of the enormous lexicon required by the mature processes. How children's word segmentation differs from adults is discussed in the following chapter.
5 Differences between spoken word recognition in children and adults

As demonstrated by the previous two chapters, adult word segmentation is not only more reliable than that by infants and toddlers, but is also qualitatively different. What is most apparent is that for adults, word segmentation and word recognition appear to be deeply intertwined, and may even be a single process (Reddy, 1976). On the other hand, for infants and young toddlers, they are clearly two separate processes, characterized by two separate lines of empirical support. The following discussion further examines these characteristic developmental differences and offers several possible explanations for this qualitative disparity.

5.1 Differences in the Lexicons of Young Children versus Adults

Adults have vocabularies that are many times larger than those of young children who are only beginning to build their lexicons. This section discusses how
this disparity in lexicon size result can account for some of the differences in spoken word recognition seen across young children and adults.

5.1.1 Effects on cohort size. As seen in the preceding chapter (Section 4.2 Models of Adult Spoken Word Recognition), many models of adult spoken word recognition propose that spoken word recognition is primarily achieved through competition of lexical candidates determined by the incoming speech signal. That is, as segments are perceived as the signal unfolds over time, various candidate segmentations are considered, with the list of possible words changing as additional components of the word (and surrounding words) are detected. This manner of segmentation necessitates a large vocabulary (numbering in the tens of thousands), encompassing most words in common use. Infants and toddlers are at an obvious disadvantage here, given their much smaller lexicons. Thus, it may be that the primary cause of the disparity between the word segmentation/recognition processes of infants and adults is adults' vast vocabulary in comparison to infant and toddlers’ smaller, still-developing lexicons. Further, while infants and toddlers lack the necessary lexicon in order to effectively segment words in the manner found with adults, they must first be able to segment words from connected IDS sentences in order to build that lexicon. Thus, infants and toddlers have no choice but to rely on bottom-up cues, be they segmental, suprasegmental, or statistical cues available in the speech stream (as described in Chapter 3 Development of Word Segmentation), that adults are able to treat as secondary and/or as a fallback.

5.1.2 Effect on use of context. The impact of the lack of a substantive lexicon is twofold. In addition to limiting the number of lexical competitors (and decreasing the likelihood that the target word is in one's lexicon), a small lexicon
also limits one's ability to use context to help inform words and word boundaries. Knowledge of surrounding words can aid segmentation and recognition of adjacent words (Bortfeld et al., 2005; Shi, Cutler, & Werker, 2006), and semantic knowledge of context can also influence word recognition (e.g., Marslen-Wilson, 1985; Marslen-Wilson & Welsh, 1978). Without this knowledge of semantic and lexical context, infants and toddlers must increase their reliance on segmental, suprasegmental, and statistical cues.

5.2 Robustness

Adults are able to segment and recognize speech across a variety of variations and impoverishments. As reviewed in Section 4.1.6 Spoken word recognition is robust, adults can segment/recognize speech in noise (Garcia Lecumberri & Cooke, 2006; Mayo et al., 1997), across segmental variation, such as that found across regional accents (e.g., Adank & McQueen, 2007), and across suprasegmental and segmental variation, such as when listening to foreign accents (Adank et al., 2009; Floccia et al., 2006). While segmenting and recognizing varied or degraded speech is not without its processing costs (e.g., Snijders et al., 2007), adults still regularly accomplish this at a competent level. There is evidence, meanwhile, that infants and young toddlers are more severely hindered in their ability to understand segmentally and/or suprasegmentally varied speech. At 9 months, children are unable to recognize familiarized words embedded in a sentence when the sentence is in a different regional accent (Schmale, Cristia, Seidl, & E. K. Johnson, 2010) or in a foreign accent (Schmale & Seidl, 2009) compared to familiarized stimuli, despite younger children's ability to do so when both
familiarization and test stimuli are in their native regional accent (Jusczyk & Aslin, 1995). It is possible that sentential context may play a role here. If children are unable to recognize the surrounding words in a sentence, this may impact their ability to segment the familiarized word. However, even older children, at 15 months, cannot recognize words in isolation when they are pronounced in another, non-native regional accent. Fifteen-month-olds are able to discriminate and show a listening preference for words having a high frequency in toddler vocabularies from words having low frequencies in adult vocabularies when the words are spoken in their native regional accent, but are unable to do so when words are spoken in another regional accent (Best et al., 2009). These results, paired with the fact that laboratory settings can control for a reduced lexicon in a young child compared to an adult (and have done so) by testing segmentation of only known words, indicates that there may be more than a difference in lexicon size behind infants' and toddlers' inability to segment words as adults do.

5.3 Phonological Knowledge of Word Structures is Required

Instead, a key difference between word segmentation in infants and young toddlers versus older children and adults involves the level of knowledge of the phonological structure of a given word. For adult-like segmentation to occur, word forms must be strict enough to be a reliable indicator of the word, but also flexible enough to account for speaker variation, mispronunciations, and regional- and foreign-accented speech. Such word forms not only allow for robust spoken word recognition across such variation, but are also less cognitively taxing compared to more detailed phonetic word forms. This reduction in cognitive load theoretically
leads to more robust spoken word recognition as it frees cognitive resources which allow for recognition in more demanding situations (e.g., in situ versus in a controlled laboratory setting). The assertion that phonologically specified word forms are less cognitively demanding than phonetically specified word forms is based on the success of computational models such as Shortlist B (Norris & McQueen, 2008), which is currently the most widely accepted model of spoken word recognition due to its ability to most accurately model spoken word recognition in adults (see 4.2 Models of Adult Spoken Word Recognition). This model employs a lexical search based on the incoming segmental stream, and is made efficient by the use of phonemic or phonological word forms, separating processing of phonetic details from later probabilistic word selection. While there is an early computational cost to categorizing phones, performing a lexical search based on the high-order speech information, and not the more detailed phonetic information, has a later payoff in word recognition efficiency. Importantly, in Shortlist B, there is allowance for variation in the phonetic-phonological mapping. This is accomplished by the incorporation of confusion probabilities in Shortlist B, with multiple phonemes activated to different levels according to their language-, if not accent-, specific confusion likelihoods. While Shortlist B makes no specific allowance for speaker or accent variation, its Bayesian architecture readily admits incorporation of these factors into the probability determination. Thus in both cases an allowance for variability both within and beyond phonemic boundaries that is not simply post-phonemic processing error recovery, but instead is built early into phonological processing, results in a more efficient and more successful computational model. Regardless of whether this model, or a descendent, is
ultimately able to fully describe naturalistic speech recognition, it is reasonable to say that the use of phonological word forms has been more efficient and less computationally demanding. In a similar fashion, it is likely reasonable to say that the same applies to human speech recognition.

Vocabulary measures enable us to determine when children develop a larger lexicon, which, as already discussed, may be a key factor influencing when children are able to recognize words in continuous speech based on the segmental word structure. What is less obvious, however, is how children develop more mature phonological representations of words, which, in addition to a substantive vocabulary, appears to be a crucial step in acquiring the robust word segmentation techniques of adults.
6 Development of Word Specificity

If adults' ability to recognize words in continuous speech in a robust manner lies in their having fully specified phonological knowledge of word structures, then the transition to mature word segmentation should coincide with the development of such adult-like word representations. This section discusses the development of word representations with the aim of drawing insights into the developmental progression toward adult-like spoken word recognition.

6.1 Requirements of Word Recognition

Fine variation in production of phonemes means that no two speakers, or even the same speaker, will produce identical acoustic signals for different utterances, or tokens, of a given word. Across regional or foreign accents, inter-speaker variations can be even more extreme. Thus, for reliable segmentation and recognition of words in speech to occur across speaker variation, the perceiver must
be able to ascertain the word intended by the speaker, despite such phonetic variation.

6.1.1 Levels of description of word structure. Acceptance of variable pronunciations requires knowledge of the abstract phonological forms of words, or the common underlying structure that is shared across native pronunciations of the word. Such phonologically specified word forms might seem to be captured well enough by listing out the sequence of phonemes in a word. However, they are more accurately described as an abstraction at a higher level than that of phonemes. Specifically, the phonological word form is an abstraction at the word level, which in addition to specifying the sequence of phonemes for that word, also incorporates additional specifications for how those phonemes may be realized in that word, governed by factors such as the surrounding phonemes and the stress patterns of the word. Importantly, phonological word forms, as defined here, are able to accommodate systematic phonetic changes that would otherwise violate the phonemic boundaries of the listeners native accent, such as in a non-native regional accent. Here, abstract phonological form is denoted via broad transcription of the phonemes in a word. While this denotation does not differentiate phonological word forms from a phonemic level of specification, and inadequately specifies the suprasegmental structure and phonological processes affecting a spoken word form, no alternative current transcription approach exists for describing phonological word forms. The phonological forms of words contrast with the simpler, stricter, phonetic forms of words, where words are known by the pattern of the phonetic features according to a native pronunciation, and are here denoted via a narrow transcription of the phonetic makeup of a word’s precise pronunciation. As well,
some models of developmental word recognition and research in this area have referred to or implied a level of word form specification more abstract than phonetic word forms. In these cases these proposed word forms derive from children’s increased native phonetic knowledge – that is, their developed phonemic categories, and in this way can be considered to be phonemically specified word forms, which are of a lower order than the phonological word forms we refer to here. We believe that a phonemic level of specification is inadequate for the type of robustness we describe here. For example, as will be revealed in Experiment 4, 19-month-olds are able to recognize familiarized words across regional accent variation that maps one or more of the accented phones to a different phonemic category in the native regional accent – a feat that would not be possible were word forms only phonemically specified.

6.1.2 Types of phonetic variation. When dealing with phonetic variation across pronunciations of a given word, there are two possible outcomes of a particular phonetic variant. First, two different utterances can contain phonetic variation that does not alter the phonological structure of a word, referred to as phonological constancy (see Best et al., 2009). Phonological constancy is reflected in the ability of an Australian English (AusE) perceiver of the American English (AmE) NICE ([naːs]) to recognize the underlying phonological word as NICE, (or /nais/) despite the vowel change from the AusE pronunciation ([naːs]²). In contrast, two different utterances can contain phonetic variation that does alter the phonological structure of a word. This is referred to as phonological distinctiveness (Best et al., 2009), and is reflected in the ability of an AusE perceiver to recognize

---

2 Australian English vowel transcriptions are based on the inventory described in (Cox & Palethorpe, 2007).
that the vowel change from [noːs] to [niːs] in his or her native regional accent (or in another regional accent) delineates two separate phonological words, NICE (/naɪs/) and NIECE (/nis/), respectively. Note that the phonetic variations do not necessarily conform to permissible variations to native phonemes in the native regional accent, which means that in order to recognize a word correctly, perceivers must determine its underlying phonological structure, at times in situations where the shift in production may signal a different phoneme in their native accent. This is discussed further in Chapter 7 Research Questions.

Robust and reliable word recognition requires that perceivers have a grasp of both phonological constancy and phonological distinctiveness, as word recognition must be flexible enough to accept speaker variation (from individual differences in vocal tract characteristics, through to differences across regional accents of the native language), but exacting enough to accept only those variations that preserve the identity of the word. These constraints in turn require that they possess knowledge of the phonological forms of words.

6.2 Accounts of the Development of Phonologically Specified Words

While the ability of adults to successfully recognize words across regional accents implies that adults possess knowledge of the abstract phonological structure of words, it does not appear that knowledge of word structures begins in this way. Just how (or indeed, whether) this knowledge develops has been debated by several theoretical viewpoints.

6.2.1 PRIMIR. In Werker and Curtin's (2005) developmental framework for Processing Rich Information from Multidimensional Interactive Representations
(PRIMIR), learning occurs via the general mechanism of statistical learning, modulated by certain attentional filters, including an innate preference for speech (especially IDS), and other linguistically related features (point vowels, proper syllable form, and ability to process metrical patterns). Developmental age and language task (e.g., segmenting speech, learning words, etc.) can also act as attentional filters. Information gathered from the speech signal, such as the phonetic and phonological structure of words, is organized in such a way that different information or combinations of information can be tapped depending on the task at hand, with certain information becoming more or less salient at different developmental ages. According to how the information is related or occurs together, it is grouped onto three planes, reflecting General Perceptual, Word Form, and Phonemic information. Importantly, access to information is not constrained hierarchically, meaning that access to all information is always available.

According to PRIMIR, phonetic and indexical (e.g., gender, affect) information is stored on the General Perceptual plane, and the information from this plane helps form the Word Form plane, which comprises tracked exemplars of informational sequences, which are then attached to concepts, to form words in the lexicon. Once a sufficient number and neighborhood density of words have been learned, the Phoneme plane emerges from a generalization of the common features shared among the exemplar tracked word forms. In this way, early word forms are exclusively phonetically based, but later word forms are both phonetically and phonemically based, as access to both planes of information persists. While PRIMIR does not explicitly propose an age at which the Phonemic plane emerges, Werker and Curtin (2005) do point out a qualitative difference in word learning
between 14 and 18 months – a span which encompasses the vocabulary spurt (an increase in the rate at which children learn words, around the time that they have achieved 50 or so words that they can produce) and which may indicate emergence of phonemic knowledge at some point between those ages. This is consistent with PRIMIR's proposal that phonemic knowledge emerges given a sufficiently large vocabulary.

6.2.2 Perceptual attunement account. Best et al. (2009), proposed a perceptual attunement account, in which a perceptual shift in attention to phonological information at around the time of the vocabulary spurt (17-18 months) results in a shift in the specificity of word forms (see also Nazzi & Bertoncini, 2003). It should be noted that the use of the term vocabulary spurt can be misleading, as not all children undergo a sudden and dramatic increase in expressive vocabulary (Ganger & Brent, 2004). However, expressive vocabulary as used here is more of a reflection of the linguistic development of the child, and reflects that the relationship between the development of phonological specificity appears to be linguistically based, rather than necessarily chronological. As early word learners, children exploit their attunement to the phonetic patterns in their language environment as a means to recognize their first words, but as they mature they become attuned instead to the higher-order, more abstract phonological forms of words (see Best, 1994b, 1995). The perceptual attunement account is distinct in its proposal that children require exposure to systematic variation in the language in order to derive a word's underlying phonological structure, and that this necessary exposure is achieved by around the time of the vocabulary spurt, at which point attention shifts to the phonological forms of words rather than their specific
phonetic details. As well as bringing awareness of phonetic variation that
distinguishes words from one another, this attentional shift allows children to
understand pronunciations they have never before encountered, as they are able to
recognize the shared phonological form of words across phonetically varied
pronunciations (Best et al., 2009).

6.2.3 Statistical learning accounts. Statistical learning accounts posit that
word structures are made up of and thus perceived as strict phonetic patterns that
are constructed through experience, either via statistical learning or exemplar-
registration mechanisms that build words towards their canonical forms (e.g.,
Saffran, Aslin, et al., 1996). This view is supported by evidence that at 14-15
months, children are sensitive to phonetic changes in words. When a picture of a
baby was paired either with the spoken word BABY or VABY, children at this age
fixated more quickly to the correctly matched BABY (Swingley & Aslin, 2002).
However, when presented with a picture representing the newly learned word BIH,
same-aged children did not notice when the item was changed to DIH (Stager &
Werker, 1997). These results are posited to demonstrate that knowledge of phonetic
detail of words builds with experience, since the children had minimal experience
with the newly learned words and thus those word forms do not contain as much
fine phonetic detail as already known familiar words. Notably, at 19-20 months,
children are able to discriminate such slight but phonologically critical phonetic
differences even between newly learned words (Swingley, 2007; Werker, Fennell,
Corcoran, & Stager, 2002), presumed to be a result of better phonetic knowledge,
and increased working memory.

6.2.4 Phonological underspecification accounts.
6.2.4.1 Lexical density account. Alternatively, lexical density accounts propose that early words begin as global, underspecified forms, with phonetic specification of these forms deriving from linguistic (rather than cognitive) processes as the expanding lexicon demands increasing contrastive phonetic feature specifications in order to differentiate words (Brown & Matthews, 1997; Metsala & Walley, 1998). For example, if the first words in a child’s lexicon are MOM and CAT, the global representations of those words would likely not overlap, and indeed research shows that 14- to 15-month-olds can differentiate even among newly learned words that differ in many features from one another, such as LIFF from NEEM (Werker & Stager, 2000). But by this view, variant words will still be recognized, so long as they can be roughly matched to the global form of the word. If the child hears the word CAP, it will not be readily perceived as a different word from CAT. However, if it becomes apparent that those two phonetic forms refer to separate concepts, by necessity the child will eventually learn to differentiate those two words, which will in turn add to each word’s form the specification of either a labial place of articulation for /p/ or an alveolar place of articulation for /t/.

6.2.4.2 Distributional account. Similar to the lexical density account, the distributional account (Thiessen, 2007) posits that early word forms are global and underspecified, with specificity of words increasing as the lexicon grows. But while the lexical density account proposes that it is the acquisition of words that are neighbors of other words (i.e., the words differ by a single phoneme) that encourages phonemic knowledge to emerge, the distributional account proposes that the statistical distribution of phonemes within words in the lexicon is what drives phonemic awareness. Thus, phonemic contrasts are more likely to arise when the
phonemic pair is found in two words that are phonologically diverse rather than two minimal pair words that differ in only a single phoneme. This is because words that are phonologically similar are more likely to activate one another compared to words that are phonologically diverse, and experiencing phonemes uniquely is more likely to result in learning that phoneme. So if a child's lexicon consists of the words DUG and TUG, the child would be unlikely to discriminate /d/ from /t/, because perception of the word DUG would activate both DUG and TUG. Instead, a child would be more likely to acquire the contrast if his or her lexicon consisted of the words DUG, TOY, and TANK, as the word DUG is sufficiently different from TOY and TANK such that neither would be co-activated.

In this way, vocabulary is proposed to play a role in phonological development, with an increased vocabulary leading to increased phonological awareness. However, it is not simply raw vocabulary size that drives phonological knowledge, but the breadth of phonological distribution among the words in the lexicon. As a proxy, the larger the lexicon, the more likely there is a large breadth of phonological distribution, and in this way the distributional approach accounts for disparate findings between younger children's (14-16 months) inability to discriminate minimal pairs in a word learning task (e.g., Stager & Werker, 1997; Thiessen, 2007), where older children (17 months and older) succeed (Thiessen, 2007; Werker et al., 2002). In addition, this account is backed by research showing that 15- to 16-month-old children are more likely to discriminate a particular contrast when they have been exposed to words containing phonemes in dissimilar contexts from one another, compared to children exposed to words containing different phonemes in similar contexts (Thiessen, 2007).
6.3 Evidence for the Development of Word Form Specificity

Considerable research has already begun to untangle the developmental course of word form specificity. This research is reviewed below, and concludes with indication of a future direction of research.

6.3.1 Emergence of phonological distinctiveness. Classically, research on the development of phonetic and/or phonological word form specificity has relied on testing children's ability to discriminate minimal pairs, such as BABY from the non-word VABY (Swingley & Aslin, 2000, 2002). Such studies have found that by 18 months, children reliably detect such minimal pair changes between familiar words (e.g., Swingley, 2003; Swingley & Aslin, 2000). For example, in a visual fixation task, in which pairs of pictures were paired with either a correct pronunciation or a minimal pair mispronunciation of one of the pictures (as in the example of BABY versus VABY), 18- to 23-month-olds fixated on the named picture for a higher proportion of time when they heard the correct pronunciation versus mispronunciations (although proportion of looking time was greater to the named image compared to the unnamed image in both situations; Swingley & Aslin, 2000).

Discrimination at this age also extends to newly learned words (i.e., words learned through the course of an experiment). In the so-called switch task, 20-month-olds were habituated to two word-object pairs in which words were non-word minimal pairs (e.g., BIH and DIH). In each pairing, a novel, moving visual referent was paired with repetitions of one of the non-words which was the name for that item. Once children were habituated to each item (indicated by a sustained decrease in looking time), they were tested on “same” trials, in which the correct
word-object pair was played, as compared to “switch” trials, in which the object was paired with the incorrect minimal pair. During switch trials, but not during same trials, 20-month-olds' looking time increased to baseline levels, indicating that they had discriminated the newly learned minimal pair (Swingley & Aslin, 2000; Werker et al., 2002). The 20-month-olds’ discrimination of minimal pairs of even newly learned words, or rather, their grasp of when phonetic variation no longer fits with the phonological form of a word, indicates an understanding of phonological distinctiveness at that age.

Children younger than 18 months do not perform as reliably in such tasks. In the same switch task mentioned above, 14-month-olds failed to discriminate the newly learned minimal pairs where 20-month-olds succeeded (and 17-month-olds showed an intermediate ability; Werker et al., 2002; see also Stager & Werker, 1997). However, 14-month-olds may show some sensitivity to minimal pair mispronunciations of known words, indicated by a difference in proportion of looking time to the named image (versus an unnamed image) in a visual fixation task (Swingley & Aslin, 2002), in a manner similar to 19-month-olds (Swingley & Aslin, 2000). And while in a switch task, 14-month-olds were not sensitive to a switch to the other word in a newly learned minimal pair associated with a visual referent (Werker, Fennell, et al., 2002), they did show a sensitivity to a switch when the word-object pairs were known words (Fennell & Werker, 2003). As well, 14-month-olds can show a sensitivity to minimal pairs of newly learned words in a switch task given sufficient reduction of task demand, whether by presenting item names in sentential context, or by preceding habituation with a task training phase in which known words are presented in conjunction with their visual referents.
(Fennell & Waxman, 2010). Fourteen-month-olds also successfully learned minimal pair words when word-object training was followed by a less cognitively demanding visual fixation task (Yoshida, Fennell, Swingley, & Werker, 2009). Together, these studies indicate while phonological distinctiveness can be reliably observed by 18 months, it possibly begins to become established as early as 14 months. However, it is also possible that the reduced cognitive demands of the studies cited above did not uncover 14-month-olds’ emerging grasp of phonological distinctiveness, but instead resulted in tasks simple enough that they could be handled with phonetically specified word forms (see Best, Tyler, Kitamura, Notley, & Bundgaard-Nielsen, 2008; Best, Tyler, Kitamura, & Bundgaard-Nielsen, 2010).

6.3.2 Lack of phonological constancy in early findings. Up to this point, the studies reviewed in this chapter have only examined children's developing grasp of phonological distinctiveness, and not phonological constancy. This is a consequence of the type of stimuli used. While studies investigating children's grasp of phonological distinctiveness have necessarily tested their discrimination of minimal pairs, this approach has been unable to reflect sensitivity to phonologically constant aspects of word forms. Phonological constancy is reflected in phonetic variation that does not alter the phonological structure of a word. When altering, for example, /b/ to /v/ in BABY to form a minimal pair with VABY (as in Swingley & Aslin, 2000, 2002), this introduces a phonetic change that also results in a phonological change to the word’s structure or form.

To gain an overall understanding of the emergence of phonologically specified word forms, both phonological constancy and distinctiveness must be considered together, as each on their own can only inform part of the picture.
Phonologically specified word forms must be strict enough so that the perceiver can recognize when phonetic variation changes the phonological form of a word, but lenient enough so that they can simultaneously recognize when phonetic variation does not alter the phonological form of a word. Testing children's grasp of phonological distinctiveness can only inform whether or not word forms are phonologically underspecified—failure to discriminate minimal pairs demonstrates word forms that are insufficiently specified (though see Fennell & Werker, 2003 for an account of how task and cognitive demands can block access to lexical detail). Conversely, testing phonological constancy on its own can only inform as to the overspecificity of word forms. Failure to recognize that certain phonetic variations maintain the phonological structure of a word indicates specifications of word forms that are too stringent, leading the perceiver to false alarm perception of alternate pronunciations as novel words, non-words, or even other known words. Thus, the appearance that phonological specificity emerges by 18 months, and possibly as early as 14 months based only on the research reviewed above may in fact mask a phonetic overspecificity in children's early word forms. To address that question, children's grasp of phonological constancy must be examined.

6.3.3 Emergence of phonological constancy.

To examine children's awareness of language-specific linguistically relevant and linguistically irrelevant phonetic detail, 18-month-old English-learning and Dutch-learning children were tested on Dutch non-words (Dietrich, Swingley, & Werker, 2007). The nonsense words were minimal pairs by vowel length, which is phonologically contrastive in Dutch, but not phonologically contrastive in English. For instance, in Dutch [stat] (city) and [statt] (stands) are two separate words, but in
English such a distinction would be treated as two pronunciations of the same word, perhaps with the perception of one having a more “drawn out” vowel, as can be characteristic in infant directed speech (e.g., “Look at the kitty!” versus “Look at the kiiitty!”). Children were habituated to one word in the pair, at which point the other word was played, with an increase in looking time from the habituation trial indicating discrimination of the minimal pair. The 18-month-old Dutch-learning children discriminated the minimal pairs, but English-learning infants did not. The results from the Dutch-learning children are consistent with previous findings that phonological distinctiveness emerges by 18 months. As English-speaking children were tested on a foreign phonological variation that was not systematic according to the English phonology, this study did not directly test phonological constancy. However, the findings are consistent with the interpretation that 18-month-old children have phonological word forms that would be capable of resolving phonological constancy, as children grasped that according to their native phonology, this particular phonetic variation did not alter the phonological word form of the newly learned word. Emergence of phonological constancy by this age would thus coincide with the emergence of phonological distinctiveness.

But while this finding appears to suggest that by 18 months child have a grasp of phonological constancy for newly learned words, results from research on phonological distinctiveness suggest that for known words, phonological distinctiveness can emerge as early as 14 months (Fennell & Werker, 2003). To determine whether this is mirrored in phonological constancy as well, children's grasp of phonological constancy in known words must be examined.
A recent study examined this by using regional accent variations in pronunciations of words (Best et al., 2009). Pronunciations of a given word across regional accents can contain phonetic variation, but still maintain the same phonological structure, as in the example of the word NICE pronounced in American English ([na's]) and Australian English ([na's]), which both share the phonological structure, /nɑɪs/. Just as the Dutch minimal pairs in Dietrich et al. (2007) contained phonetic variation, but not variation to the phonological form of the word according to the English-speaking participants, making them suitable for examining children's grasp of phonological constancy, pronunciations of words across regional accents also separate phonetic and phonological variation, with the benefit of ecological validity, as they are natural variations of real words in the perceiver’s native language.

In a listening preference task based on the procedure developed by Hallé and de Boysson-Bardies (1994, 1996), Best et al. (2009) examined children's preference for listening to words that have a high frequency in toddler vocabularies (e.g., BOTTLE) over that for words having a low-frequency in adult vocabularies (e.g., BAUXITE), thus comparing their preference for familiar versus unfamiliar words. In two separate tests, one per accent, words were presented either in the children's native regional accent, American English (AmE), or in an unfamiliar regional accent, Jamaican Mesolect English (JaME). Nineteen-month-olds were tested, as based on previous results from Dietrich et al. (2007) they should have been expected to show an emergent grasp of phonological constancy, along with the complementary skill of phonological distinctiveness (Swingley, 2003; Swingley & Aslin, 2000; Werker et al., 2002). Fifteen-month-olds were also tested, as the
literature on phonological distinctiveness suggests that it is possible they are on the
cusp of developing phonologically specified word forms (Fennell & Waxman, 2010;
Fennell & Werker, 2003; Swingley & Aslin, 2002; Yoshida et al., 2009), though
see Section 6.3.1 *Emergence of phonological distinctiveness*, above, for our
interpretation of these findings as allowing access to otherwise less effective
phonetically specified word forms.

While it was found that both 15- and 19-month-olds preferred listening to
high-frequency toddler words over low-frequency adult words (i.e., familiar over
unfamiliar words) in their native regional accents, only for the 19-month-olds did
this preference extend to the non-native regional accent, suggesting a command of
phonological constancy at 19, but not 15 months. These results support the proposal
by Best et al. (2009) that phonological constancy emerges by 19 months, congruent
with the timeframe proposed for a command of phonological distinctiveness
(Swingley, 2003; Swingley & Aslin, 2002; Werker et al., 2002). Further, the study
failed to support the possibility that phonological constancy may be present as early
as 15 months for familiar words, as has been suggested by comparable studies on
phonological distinctiveness. Thus, it appears that children's word forms are fully
phonologically specified at 18-20 months, and that these forms are in a transitional
state at 14-15 months, with even familiar word forms specified at least partly on a
detailed, native-accent-specific phonetic level.

However, the interpretation remains that these results may not reflect a
perceptual shift in attention from phonetic to phonological detail between 15- and
19-months, but may instead reflect a developmental increase in cognitive capacity,
which can be applied to the task (e.g., Werker & Curtin, 2005). This issue will be further addressed in Experiment 1 (Chapter 8), as outlined in the following chapter.
7 Research Questions

7.1 Experiment 1

The motivation of this research project is to explore the development of spoken word segmentation and recognition in children, and in particular, its emergence as robust across segmental and suprasegmental variation, such as in adults (see Section 4.1.6 *Spoken word recognition is robust*). In order to accommodate efficiently to such variation, word forms must be phonologically specified (see Sections 5.3 *Phonological Knowledge of Word Structures is Required* and 6.1 *Requirements of Word Recognition*), and able to resolve between when phonetic variation changes a word's phonological form (phonological distinctiveness) and when it does not (phonological constancy). The ability to segment words despite both segmental and suprasegmental variation is expected to coincide with the emergence of phonological word forms. Research on phonological distinctiveness demonstrates that children can resolve such phonetic variation by
18-20 months (Swingley, 2003; Swingley & Aslin, 2000; Werker et al., 2002),
given that at this time children reliably discriminate minimal pairs, recognizing, for
example, that altering the initial /b/ in BABY to /v/, changes the word to the
nonword, VABY (Swingley & Aslin, 2000, 2002). Similarly, research on
phonological constancy also suggests a reliable grasp at 18-19 months (Best et al.,
2009; Dietrich et al., 2007), given that at that age children accept regionally
accented pronunciations as valid pronunciations of words, recognizing that that
phonetic variation maintains the word's phonological form (Best et al., 2009).

However, the possibility still remains that phonologically specified word
forms emerge as early as 14 months, at least in some form. Evidence from studies
that limit cognitive demand, either through providing more context or using simpler
tasks suggests that phonological knowledge may be present at 14 months (Fennell
& Waxman, 2010; Fennell & Werker, 2003; Swingley & Aslin, 2002; Yoshida et
al., 2009). It has been suggested that the reason that 14- to 15-month-olds fail to
demonstrate a grasp of phonological distinctiveness in some studies is due to the
difficulty of the experimental tasks masking their access to this information, rather
than to a poor grasp of the information itself (Fennell & Waxman, 2010; Fennell &
Werker, 2003).

In the exploration of the emergence of phonological constancy, then, it
remains possible that children's developing grasp of phonological constancy follows
a similar trend. That is, the results from Best et al. (2009) may not reflect a
perceptual shift in attention from phonetic to phonological detail between 15- and
19-months, but may instead reflect a developmental increase in available cognitive
resources required to recognize the non-native pronunciations sharing the same
underlying phonological structure of their own pronunciations, and an increased ability to access that knowledge, given the context-free listening preference task used in the experiment.

However, an alternative interpretation is that reduced cognitive demands did not uncover 14-month-olds’ emerging grasp of phonological distinctiveness, but instead facilitated access to their cognitively taxing phonetically specified word forms. The tasks that have reported success by 14- to 15-months-olds have all been tasks measuring phonological distinctiveness – that is, success in the task required children to grasp that a minimal pair change altered a word’s phonological form. Success in the familiar-word preference task can therefore result from either phonologically specified word forms, or phonetically specified word forms, as the difference between, for example, /b/ and /v/ in BABY and VABY simultaneously contains a phonological change that would be detected by phonologically specified word forms, and a phonetic change that would be detected by phonetically specified word forms. This possibility highlights the necessity of examining children’s grasp of phonological constancy as well, in order to reign in interpretations of the findings on phonological distinctiveness. Evidence about the emergence of phonological constancy would clarify whether children do in fact have phonologically specified word forms that can be accessed provided the task is simple enough, or instead overspecified phonetic forms of words that can likewise be used effectively when task demands are low. By our account, accessing phonologically specified word forms is less cognitively demanding than accessing overly detailed phonetic word forms. Accordingly, we believe that acquisition of phonological word forms provides the child with word recognition skills that are more robust to task
demands. Unlike tasks examining phonological distinctiveness, tasks that examine phonological constancy cannot be achieved with phonetically specified word forms. Such word forms would false alarm a phonological change, failing to resolve phonetic changes that do not alter a word’s phonological form. Thus, even if task demands are greatly reduced, 15-month-olds should be unable to successfully perform in a phonological constancy task if they, as we predict, do not yet have phonological word forms at that age. However, if 15-month-olds are able to demonstrate a grasp of phonological constancy in a task that has taken steps to reduce task demands, then the alternative interpretation would be correct, that phonologically specified word forms are present at 15 months, but are masked by a developmental deficit in general cognitive resources as compared to slightly older children (see also Best et al., 2010, 2008).

The aim of Experiment 1, therefore, is to clarify when children's grasp of phonological constancy has emerged – whether it has emerged by 19 months, as suggested by previous results (Best et al., 2009; Dietrich et al., 2007), or whether it actually emerges earlier, at 15 months. Merging the results from Experiment 1 with those from prior examinations of phonological distinctiveness will inform when children have developed fully specified phonological word forms – that is, word forms capable of resolving both phonological distinctiveness and phonological constancy, which will be the key to allowing reliable segmentation of words across phonetic variation.
7.2 Experiment 2

The aim of the remaining experiments is to examine the children’s progression towards mature methods of spoken word recognition. Experiment 2 examines children’s ability to segment and recognize words across suprasegmental variation that departs from native metrical patterning in connected speech. Current models of mature spoken word recognition consider suprasegmental cues to word boundaries to serve as supplementary cues to word recognition, with segmental correspondence to the lexicon being the primary source of information for lexical selection (e.g., Norris & McQueen, 2008). However, to segment words in this way, children must first build up a lexicon by segmenting words from the speech stream via other means, so that they can learn the word forms and attaching meanings to the segmented words. Thus for early word segmentation, the suprasegmental cues that serve as supplementary cues in mature spoken word recognition instead play a primary role for younger children. Infants demonstrate an early sensitivity to native language meter (Christophe & Morton, 1998; Mehler et al., 1988; Moon et al., 1993), and go on to exploit this sensitivity as a primary means to segment early words, by treating metrical boundaries as potential word boundaries, and segmenting words that adhere to metrical patterns in their language. At 7.5 months, children segment only (disyllabic) words that align with the predominant metrical pattern in their native language (e.g., trochaic stress in English), but by 10.5 months can segment words following other metrical patterns as well (e.g., adhering to an iambic stress pattern; Jusczyk, Houston, et al., 1999; see also Section 3.4.2 Children’s use of the MSS).
Thus Experiment 2 addresses when children transition from a primary to a more supplementary reliance on suprasegmental cues for word segmentation, by testing their ability to segment words across a synthesized accent comprised of native segmental information, but non-native suprasegmental variation to native metrical patterns. Variation to these patterns should affect their reliability as cues to word boundaries, and thus segmentation in spite of the imposition of non-native metrical patterning would suggest a shift in attention to segmental information.

Children's ability to tolerate suprasegmental variation in spoken word recognition in continuous speech may also reflect a point when their lexicon has reached a size such that segmenting words via correspondence of the segmental information for lexical items is now viable, but may not necessarily require phonologically specified word forms. We therefore suspect that the ability to handle suprasegmental variation may occur earlier than the ability to handle segmental variation, as suprasegmental cues for spoken word recognition can take on a supplementary role even before word forms become phonologically specified. Twelve-month-olds were tested as they have a parent-reported receptive vocabulary of about 39 content words (Dale & Fenson, 1996), and more importantly, are at the point where they are beginning to say their first words, signaling a developmentally significant point in language acquisition. Further, they are past the point when they only segment words according to the predominant metrical pattern of their language (Jusczyk, Houston, et al., 1999), and thus this age may represent a time when the MSS is beginning to take on a more supplementary role for spoken word recognition.
7.3 Experiments 3 and 4

The aims of Experiments 3 and 4 are twofold, with both experimental aims working towards clarifying the development and nature of phonological word forms, and their relationship with spoken word recognition. As will be further discussed below, the first aim is to determine whether the specification of phonological word forms differs across vowels and consonants. Secondly, Experiments 3 and 4 determine the robustness of phonological word forms – that is, whether they are able to overcome varying degrees of segmental variation from native pronunciations.

As discussed in Section 2.2 *Attunement to Native Consonants and Vowels*, vowel and consonant perception take different developmental paths. While at birth children are able to discriminate almost all consonant and vowel contrasts, by 6-8 months they begin to demonstrate some attunement to the vowels of their native language, showing directional asymmetries in discrimination of certain non-native vowel contrasts (Polka & Werker, 1994). A similar attunement follows later for consonants, at 10-12 months. Moreover, while this categorical language attunement for consonants persists, vowel perception takes on a qualitative change as children develop, so that vowels instead are perceived more gradiently, with a loss of the directional asymmetries present at 6-8 months (Polka & Werker, 1994), and with less of a clear divide between the perception of one vowel and another (Eimas, 1963). By this principle, perception of vocalic variation within words may be more forgiving than perception of consonant variation.

Thus it is possible that phonological specification of words differs for vowels and consonants. It may be that the specification of phonological word forms
follows a similar developmental disparity between vowels and consonants, or alternatively, perhaps phonological specification of vowels in phonological word forms is not even necessary. To address these issues, Experiment 3 and 4 test children's ability to recognize words from continuous speech across the native accent and another regional accent that contains only vocalic variation from the native regional accent (Experiment 3), and compare it to children's ability to do so in a regional accent that contains only consonantal variation from the native regional accent (Experiment 4). Based on experimental outcomes of Experiment 2, Experiment 3 tested 12- and 15-month-olds, whereas given the results of that study, Experiment 4 in turn also included 19-month-olds.

As discussed in Section 6.1.2 *Types of phonetic variation*, while pronunciations of words across regional accents share the same phonological structure, they can contain differing degrees of phonetic variation from one another. This degree of variation can be described according to whether it is within or beyond the variation allowed to the individual phonemes of the language. This level of specificity is below that of phonological word specificity. For example, to the native perceiver, regional accent pronunciations can contain phonetic variations that are sub-phonemic (Category-Goodness variation; CG), and map categorically to the corresponding phonemes to the native perceiver. Alternatively, phonetic variation can be phonemic, in that it maps to a different phoneme according to the native perceiver (Category-Shifting variation; CS). These two types of cross-accent variations and their predicted perceptual assimilation or non-assimilation to the corresponding phoneme in the native accent are based on the contrast assimilations put forth by the Perceptual Assimilation Model (Best, 1994b, 1995). Phonological
word forms are able to resolve both types of variation, whereas failure to resolve CS variation may indicate that the specification of word forms is of a lower order, i.e., remains phonetically overspecified. Thus, both Experiments 3 and 4 compared children's ability to recognize words in the face of CG versus CS between-accent variation.

The following experimental chapters have been written as stand-alone articles for journal submission. Experiment 1 (Chapter 8), has been submitted to a journal, and is now revised and will shortly be resubmitted, and Experiments 2-4 will also be submitted for publication once the thesis has been lodged.
Development of phonological constancy: 19-month-olds, but not 15-month-olds, identify familiar words spoken in a non-native regional accent

The pronunciation of a given word can display notable phonetic variation across utterances, which adult native perceivers are able to “hear through” in order to accurately and rapidly recognize the word. The sources of variation that confront speech perceivers range from between-speaker differences in vocal tract characteristics, to within-speaker differences in emotional state and speech style, through to pronunciation patterns that differ systematically between regional accents of the same language but are shared among speakers within each accent. To recognize spoken words despite this range of phonetic variation, perceivers must identify the underlying phonological form of the word that remains constant across the variations.

To discover the more abstract phonological form of a word, perceivers must determine whether a given phonetic difference alters the word’s underlying identity,
or leaves its identity intact. For example, an adult American English (AmE) perceiver who recognizes the Australian English (AusE) pronunciation of the word NICE [nɑːs] (low back vowel with mid front offglide) as equivalent to the AmE pronunciation [naɪs] (low central vowel with high front offglide) has grasped the phonological constancy (Best et al., 2009) of the word, which is not altered by the phonetic differences between AusE [ɑː] and AmE [a]. Conversely, perceivers must also recognize the complementary type of phonetic variation, that is, that which does signal a lexical distinction. For example, if the same American English perceiver recognizes that the AusE pronunciation [nəɪːs] (high front vowel with centralized onglide) is not NICE but the contrasting word NIECE (pronounced [niːs] in AmE, without an onglide), they have grasped the concept of phonological distinctiveness (Best et al., 2009) and generalized the word across the two accents.

Adults are experts at spoken word recognition because they make efficient use of phonological constancy and distinctiveness, the complementary principles that relate surface phonetic variations to more abstract phonological forms. Expert word recognition is thus flexible enough to accept speaker and accent variation, but exacting enough to accept only those variations that preserve the identity of the word.

By these definitions, the phonological form of a word is not simply a sequence of individual phonemic categories. In grasping phonological constancy, the perceiver recognizes that phonetic variation in a word can violate native-accent phonemic boundaries without changing the identity of the word, provided that it does so systematically. That is, perceivers can adapt to new phonemic boundaries in a non-native accent that are systematic in their relationship to the native regional
accent. As well, it is not necessary that systematic variation maintain the same
distinctions across regional accents, that is, they do not always have a one-to-one
equivalence across regional accents. This can be seen in cross-accent variation that
contains phonemic mergers that are still maintained as distinct contrasting
phonemes in the perceiver’s native regional accent. While this would not stop
perceivers from adapting to the variation, it may cause difficulties in their
recognition of novel words and lexical ambiguities.

Unlike adults, young children are word recognition novices. The ability to
discern the abstract phonological form of words takes time to develop based on
experience with the specific phonetic patterns of the native language as spoken in
their environment. Until they discover phonologically specified word forms, they
must recognize words as phonetic patterns, specifically those of the pronunciations
they have previously encountered. Phonetic patterns refers to the fine-grained detail
of the phones involved regardless of how they would be categorized phonemically
in the native accent, or when put in the context of a phonological word. This
includes any subset of the panoply of nuances that can be incorporated into
definitions of phonemes, including the results of coarticulation and allophony. This
is contrasted with phonologically specified word forms, which are word forms that
are represented without respect to purely concrete phonetic detail or to native
accent phonetic categories, but instead to more abstract phonological structure that
allows for supposition of systematic variation in phonemic categories (i.e., in a
regional accent), as described above. Thus, it can be seen that phonologically
specified, rather than phonetically detailed, word forms are required for cross
regional-accent perception.
Many theories on the development of word learning and recognition propose that as young children learn their native language, their recognition of spoken word forms does not become based on phonological principles until 18-20 months (e.g., Swingley, 2008; Thiessen, 2007; Werker & Curtin, 2005). That premise, however, is largely based on discrimination tests of minimal pair phonetic differences. At 14 months, discrimination of minimal pairs is unreliable, with some studies showing successful discrimination (Swingley & Aslin, 2002), and others showing a failure to discriminate (Stager & Werker, 1997; Werker et al., 2002). By 18-20 months discrimination of minimal pairs has become reliable (Swingley, 2003; Swingley & Aslin, 2000). However, if task demands are reduced relative to other studies in the literature, either by testing children’s discrimination of familiar rather than newly learned words (Fennell & Werker, 2003; Swingley & Aslin, 2002), or by increasing contextual support in studies using novel words (Fennell & Waxman, 2010; Yoshida et al., 2009), it appears that even 14-month-olds show some sensitivity to minimal pair differences. From these results it might be concluded that phonologically specified word forms may be emerging yet fragile at 14 months, and have become robust by 18-20 months.

Studies of minimal pair discrimination, however, can provide only a partial picture of the nature of early word form recognition. Such results can suggest but cannot confirm whether the older or especially the younger children have even achieved an understanding of phonological distinctiveness alone. To fully address the proposition that 18- to 20-month-olds have phonologically specified word representations, it is necessary to also examine the complementary principle of phonological constancy.
Efficient phonologically based word representations cannot be overly stringent and detailed, that is, phonetically *overspecified*, as this would lead to rejection of phonologically constant phonetic variations (e.g., an AmE-perceiver mustn't reject the AusE pronunciation of NICE as an acceptable variant of that word). On the other hand, they also cannot be too flexible and lacking sufficient differentiating detail, that is, phonetically *underspecified*, as this would lead to acceptance of phonetic variations that actually signify phonologically distinct words (e.g., the AmE perceiver must not accept the AusE pronunciation of NIECE as an acceptable pronunciation of NICE). Minimal pair discrimination tests can tell us only whether or not children's early word forms are *underspecified*; they cannot tell us whether they are *overspecified*. Critically, this means that minimal pair tests alone cannot pinpoint whether the children who succeed are attending to phonetic details or to phonological structure, as both types of difference are involved in the distinction between BABY and VABY. And as for the younger children who do not reliably succeed, we again can conclude only that they fail to reliably detect *either* the phonetic difference or the phonological difference.

Therefore, the possibility remains that children’s early word forms may be *overspecified*, comprised of very specific and even richly detailed phonetic patterns built up through experience with the native accent input via statistical learning or exemplar registration mechanisms. This is supported by studies showing that 7.5-month-old children are unable to recognize previously familiarized words when uttered by a different speaker of the same regional accent and same gender (Houston & Jusczyk, 2003), and do not segment words from passages when familiarized to phonetically similar foils (e.g., failing to segment BIKE from test
sentences when familiarized to GIKE: Jusczyk & Aslin, 1995). Where reduced task
demands have permitted 14-month-olds to discriminate minimal pairs (Fennell &
Waxman, 2010; Fennell & Werker, 2003; Swingley & Aslin, 2002; Yoshida et al.,
2009), this may have simply facilitated the children's access to less efficient,
phonetically (over)specified word forms, rather than uncovered an emerging grasp
of phonological distinctiveness.

To resolve these issues, the complementary skill of phonological constancy
must be examined. As phonological constancy is reflected in phonetic variation that
does not alter a word's underlying phonological form, testing phonological
constancy calls for word pronunciation differences that contain phonetic variation,
but not phonological changes (that is, stimuli that do not violate the phonological
form of the word). Using regional accent variations in pronunciations of words, a
recent study began the investigation of phonological constancy in young toddlers
(Best et al., 2009). Pronunciations of a given word in two different regional accents
of the same language generally share the same abstract phonological structure, but
still contain perceptible phonetic differences, as in the example of NICE in AmE
[naɪs] versus AusE [nɒs]. Because of this, phonetic variations can be separated
from phonological distinctions by careful selection of words according to the
naturally occurring pronunciation differences between two regional accents,
providing an ideal tool for examining recognition of phonological constancy.

Best et al. (2009) examined 15- and 19-month-olds' listening preference for
high frequency toddler vocabulary words versus low frequency adult words spoken
either in their native accent (AmE) or in a phonetically quite different regional
accent that the participants had not experienced previously (Jamaican Mesolect
English: JaME). They found that both age groups preferred listening to frequent
toddler words over unfamiliar low frequency adult words in AmE, indicating
recognition and preference for familiar words. However, only the 19-month-olds
also displayed this preference for the JaME-accented words, suggesting they had
gained some command of phonological constancy that the 15-month-olds had not
yet achieved.

To explain these results in relation to the prior developmental findings on
discrimination of minimal pair word modifications, Best et al. (2009) proposed a
perceptual attunement account, in which word forms are more phonologically
specified by 19 months, but not yet at 15 months, when they are still phonetically
defined and specific to the child’s native accent input. The younger children exploit
their perceptual attunement to the specific phonetic patterns in their language
environment as a means to recognize their first words, but this hinders their ability
to recognize words spoken in other regional accents. Thus they are able to
recognize words across speakers, affects, and the phonetic variations permissible
within the phonemic definitions of their native accent. However, increasing
exposure to between- and within-talker variation soon fosters the emergence of
more abstract phonological knowledge of the word structure by around the time of
the vocabulary spurt (around 18 months). The older children’s shift of focus to the
abstract phonological structure of words allows them to accept a range of phonetic
variations in word pronunciation that nonetheless preserve the invariant
phonological structure of the words. Thus, according to the Perceptual Attunement
account, as they mature children shift their attention from environment-specific
phonetic patterns to the higher-order, more abstract phonological structure of
words. This is what allows them to understand pronunciations they have never before encountered (Best, 1994b, 1995).

However, that may not be the only possible explanation of Best and colleagues’ findings. Alternatively, as noted earlier, phonological knowledge may emerge as early as 14 months for already known words, if the cognitive demands of the experimental task are reduced (as in Fennell & Waxman, 2010; Fennell & Werker, 2003; Yoshida et al., 2009). If that interpretation is correct, then the developmental shift observed by Best and colleagues (2009) could simply reflect a developmental increase in the cognitive resources children can bring to bear on recognizing phonological structure in non-native pronunciations of words they know.

But this alternative account was based solely on children’s discrimination of minimal pairs in their native accent (e.g., a change from /b/ to /v/ in BABY and VABY). As we have argued, success on minimal pair discrimination tasks can be achieved through a focus on either phonological structure or detailed phonetic patterns. Thus, Fennell and colleagues interpret these findings as indicating simply that reduced cognitive demands allow 14-month-olds to more easily access the full phonetic details of known words in their own accent. Importantly, tasks that examine phonological constancy cannot be successfully performed with phonetically detailed, input-specific word representations, as the phonetic differences between regional accents would false alarm that the word forms are different. Therefore, even if task demands were greatly reduced, 15-month-olds would fail on a phonological constancy task if, as the Perceptual Attunement account predicts, they retain a focus on specific phonetic details (overspecified word
forms, with regard to the native accent) and do not yet attend to phonological structure in words. However if they can detect phonological information under reduced task demands, then they should display knowledge of phonological constancy and recognize familiar words even when spoken in an unfamiliar regional accent. Such a finding would suggest that 15-month-olds do have phonologically specified word forms, which are fragile and easily masked by demands on their more limited cognitive resources than 19-month-olds are able to access for the task (see also Best et al., 2010, 2008).

To tease these possibilities apart, the present study examined whether 15- and 19-month-olds can identify familiar spoken words with their meaningful real-world referents in a task with relatively low cognitive demands. Specifically, we assessed whether they can match words spoken in the native accent versus an unfamiliar regional accent to their visual referents, as reflected in their direction of gaze between an image of the named word (target image) versus an unnamed distractor image (of a different known word). To reduce cognitive demands, we tested children’s identification of familiar words rather than newly learned words (see Fennell & Werker, 2003). Further, as research shows that the more contextual information available to a child, the more they are able to access known word forms (Fennell & Waxman, 2010), words were presented in a sentence context, which had the added benefit of increasing the child’s exposure to the systematic variation of the non-native regional accent. As well, a reward stimulus that played at the end of each trial regardless of performance reinforced the objective of gazing at the named picture, further reducing cognitive demands and helping to maintain the children's interest in the task. We also assessed each child’s expressive vocabulary, as
development of phonological distinctiveness in children's spoken word recognition has been linked with vocabulary size (e.g., Werker et al., 2002), as has their development of phonological constancy in the word-preferences task (Best et al., 2010).

If 15-month-olds do have a nascent grasp of the phonological structure of known words, but it is masked by their difficulties with cognitively taxing tasks, then the increased contextual support of the current lower-demand word recognition task should facilitate their access to phonological word structures, resulting in identification of the non-native pronunciations. Such findings would suggest that in Best et al. (2009) the 15-month-olds’ failure to recognize the accented pronunciations of familiar words was due to the demands of the listening preference task with sets of numerous known words versus sets of phonetically similar low frequency adult words, rather than being due to a lack of phonological knowledge.

On the other hand, if reduced task demands instead only facilitate young children's access to their more detailed phonetically specified word forms, then we would expect 15-month-olds to show above chance identification of native-accented words, but fail to identify the unfamiliar non-native accented pronunciations. By contrast, the 19-month-olds should be above chance across both regional accents as we expect them to have become attuned to the phonological form of the words, that is, we expect them to show phonological constancy across the accents.

As past results show a link between increasing vocabulary size and increasing phonological knowledge (Best et al., 2010; Werker et al., 2002), we further predicted that vocabulary size would be positively correlated with the ability of the 15-month-olds to identify words in the non-native regional accent. A
correlation with the native regional accent is not predicted, as phonological constancy is not required for word recognition in the native accent, where identification can be achieved through recognition of the familiar phonetic patterns. No correlations between vocabulary size and word recognition in the native or non-native regional accent are predicted for the 19-month-olds, on the other hand, as we postulate they have already gained a competent grasp of phonological constancy, which is no longer changing in tandem with the growth of their lexicon.

8.1 Method

8.1.1 Participants. Participants were 16 14.6- to 15.5-month-olds ($M = 15$ months, 3 days; $SD = 11$ days; 8 per gender) and 16 18.6- to 19.7-month-olds, ($M = 19$ months, 3 days; $SD = 11$ days; 8 per gender) from AusE-speaking households in Sydney, Australia, whose amount of exposure to non-native languages or non-AusE accents ranged from 0 to no more than 4 hours per week ($M = 0.34$ hrs/week). Participants were primarily Caucasian, from middle- to upper-middle-class households, and were recruited via advertisements in a regional parents’ magazine. Data from an additional 19 15-month-olds and 18 19-month-olds were collected but not included in the analysis due to fussiness or inattentiveness resulting in < 45% eye tracking for either test ($n = 17$), failure to complete both the native and non-native accent tests ($n = 12$), technical problems ($n = 7$), or parental interference ($n = 1$).

8.1.2 Stimuli and Materials.

8.1.2.1 Audio target words and sentences. Stimuli were produced by a male native speaker of general middle AusE, the children's native regional accent, and a
male native speaker of the non-native regional accent (JaME). During recording, the
speakers were instructed to produce the tokens as if they were speaking to a
toddler. JaME was selected as the non-native regional accent as it differs markedly
from AusE in phonetic realizations of its vowels, consonants, and prosody (Patrick,
1999; Wassink, 2006; see Best et al., 2009). Vowels in Jamaican Mesolect English
appear to retain a length distinction not present in Australian English, and many
consonants differ from Australian English consonants either in place or manner of
articulation, or both. Furthermore, using JaME in this study allows direct
comparison to previous findings on development of phonological constancy, which
used the same non-native accent but different word sets and recordings by different
female speakers (Best et al., 2010, 2008) and a different native accent (Best et al.,
2009). Eighteen target words were selected on the basis that they are easily
depictable, and appear in most toddlers’ early vocabularies, having a mean
frequency of 71% in 15-month-olds’ receptive vocabularies\(^3\) (Dale & Fenson,
1996). While receptive frequencies are not available at 19 months, it is expected
that the target words have a higher mean frequency in 19-month-olds' receptive
vocabularies. However, since all target words are classified as high-frequency
words for both age groups, it is not expected that the difference in frequency would
be responsible for any discrepancies in performance across age groups. The target
words and their phonetic realizations in AusE and JaME are presented in Table 1.
Phonetic realizations are given here as narrow IPA transcriptions, based on
judgments of three phonetically trained listeners. Target words were presented in

\(^3\) Receptive frequency for target word \textit{boat} at 15 months is not available. However, it has a 41% occurrence in 16-month-olds’ expressive vocabularies, which implies a notably higher frequency in receptive vocabulary. Moreover, it corresponds exactly to the average frequency for all our target words in 16-month-olds’ expressive vocabularies: 41% (Dale & Fenson, 1996).
four carrier sentences (Can you see the ____?; Where is the ____?; Let’s find the ____; Look at the ____). Four reward sentences (That’s the one!; There it is!; Yeah that’s right!; You got it!) were selected to serve as task reinforcement at the end of each trial. Multiple recordings of each carrier sentence, reward sentence, and target word were produced by the same AusE and JaME speakers. The final tokens of each target word (two per speaker per word) were selected based on similarity across accents in voice quality and IDS speech quality, by consensus among the authors and informal verification by other lab personnel.

Table 1

<table>
<thead>
<tr>
<th>Target Word</th>
<th>Phonetic realization</th>
<th>Phonetic realization</th>
</tr>
</thead>
<tbody>
<tr>
<td>baby</td>
<td>[pæibii]</td>
<td>[peibii]</td>
</tr>
<tr>
<td>cat</td>
<td>[kæt]</td>
<td>[kʰæt]</td>
</tr>
<tr>
<td>ball</td>
<td>[pɔl]</td>
<td>[paːl]</td>
</tr>
<tr>
<td>doggy</td>
<td>[tægi]</td>
<td>[tægi]</td>
</tr>
<tr>
<td>bathtub</td>
<td>[pətʰtəb]</td>
<td>[paːtʰtəb]</td>
</tr>
<tr>
<td>door</td>
<td>[toː]</td>
<td>[toɹ]</td>
</tr>
<tr>
<td>bear</td>
<td>[pɛɾ]</td>
<td>[piːɾ]</td>
</tr>
<tr>
<td>flower</td>
<td>[fləʊəɾ]</td>
<td>[fləʊəɾ]</td>
</tr>
<tr>
<td>birdy</td>
<td>[پɜːdiː]</td>
<td>[پɛɾdi]</td>
</tr>
<tr>
<td>hair</td>
<td>[heː]</td>
<td>[iːɾ]</td>
</tr>
<tr>
<td>boat</td>
<td>[pɔt]</td>
<td>[pɔt]</td>
</tr>
<tr>
<td>mouth</td>
<td>[məʊθ]</td>
<td>[məʊθ]</td>
</tr>
<tr>
<td>bottle</td>
<td>[pɔtɭ]</td>
<td>[pɑk]</td>
</tr>
<tr>
<td>paper</td>
<td>[pʰεɪpəɾ]</td>
<td>[pʰεɪpəɾ]</td>
</tr>
<tr>
<td>button</td>
<td>[pɛɾtən]</td>
<td>[pɔʔən]</td>
</tr>
<tr>
<td>spoon</td>
<td>[spʰən]</td>
<td>[spʰən]</td>
</tr>
<tr>
<td>car</td>
<td>[kʰɛɾ]</td>
<td>[kʰɛɾ]</td>
</tr>
</tbody>
</table>

*Note.* AusE = Australian English; JaME = Jamaican Mesolect English

In the original recordings, gathered for a separate experiment, many of the carrier sentence/target word combinations required for this task had not been recorded. In order to maintain consistency across stimuli, all carrier sentence/target combinations for the present study were created by splicing a token of the target word into the final position in a carrier sentence. Care was taken to ensure that
targets were spliced into carrier sentences that originally contained a word having a phonetically similar or identical onset (e.g., the target word BABY was spliced into a sentence originally ending with the word BALL). An additional token of each target word was selected to serve as the second repetition of the word in isolation. Due to differences in speaker rate, the resulting JaME sentences consisting of the carrier sentences and spliced target words were substantially shorter in duration (M = 1224 ms; SD = 150 ms) than the AusE sentences (M = 1668 ms; SD = 188 ms). Therefore, to ensure that the mean sentence duration did not differ significantly across the stimuli sets, the JaME sentences were lengthened (M = 1456 ms; SD = 187 ms), and the AusE sentences were shortened (M = 1439 ms; SD = 158 ms) using the duration manipulation function in Praat version 5.0.11 (Boersma & Weenink, 2008). The durations of the tokens used for the second repetition of the target words did not differ significantly across the AusE (M = 696 ms; SD = 150 ms) and JaME (M = 658 ms; SD = 131 ms) stimulus sets, and therefore were not altered. Naturalness of the resulting modified phrases and words was verified auditorily by three phonetically trained judges (co-authors CTB, MDT, and CK). A given item was accepted for the final stimulus set only if at least two of these judges agreed that it sounded natural.

8.1.2.2 Target and distractor images. For each word, two color photographic images were selected for visual depiction of the word. The two images depicting a given word were matched for size, clarity, and impressionistically for visual salience. Individual images were then paired with another image depicting a different word from the target familiar-words list to create target-distractor image pairs (e.g., the first CAR image was paired with the
second SPOON image, and the second CAR image was paired with the first BOAT image). Each target-distractor image pair was matched as best as possible on visual (e.g., color, complexity) and semantic (e.g., animals, objects) characteristics. Further, monosyllabic item names were paired only with monosyllabic distractors, and disyllables were paired only with disyllabic distractors. Each pairing was placed on an 800 x 600 pixel 10% gray background, placing one image on the left (center pixel: x = 160, y = 300), and one on the right (x = 640, y = 300; see Figure 1). Two such arrangements were created for each pairing so that an image appeared on the left and right side of the display equally often, and with equal designation of each item as the target and distractor image in a given condition (see Procedure and counterbalancing, below). All images measured 280 x 274 pixels, apart from the toothbrush and spoon, which were oriented diagonally and therefore were scaled to 200 x 196 pixels so that the length of the diagonal equaled the horizontal width of the other images. From a 95 cm viewing range, the full display subtended a 20.17° x 16.18° visual angle. The visual stimuli (apart from the toothbrush and spoon images) subtended a 5.89° x 5.82° visual angle, and the diagonally oriented toothbrush and spoon images, a 4.22° x 4.16° visual angle.
8.1.2.3 **Trial videos.** Audio and visual stimuli were combined into audiovisual videos, as detailed in *Procedure and counterbalancing*.

8.1.2.4 **The Australian English vocabulary inventory (OZI).** To measure the participants’ expressive vocabularies, the Australian English vocabulary inventory (OZI) was completed by the parent who brought the child in for the test session. The OZI is an adaptation of the MacArthur-Bates Communicative Development Inventory (MCDI): Words and Sentences form that is designed for use with 16- to 30-month-old children (Fenson et al., 2007). The OZI was developed at the Marcs Institute at the University of Western Sydney, and comprises two sections: an
expressive vocabulary measure, and a measure of developing grammar skills (Schwarz & Burnham, 2006). While the MCDI measures the vocabulary of AmE-learning infants and toddlers, the OZI was adjusted to better reflect contemporary AusE, as well as to shorten the overall administration of the measure by including only nouns, verbs, and descriptives in the inventory. Although normative data are still being collected for children in the age range tested here, preliminary results show that Australian children's scores on the MCDI and OZI are highly correlated for 24- and 30-month-olds (Schwarz & Burnham, 2006).

8.1.3 Apparatus and Setup. Participants’ gaze during the experimental task was measured using a Tobii X120 eye tracker (Tobii Technology AB) sampling at 120 Hz. This eye tracker is accurate within 0.5° and has a 0.2° compensation error for head movements, and has a 100 ms recovery time when tracking is lost. It implements both dark-pupil and bright-pupil technology to minimize data loss, and tracks both eyes simultaneously, which allows for data collection even when one eye is not being tracked. This binocular tracking also allows for correction of drift through continuous averaging of drift effects between the two eyes.

Two hand-drawn areas of interest (AOIs) were defined that coincided with the left and right images on the video monitor the infants viewed. Each AOI measured 323 x 468 pixels, and the pairs were positioned such that the left AOI had its center at 161.5 x 300 pixels, and the right at 638.5 x 300 pixels (see Figure 1). The same AOIs were used for all trials.

The testing room was set up with a 17” Diamond Digital LCD monitor (Mitsubishi Electric) 25 cm behind the back of the eye tracker, and with its lower edge positioned 26.5 cm above the table on which the eye tracker sat. The monitor
was angled at approximately 5° backward tilt (top tilting away from the child). Two Edirol MA-15D speakers were centered 41 cm below the tabletop. A Logitech QuickCam Orbit AF camera was placed 15 cm to the right of the eye tracker, allowing the experimenter to view participants from the adjoining room, and verify that participants' gaze was being tracked when they were oriented toward the screen. Stimuli were presented using Tobii Studio 2.0.2 software (Tobii Technology). So that the lag between presentation of the stimuli would not be dependent on the computer's speed, Tobii Studio was set to load one video in advance.

8.1.4 Procedure and counterbalancing. Participants were seated on their caregiver’s lap so that their eyes were 70 cm from the front of the eye tracker. For the duration of the study, caregivers wore Macally MTUNE headphones that played a mixture of music and speech, and were instructed to look down or to the side during the experiment. This kept caregivers blind to the experimental conditions, and also served to assure that it was the children's and not the caregivers' gaze that was tracked. Before testing began, the participant’s gaze was calibrated to a moving object presented once at the top left and once at the bottom right corner of the screen. For this purpose, a dynamic cartoon measuring 6.35 cm x 6.35 cm was paired with sound to attract and maintain the children's interest. The experimenter determined participants to be looking at the calibration stimuli when their gaze was fixed at a point on the screen at or in close proximity to the calibration object.

Following calibration, each participant completed two tests, one per regional accent, with 18 trials (one for each target word) per test. The entire procedure lasted about 8 minutes. A trial began with a central looming bull's-eye video clip that
played until the child fixated on it, at which point they saw a four-second silent video depicting the target-distractor image pair, which served to familiarize them to the pictures. Immediately following that, the test video began with the looming bull's-eye. Once the child fixated on the bull’s-eye, the carrier sentence began. The looming bull's-eye continued to play during the sentence, until the onset of the target word at the end of the sentence, at which point the target-distractor image pair replaced the bull’s-eye. In order that the onset of the target word and the corresponding image pair would occur at 1160 ms after fixation of the bull’s-eye for all trials, silence was inserted at the beginning of each carrier sentence as needed to compensate for the duration variations among the carrier sentences. Onset of the second target word repetition occurred 2000 ms after the onset of the word within sentence frame (at 3160 ms from the start of the test video). This created two time windows of 2000 ms for fixation analyses, one per word repetition. Animation (e.g., spinning, blinking) of the target image began 2000 ms after the onset of the second word repetition (5160 ms into the test video), which was on average 1302 ms (SD = 144 ms) after the word had ended (the words had a mean duration of 696 ms in AusE, and 658 ms in JaME). The reward sentence began playing 480 ms after onset of the animation (at 5640 ms); the animation continued until the reward sentence finished, at which point the trial ended and the looming bull’s-eye beginning the next trial began. As reward sentences varied in length, the total durations of the test videos ranged from 6600 ms to 7520 ms (M = 7001 ms; SD = 285 ms).

Trial order and regional accent order were counterbalanced across participants. Images were counterbalanced such that in each test the participant saw
all 18 target-distractor image pairings, arranged so that one of the two images corresponding to a target word (e.g., one of the two CAR images) appeared on the left side, and the other appeared once on the right side, with equal designation of each image as the target or distractor image across the two tests in a given condition. Raw gaze points to the computer screen were recorded throughout the trial by the Tobii X120 eye tracker.

8.2 Results

For the duration of the experiment, participants' gaze was tracked on average 72.5% of the time ($SD = 9.8\%$). The time participants’ gaze was not tracked was due primarily to looking away from the screen. Raw gaze points were converted to fixations by applying a fixation filter that compiled groups of gaze points occurring within 50 pixels and within at most 200 ms of each other. This is the recommended fixation filter for use with still images (as opposed to text or videos; Tobii Studio 1.2 user manual, 2008). This removed gaze points associated with saccades, so that only gaze points associated with fixations were measured. The fixation filter also served to compensate for missing data points over a short time period (e.g., when the participant was blinking) by replacing missing points with the mean point of the fixation.

The percentage of total fixation time that fell within the AOI of the target image was calculated by examining the durations of fixations from word onset to 2000 ms after word onset for each of the two word repetitions per trial, creating a 4000 ms window of analysis. The total duration of fixations to the target image in this interval was then divided by the total fixation duration to the target and
distractor image during the interval. Some analyses of infant gaze data for this type of task exclude the first 367 ms post word onset from analysis, as that is the presumed time it takes young children to carry out a saccade (e.g., Swingley & Aslin, 2000). Since we instead filtered the raw gaze data to only count fixation times, this approach simultaneously excluded gaze points belonging to saccades at any point during the analysis interval.

Percent fixation time to the target image was examined in 2 x 2 ANOVA comparing accent (AusE, JaME) as a within-subject factor and age group (15 months, 19 months) as a between-subjects factor. A main effect of accent was found, $F(1, 30) = 13.723, p = .001, \eta^2_p = .314$, reflecting that there was more looking to the target image in the native regional accent compared to the non-native regional accent overall. No other effects were found. Percentage of fixation time to the target image was then compared against chance (50%) for each age group and accent condition using one-sample $t$ tests, and percentage of fixation time to the target image across accent conditions for each age group was compared using paired-samples $t$ tests. Fifteen-month-olds' percentage of fixation time on the target image was above chance only for the native regional accent, $t(15) = 3.893, p = .001, 95\% \text{ CI} (4.212, 14.405)$; Figure 2). Their fixation to the target image in the non-native regional accent was at chance level, and the percentage of target fixation time differed across regional accents, $t(15) = 3.547, p = .003, 95\% \text{ CI} (3.232, 12.965)$. In contrast, 19-month-olds' percentage of fixation time on the target image exceeded chance in both the native, $t(15) = 5.352, p < .001, 95\% \text{ CI} (7.183, 16.691)$, and non-native regional accent, $t(15) = 4.474, p < .001, 95\% \text{ CI} (7.183, 16.691)$,
(4.370, 12.321), with the percentage of fixations to the target image not differing across regional accents, $t(15) = 1.649$, $p = .120$.

To examine the roles of vocabulary and/or age in predicting children's performance on the task, two stepwise regressions were carried out comparing log-transformed vocabulary and age (in days) against percentage of fixation time to the target in the native and the non-native regional accents, with $p < .1$ as the entry criterion. Vocabulary scores were log transformed to minimize any distorting effect of the rapid expansion of expressive vocabulary around the time of the vocabulary spurt. For the non-native regional accent, the log-transformed vocabulary score reached significance across the two age groups $\beta = .388$, $t(31) = 2.306$, $p = .028$, 95% CI (0.362, 5.974), explaining 15.1% of the variance in performance, $F(1, 31) = 5.317$, $p = .028$, while age did not add any predictive power to the

Figure 2. Mean percentage of fixation time to the target image. This was calculated by dividing total fixation time on the target image by the sum of fixation time on the target and distractor images. Both 15- and 19-month-olds fixated to the target image more than chance (50%) when target words were spoken in the native regional accent (AusE), but only 19-month-olds fixated to the target image more than chance also when target words were spoken in the non-native regional accent (JaME). Error bars represent standard error.

To examine the roles of vocabulary and/or age in predicting children's
Neither factor predicted children's performance in the native regional accent. Separate regression analyses were then performed for each age group, to examine vocabulary scores as a predictor of performance in the non-native regional accent. There was an effect of log-transformed expressive vocabulary size against 15-month-olds’ performance in the non-native regional accent, $\beta = .561$, $t(15) = 2.535$, $p = .024$, 95% CI (1.239, 14.867), explaining 31.5% of variance in performance, $F(1, 15) = 6.425$, $p = .024$ (Figure 3). Vocabulary size did not predict the 19-month-olds’ performance in the non-native regional accent (Figure 4).

![Figure 3](image)

*Figure 3. Mean percentage of fixation time to the target image versus log-transformed expressive vocabulary size at 15 months. Although 15-month-olds did not fixate on the target image more than chance in the non-native regional accent (Jamaican Mesolect English), log-transformed expressive vocabulary size is positively correlated with total fixation time on the target image in the non-native regional accent. Log-transformed expressive vocabulary size did not predict their performance in the native regional accent (Australian English).*
Fine-grained analyses were also conducted on looking patterns for the two accents, to determine whether there were more subtle differences in the looking behavior of the two age groups. In particular, if 15-month-olds did identify the words in the non-native regional accent, but at a slower rate than the 19-month-olds, their correct identifications could have been masked by examining total fixation time over the whole trial. Because application of a fixation filter to data disrupts the relationship between time and spatial location of participants’ gaze, we analysed the raw gaze data and normalized the participants’ gaze so that a value of 400 (pixels) represented the vertical midpoint of the screen, and looks with horizontal gaze location values above 400 corresponded to looking to the target image side of the screen, and those with values below 400 corresponded to looking at the distractor image side of the screen. These values were then averaged across all trials for each participant, and collected into bins with separate analyses.

Figure 4. Mean percentage of fixation time the target image versus log-transformed expressive vocabulary size at 19 months. Nineteen-month-olds fixated on the target image more than chance in both the native (Australian English) and non-native (Jamaican Mesolect English) regional accent, but expressive vocabulary size did not predict their performance in either regional accent.
conducted for bins of length 100 ms, 200 ms, 400 ms, and 500 ms. Bin values were then compared to chance via t test analysis across age group and accent to determine at which points in time participants' gaze was directed at the side of the screen containing the target or distractor image. No systematic differences between age groups emerged from these analyses, primarily due to the fact that the 15-month-olds' raw gaze data never differed from chance. Additional analyses compared the time point at which the maximum horizontal gaze towards the target image was achieved. A 2 x 2 ANOVA comparing time of maximum value for the between factors of age group (15 and 19 months) and within factor of accent (AusE and JaME) found only a main effect of age group, $F(1, 30) = 4.325, p = .046, \eta^2_p = .126$, reflecting that 19-month-olds reached their maximum gaze point sooner than 15-month-olds ($M_{19 \text{ mos}} = 2228.13 \text{ ms}, SD = 910.24 \text{ ms}; M_{15 \text{ mos}} = 2737.5, SD = 1138.97 \text{ ms}$). This suggests that 19-month-olds identify the target word sooner than 15-month-olds, but this interpretation is mitigated by the fact that 15-month-olds' raw gaze data never differed from chance. Gaze point data are presented in Figures 5 and 6. The patterns in the non-native regional accent condition seem to suggest that 15-month-olds show a similar looking pattern to 19-month-olds, but that this pattern is both delayed and weaker in comparison, likely reflecting 15-month-olds' inability to identify words in the non-native regional accent. However, the younger age displays too much variance in gaze to draw clear conclusions.
Figure 5. Horizontal location of gaze for 15- and 19-month-olds in the naive regional accent condition from onset of the first repetition of the target word. Onset of the second repetition of the target word occurred at 2000 ms, and shaded areas indicate the presentation of auditory stimuli. Data is corrected so that gaze above 400 reflects looking to the target image, and gaze below 400 indicates looking to the distractor image. While analysis did not reveal any differences in looking patterns across ages, analysis of total fixation time on the target image revealed both 15- and 19-month-olds identified target images in the native regional accent (Figure 2).
When compared against chance, 19-month-olds identified words spoken in both the native and non-native regional accent, with total fixation time on the target images exceeding chance in both conditions. Fifteen-month-olds, on the other hand, only identified words spoken in the native regional accent; their recognition of words in the non-native regional accent did not exceed chance. Word recognition did not differ across regional accents for the 19-month-olds, but did for the 15-month-olds. While these results do not demonstrate a clear effect of age, they are suggestive of a possible link between performance and age. Rather than age, expressive vocabulary may be a more suitable proxy for phonological development.

Figure 6. Horizontal location of gaze for 15- and 19-month-olds in the non-native regional accent condition from onset of the first repetition of the target word. Onset of the second repetition of the target word occurred at 2000 ms, and shaded areas indicate the presentation of auditory stimuli. Data is corrected so that gaze above 400 reflects looking to the target image, and gaze below 400 indicates looking to the distractor image. While analysis did not reveal any differences in looking patterns across ages, analysis of total fixation time on the target image revealed that 19-, but not 15-month-olds, identified target images in the non-native regional accent (Figure 2).

8.3 Discussion

When compared against chance, 19-month-olds identified words spoken in both the native and non-native regional accent, with total fixation time on the target images exceeding chance in both conditions. Fifteen-month-olds, on the other hand, only identified words spoken in the native regional accent; their recognition of words in the non-native regional accent did not exceed chance. Word recognition did not differ across regional accents for the 19-month-olds, but did for the 15-month-olds. While these results do not demonstrate a clear effect of age, they are suggestive of a possible link between performance and age. Rather than age, expressive vocabulary may be a more suitable proxy for phonological development.
Accordingly, expressive vocabulary size was correlated with the 15-month-olds' recognition of words in the non-native regional accent. It was predictive of their performance on words in the non-native accent, but not with the native accent nor with 19-month-olds’ performance on either accent.

These findings are inconsistent with the hypothesis that phonological knowledge emerges as early as 14 months. Given the simple nature and relatively low cognitive demands of the current task, the current findings support the notion offered by some researchers that 14-month-olds’ ability to discriminate minimal word pairs in previous studies suggests that low cognitive demands permit them to access phonetically detailed word representations, and not that they already possess phonologically specified word forms (e.g., Fennell & Waxman, 2010; Fennell & Werker, 2003; Yoshida et al., 2009). If they had discriminated minimal pairs via phonologically specified word forms, then they should have identified the non-native regionally accented words in the present study. Thus, the reduced demands simply appear to allow them better access to inefficient, detailed phonetic specifications of known words that they have experienced in their native accent.

The disparity between the 15- and 19-month-olds' recognition of words in the non-native regional accent in our procedure is instead compatible with the perceptual attunement hypothesis (Best et al., 2009) that children first become attuned to the phonetic information in native-accented words, but that further experience with natural variations shifts their attention to the more abstract phonological structure of words by around 17-18 months, a time at which children’s expressive vocabulary has begun to grow at a rapid rate. Early phonetically specific native accent lexical representations are too rigid to accept phonetically different...
non-native regional accent pronunciations, whereas later phonologically specified
lexical representations can accommodate the phonetically varied but phonologically
invariant non-native pronunciations. The important distinction is that the 15-month-
olds in this experiment are attending to phonemic categories that appear to be
activated and defined by fine-grained phonetic detail, as in the manner traditional
conceptualized by definitions of phonemic categories, whereas 19-month-olds have
made a further leap of abstraction where they can consider phonetic violations as
still belonging to a phonemic class in the context of word recognition. By this
account, of course, word recognition requires more than an understanding of native
phonemic categories, but this is not a controversial claim.

The present findings echo the challenge to phonological underspecification
accounts (Brown & Matthews, 1997; Metsala & Walley, 1998) set forth by previous
studies on phonological distinctiveness. The underspecification accounts propose
that children’s early lexical representations begin as global and underspecified, with
more detailed phonetic specification arising as needed to resolve ambiguities in the
expanding lexicon. Earlier findings that 14-month-olds can discriminate pairs of
words based on contrasts that are not required to separate minimal pairs in their
existing lexicon demonstrates that their lexical representations are more richly
phonetically specified than the phonological underspecification account requires.
Similarly, if 15-month-olds had underspecified lexical representations, they should
be more likely than 19-month-olds, who have more fully specified lexical
representations, to accept the non-native regional accent pronunciations. However,
exactly the converse was found.
In addition to being in line with the Perceptual Attunement approach, the current results are also consistent with other contemporary theories that posit phonological knowledge emerges by 18-20 rather than by 14-15 months (Swingley, 2003; Thiessen, 2007; Werker & Curtin, 2005). Further, while the emergence of phonological knowledge is unrelated to the presence of phonological neighbors in the child’s lexicon (as proposed in Brown & Matthews, 1997; Metsala & Walley, 1998), phonological knowledge is thought to emerge once children have learned a sufficient number of words from which they can draw phonological generalizations (Best et al., 2009; Swingley, 2003; Thiessen, 2007; Werker & Curtin, 2005). This knowledge is posited to derive either from the abstraction of phonemes from experienced native accent exemplars (Werker & Curtin, 2005), or from the distribution of phonemes in diverse lexical contexts (Thiessen, 2007), and/or from exposure to systematic variation in speech that allows children to derive the underlying phonological structure of the words (Best et al., 2009). This relationship with vocabulary size is compatible with our finding that although 15-month-olds overall did not identify target words in the non-native regional accent overall, their performance with the non-native accent was positively correlated with their expressive vocabulary size. It was not correlated with the 15-month-olds' performance in the native regional accent, which may be interpreted as evidence that phonological knowledge is not required for them to identify native pronunciations. This is important, as it suggests the emergence of a specific phonological skill – phonological constancy – rather than a simple improvement in cognitive skills. The lack of correlation between vocabulary and performance in either regional accent for the 19-month-olds suggests that they had already achieved
phonological constancy, such that their recognition of the non-native pronunciations was no different than that for the native pronunciations. This is consistent with recent findings linking lexical development to the presence of phonological constancy via children's ability to recognize non-native pronunciations of familiar words (Best et al., 2010). A similar correlation between expressive vocabulary size and the presence of phonological distinctiveness, via children's ability to learn new words containing minimal pair differences (Werker et al., 2002), demonstrates that vocabulary growth is linked with overall development of phonological knowledge, and is consistent as well with recent evidence that expressive vocabulary size is linked to phonological knowledge even in adult second-language learning (Bundgaard-Nielsen, Best, & Tyler, 2011a, 2011b).

An issue related to the development of phonological specification is how children learn to handle allophonic variation within native phoneme categories. In one study, Canadian English (CanE)-learning and Dutch-learning 18-month-olds’ awareness of language-specific linguistically relevant versus irrelevant phonetic detail was examined in a discrimination test using newly learned nonsense words that were minimal pairs by a vowel length difference that is contrastive in Dutch but not CanE (Dietrich et al., 2007). Short [ɑ] and long [aː] are lexically contrastive in Dutch, for example, in the words /stɑt/ (city) versus /staːt/ (stands), whereas they are both variants of a single vowel /a/ in CanE. In Dutch, therefore, the target stimuli used in the toddler study were minimal pairs, but in CanE they were simply phonetically varied tokens of the same phonological words. Children of each language group were habituated to word-object pairings for each item of the minimal pair (e.g., /stɑt/ paired with object A, and /staːt/ paired with object B), and
then tested on a switched pairing (/statt/ paired with object B). A change in looking
time indicated discrimination of the minimal pair. The Dutch-learning toddlers
discriminated the minimal pairs, but the CanE-learning children did not, suggesting
the former but not the latter group perceived them as phonologically distinct. This
study was not a direct test of phonological constancy, in part due to the use of
foreign words and a foreign speaker. Furthermore, the distinction involved may be
considered comparable to that between allophonic variants. This raises the question
of how children come to master perception of allophonic variation, and how this is
related to phonological perception. It is our contention that phonemic perception
includes the variation allowable within a category both in terms of its canonical
production and in terms of variations that follow from phonotactics and
cocarticulation. It is important to be clear that this encompasses variants such as non-
aspirated voiceless stops that appear after a sibilant in many varieties of English,
which are perceived as being members of the same phonemic category as initial
voiceless aspirated stops. The crucial distinction between allophonic variation and
cross-accent variation is that allophonic variation is experienced regularly and
systematically as part of quotidian exposure to the native regional accent, and is
thus part of what is learned in the development of a phonemic category. In contrast,
phonological constancy allows successful perception of a word even in cases where
the variation has never been experienced in the native regional accent.

Considering the results of the current study together with other research on
cross-accent word perception gives view to a more complete account of the
development of phonological constancy. Recent research on cross-accent speech
perception in infants has revealed that 12-month-olds, but not 9-month-olds, are
able to segment words familiarized in one regional accent from sentences in another regional accent (Schmale et al., 2010). This suggests that a grasp of phonological constancy appears to have begun to emerge at this age. However, this skill is likely fragile at this point, and limited to cognitively undemanding tasks like the recognition of word forms, which does not require any semantic processing. This notion is supported by recently completed research that shows an inability of 12-month-olds to segment word forms across accents when the stimulus load is increased (Mulak, Best, Tyler, & Kitamura, in preparation). The current study demonstrates that by 19 months, however, children’s grasp of phonological constancy has strengthened to the point that they are able to recognize known words across accents, in a paradigm that requires both semantic and word form processing. This development of phonological constancy continues, however. In an even more cognitively demanding task that required participants to learn novel word-object associations, only 30-month-olds, and not 24-month-olds, were able to generalize newly learned words across accents (Schmale, Hollich, & Seidl, 2011). This study tested children’s ability to generalize across a foreign accent, which may contain more non-systematic variations and thus be more cognitively demanding to resolve. Thus, children’s developing grasp of phonological constancy can be seen in the children’s performance on increasingly demanding tasks. This sensitivity to task difficulty is consistent with results showing pre-exposure to a specific accent facilitates cross-accent word recognition (White & Aslin, 2010; see also Schmale et al., 2011), and is a characteristic mirrored in the literature outlining children’s developing ability to access phonetic detail in speech (see Fennell & Werker, 2003; Werker & Curtin, 2005).
Overall, the current study together with other recent research on cross-accent word perception (Schmale et al., 2010; Schmale & Seidl, 2009; White & Aslin, 2010) underline the importance of examining phonological constancy as the complement to phonological distinctiveness in order evaluate overall phonological knowledge. Both of these phonological abilities need to be considered in current theories of language development. The concept of phonological constancy appears to have particular utility for separating the use of phonologically versus phonetically specified word forms. The Perceptual Attunement account (Best et al., 2009) specifically addresses the complementary relationship between these two aspects of phonological knowledge. The results of the present study support a developmental transition in perceptual attunement from attention specifically to phonetic detail, toward greater reliance on phonological structure by 19 months. That this point in development coincides with the vocabulary spurt, and that expressive vocabulary size correlates with recognition of non-native accented words during the transitional period, taken together, are highly suggestive of a link between an increasing lexicon and developing phonological knowledge. This relationship may be a complex one, in which the need to rapidly expand one's lexicon encourages attunement to the systematic phonological properties of words, which in turn reduces the cognitive load required to access a known word. Attunement to the more cognitively efficient phonological properties might then bootstrap further vocabulary growth, allowing children in the midst of the vocabulary spurt to go on to increase their lexicon at a previously unmatched rate as they form more abstract and versatile lexical representations.
In this experiment we examined variation to vowels, consonants, and suprasegmental features. There is reason to believe that there may be differences in how each of these, and in particular vowels and consonants may emerge in the specificity of children’s word recognition. For instance, vowels are perceived less categorically than consonants, with less of a clear divide between the perception of one vowel and another compared to consonant perception (Eimas, 1963), and consonants and vowels appear to serve different roles in word perception by young children (e.g., Bonatti, Peña, Nespor, & Mehler, 2007; Hochmann, Benavides-Varela, Nespor, & Mehler, 2011), with consonants being tied to word identity, and vowels linked more to structural features. Recent results suggest that in terms of developing phonological specificity, vowels, consonants, and suprasegmental features may follow different developmental paths (Mulak et al., in preparation; Mulak, Best, Tyler, & Kitamura, 2012).
9 Experiment 2

Twelve-month-olds may segment words from connected speech in the face of non-native suprasegmental features

Reliable comprehension of spoken language requires the mastery of two tasks: word recognition and word segmentation. Perceivers must be able to segment the words from running speech, which is not a straightforward task as word boundaries are not reliably marked by pauses or other acoustic cues (Cole & Jakimik, 1980). They must also be able to recognize the segmented word. This requires a perceiver to possess a substantial lexicon, as well as mature phonological knowledge of those words in order to overcome phonetic variation, from that between different speakers of the native language, through to the systematic variation that can occur between cross-regional accent pronunciations.

In adults, word segmentation occurs primarily in tandem with word recognition. Models such as Shortlist B (Norris & McQueen, 2008) and TRACE (McClelland & Elman, 1986) assume that knowledge of word structure, or the
sequential order of segments comprising a word, is in place and highly automatized. These models invoke attention to segmental cues and lexical competition as the main methods for determining word boundaries. In this way, the incoming stream of phonemes leads to activation of candidate words, according to both their probability and the compatibility of the segmentation with known words. Thus, once a word is recognized it is segmented, and the process continues with the subsequent phoneme stream.

Without a substantial lexicon of common words, adults would be unable to segment spoken language in a way that is useful for natural communication in realtime. But infants learning their first language must segment words from running speech so that they can attach meanings to the words and add them to their lexicon. Thus children are faced with the paradoxical task of requiring a lexicon to segment the speech stream, but needing to first segment the speech stream to build their lexicon.

Instead, infants use a variety of lower-level (phonetic) and statistical cues to discern word boundaries, separating out words and slowly beginning to associate these words with concepts. One tool they use is tracking transitional probabilities of syllables in running speech to determine word boundaries. For example, in the phrase “pretty baby,” the probability that “ty” follows “pre” is higher than that of “ba” following “ty,” and a dip in transitional probabilities between syllables may provide evidence of a word boundary (Saffran et al., 1996). Children (as well as adults [Saffran, Newport, et al., 1996]; and other animals [Hauser, Newport, & Aslin, 2001]) are able to track and make use of these distributions (Friederici & Wessels, 1993; Goodsitt et al., 1993; Jusczyk et al., 1994; Mattys & Jusczyk,
2001a; Mattys et al., 1999; Saffran, Aslin, et al., 1996). Another strategy infants can employ is the use of phonotactics – the rules that govern permissible sequences of phonemes and where they can appear in words in a given language – to help narrow down possible word boundaries (Mattys & Jusczyk, 2001a, 2001b; Mattys et al., 1999). Sequences of phonemes that violate native language phonotactics appear only across syllable and/or word boundaries (e.g., the sequence /mdpr/ cannot occur within a syllable or across syllable boundaries, but can occur across word boundaries, as in JAMMED PRINTER). Similarly, infants may use allophonic variations to cue word boundaries (Jusczyk, Hohne, et al., 1999), as some allophones only occur at certain lexical and/or syllable positions – for example, in American English /k/ is realized as the aspirated long-lag VOT allophone [kʰ] word-initially, but as unaspirated short-lag VOT [k] following /s/ or /ʃ/ or in stress- or word-final position (e.g., ACTIVE [ˈaktɪv] or TACK [tæk]).

Infants also make use of suprasegmental, or prosodic cues, such as rhythm, metrical patterning, and intonation of the utterance. The effectiveness of this approach is likely enhanced by the exaggeration of suprasegmental features in infant-directed speech (Fernald & Kuhl, 1987; Kuhl et al., 1997). Evidence shows that infants attend to these cues, for example via their preference for prosodically exaggerated infant-directed speech over adult-directed speech (R. P. Cooper & Aslin, 1990; Fernald, 1985; Pegg et al., 1989), their ability to discriminate languages belonging to different rhythmic classes (e.g., Mehler et al., 1988), their sensitivity to clause (e.g., Hirsh-Pasek et al., 1987) and phonological phrase (Gout et al., 2004; Soderstrom et al., 2003) boundaries, and their preference for the predominant metrical pattern in their native language (e.g., Houston et al., 2000;
Jusczyk et al., 1993; Jusczyk, Houston, et al., 1999). By using suprasegmental cues, infants are able to home in on likely word boundaries, reducing the number of possible words into a much more manageable set of those fitting within the suprasegmental boundaries (e.g., Shukla et al., 2007), with word segmentation then further supplemented by clues brought forth by the tracking of segmental features.

In infants, the metrical patterns of a language provide the primary suprasegmental cues for informing locations of likely word boundaries (Cutler, 1994; Cutler & Carter, 1987; Jusczyk, Houston, et al., 1999), and using these patterns to inform word boundaries is referred to as the Metrical Segmentation Strategy (MSS; Cutler & Norris, 1988). Languages can be broadly categorized as having one of three classes of metrical patterns: syllable-timed, mora-timed, and stress-timed (see Cutler & Otake, 2002, for a review). In syllable-timed languages, such as French and Italian, native listeners are sensitive to the boundaries between syllables as possible places for word boundaries. Likewise, in mora-timed languages listeners are sensitive to boundaries between morae. In stress-timed languages, such as English, listeners are sensitive to and use the boundaries between stress units, which in English most typically include an initial strong syllable and any following weak syllables (Cutler & Otake, 2002).

Infants begin to use their language-specific metrical patterns for word segmentation around 6 months (Morgan, 1996), and by 7.5 months rely heavily on those cues for segmenting words (Jusczyk, Houston, et al., 1999). The majority of polysyllabic words in English are disyllabic, and have a trochaic stress pattern (a stressed syllable followed by an unstressed syllable). English-learning infants at this age typically segment words based on a trochaic rhythm, so much so that words not
following this pattern are mis-segmented. For example, when presented with the spoken phrase, “guiTAR is” (with capital letters indicating the stressed syllable), infants incorrectly perceived TARis as a word they had heard before (Jusczyk, Houston, et al., 1999). This native-stress preference for segmentation of possible words has been found across languages, for example with Canadian-French-learning infants segmenting words based on *iambic* stress patterns (an unstressed syllable followed by a stressed syllable), the principal stress pattern in that language (Polka et al., 2002). While children continue to segment words via the predominant rhythmic pattern of their native language at nine months (Jusczyk et al., 1993), infants begin to segment words that do not demonstrate the predominant rhythmic pattern in their language by 10.5 months (E. K. Johnson, 2005; Jusczyk, Houston, et al., 1999), suggesting a progression toward a more adult-like word segmentation method by as early as 10.5 months.

Although access to metrical patterns is still available to adults, these cues are no longer crucial for word segmentation, with such cues instead taking on a supplementary role. This is apparent from adults' ability to segment words correctly despite segmental and suprasegmental variation. Adults are capable of understanding spoken words in a non-native accent – that is, spoken by someone who is not a native speaker of the language (e.g., Australian English speakers listening to English words produced by a native speaker of Italian), though this can be more cognitively taxing compared to perception of words spoken in a person's native language and accent (Snijders et al., 2007). Non-native accented speech can contain considerable segmental and suprasegmental variation from native pronunciations, and is often produced according to the speaker's native metrical
class – for example, a native Italian speaker speaking English may produce it as syllable-timed English, rather than stress-timed (see Gut, 2003). Moreover, segmental and suprasegmental variation can introduce many systematic irregularities into non-native accented speech, as it is based on the phonology and suprasegmental properties of the speaker's native language, which may not map systematically to the target language.

As early word segmentation relies strongly on suprasegmental information, particularly meter, compared to mature word segmentation, we examined at what point children can segment words in the face of non-native suprasegmental variation, to help inform when children have progressed to this more adult-like word segmentation. One study has looked at children's ability to handle suprasegmental (along with segmental) variation in segmenting words from continuous speech, across a non-native accent. Schmale and Seidl (2009) exposed English-learning 9- and 13-month-olds to words spoken in either their native regional accent (American English) or a foreign accent (Spanish-accented English), and then tested their listening preference for passages in the unfamiliarized accent that contained the familiarized words versus those that did not. They found that 13-month-olds preferred listening to passages containing the familiarized words regardless of the accent, indicating that they were able to recognize words across a non-native accent.

However, interpretation of the Schmale and Seidl (2009) study is limited in that the stimulus materials conflated segmental with suprasegmental variation, thus the findings do not straightforwardly address our specific research aims. As well, while the study did limit the amount of exposure to a second language a participant
could have, it did not report limiting the amount of exposure to accented English, and the amount of additional language exposure allowed was considerably high (30%). Therefore, while the results are intriguing, it cannot be ruled out that many of the children may have had mixed language and/or accent exposure, especially to the Spanish accent which is increasingly frequently encountered in the U.S., and research suggests that exposure to variation in words can alter the word forms that young children can recognize (e.g., Singh, 2008). Accordingly, in this study we have taken care to ensure that participants have minimal-to-no exposure to accented varieties of English.

In the present study, we examined children's ability to overcome suprasegmental variation independent of their ability to overcome segmental variation by using sentences containing native segments, but non-native suprasegmental variation. Recordings were made in the children's native accent (AusE) as well as in French-accented English (FrE), and the native segmental information was synthetically combined with the non-native suprasegmental information. French was chosen as the non-native language for this purpose as it is a syllable-timed language, in contrast to English, which is stress-timed. Additionally, whereas disyllabic words tend to have trochaic stress in English, disyllabic words in French have a tendency towards iambic stress.

A serial preference procedure was used, in which children were familiarized to trochaic disyllabic words spoken in their native accent (AusE), and were then tested on resynthesized sentences that retained AusE segmental structure, but had the prosodic contour imposed from either another speaker of AusE, or from a native speaker of French, who produced the English sentences as their late-learned second
language (post-puberty). The two types of comparison sentences in each case contained either the familiarized words or new, unfamiliarized disyllabic words.

Twelve-month-olds were tested because while they tend to have small receptive vocabularies (parents report a receptive vocabulary of about 39 content words [Dale & Fenson, 1996]), they are beginning to say their first words, signaling a developmentally significant age. Further, they are past the point when they only segment words from connected speech according to the predominant metrical pattern of their language (Jusczyk, Houston, et al., 1999), and thus this age may represent a time when this approach is beginning to take on a more supplementary role for spoken word recognition. We thus predicted that along with segmenting words in the native accent condition, 12-month-olds would also be able to segment words in the non-native suprasegmental information condition.

9.1 Method

9.1.1 Participants. Participants were 22 11.6- to 12.5-month-olds, \((M = 12\) months, \(SD = 9\) days; 10 females, 10 in native condition) from AusE-speaking households in Sydney, Australia, whose amount of exposure to non-native languages or non-AusE accents ranged from 0 to no more than 3 hours per week \((M = 0.4\) hrs/week). Data from an additional 9 toddlers were collected but not included in the analysis due to fussiness.

9.1.2 Stimuli and Materials.

9.1.2.1 Target words. Eight target words appearing in early-toddler vocabularies were selected from the Australian English vocabulary inventory (OZI; Schwarz & Burnham, 2006). All target words were selected on the basis of their
trochaic (strong-weak) syllable pattern in AusE. To ensure the AusE and French-accented pronunciations contained the same number of segments and thus could be morphed, only AusE pronunciations containing a vowel in the weak syllable (versus a syllabic consonant, e.g., [l] or [n]) were selected, as none of the French-accented pronunciations that we recorded contained any syllabic consonants. Word frequencies were balanced across groups as having a 30.1% (SD = 8.4%) occurrence in 15-month-olds' expressive vocabulary. A summary of the target words can be found in Table 2.

9.1.2.2 Carrier sentences. Carrier sentences were also constructed from words available on the OZI, in order to match average sentence word frequency with target word frequency. As the OZI does not contain articles, pronouns, or auxiliary verbs, the 100 most common English words (as based on the Oxford English Corpus) were also allowed in sentence construction, though information on the occurrence of these words in early vocabularies was not available, and thus not considered in the sentence averages. In the original construction, care was taken to limit contextual information for the target word by pairing two words together and creating six nonsense sentences into which either word could be placed. Thus, for the current studies, while each target word contains a unique set of six carrier sentences, they remain context-free. Sentences contained 8-10 syllables ($M = 8.7$), and were controlled so that a carrier sentence's words did not differ in frequency from its target word by more or less than 20% ($M = |7.5|\%$), and had an average frequency of 32.1% ($SD = 8.0\%$). The target word occurred equally at either the

---

4 Word frequencies were balanced according to a 15-month-olds' vocabulary as that is the median age group tested throughout the experiments presented in this thesis. It is not expected that the relative frequencies between words in the set would differ across months.

5 http://oxforddictionaries.com/page/oec
beginning (with onset at the second syllable of the sentence), middle, or at the end of the sentence. As well, target words were preceded equally by either a strong or weak (e.g., preceded by an article) syllable. A summary of the carrier sentences for each target word can be found in Table 2.

Table 2

*Target words and carrier sentences*

<table>
<thead>
<tr>
<th>Target Word</th>
<th>Carrier Sentences</th>
</tr>
</thead>
</table>
| orange      | Five orange shops were on my street.  
My orange was put on my plate.  
You can get a nice orange at some shop.  
I got my orange from a policeman.  
Wipe Vegemite off Pop's orange.  
You got wet soup on my orange. |
| present     | Bear's present stands on the new stairs.  
The present was as long as a broom.  
The man and blue present flew on a goose.  
Nanna looked for the present from the zoo.  
On the porch there was a good present.  
A dog ran and scratched a present. |
| pizza       | Good pizza was on the new shoes.  
The pizza was on the chair for you.  
A boy wants pizza for a dog.  
A man had pizza and corn for Nanna.  
Pour blue sauce on the bear's pizza.  
Juice splashed and flew on the pizza. |
| picture     | Four picture rooms stand on the roof.  
Their picture looks good because of the bush.  
You would have a new picture of corn.  
Thank you for the picture and shorts.  
The mad deer ran on the frog's picture.  
Nanna splashed sauce on the picture. |
| garbage     | Bad garbage was food for the bear.  
The garbage flew from the blue bush.  
Four dogs share garbage on a roof.  
A broom could push the garbage on the porch.  
The sad kangaroo could want garbage.  
A horse and lamb looked for the garbage. |
**raisin**
- One raisin sat on Nanna’s street.
- My raisin was put on a cot.
- Soup splashed and got Pop’s raisin so wet.
- On my raisin was a cute butterfly.
- Come and look at my cat's white raisin.
- At lunch I sat on my raisin.

**tiger**
- Pop’s tiger sat on some white plates.
- My tiger was lost on a boat.
- You’ve got one tiger and a wet cat.
- Someone put a tiger on a cot.
- Please get me a cute new tiger.
- Don’t put Vegemite on my tiger.

**hammer**
- No hammer can do what my friend’s does.
- My hammer was broken by a boy’s bat.
- What a nice hammer my nurse gave me.
- I can use a hammer or a boat.
- I will get you just one hammer.
- A policeman does not want my hammer.

*Note.* Shading indicates paired words.

### 9.1.2.3 Speakers
Target words were produced by a female native speaker of AusE from Western Sydney, Australia. The sentences were produced by two additional female native speakers of AusE, who were from Western Sydney and Northern Sydney, Australia, respectively. The sentences were also produced by a native speaker of French from Moissac, France, who could read English at a fluent level and produced the sentences in French-accented English. All speakers were of similar age ($M = 25$ years, 7 months, $SD = 2$ years, 3 months) and were selected from a total of four recorded AusE and four French-accented speakers based on similar vocal quality and infant-directed speech (IDS) affect.

It should be noted that varieties of AusE are not completely homogeneous. For instance, high, middle, and broad varieties are widely recognized. Increasingly, distinctions are being made between regional accents of Australian English as well (see Cox & Palethorpe, 2001). However, compared to the grosser variations found
in British English or AmE varieties, the variations found in AusE are generally accepted to be quite minor, particularly in regard to suprasegmental variation. The AusE accents used in this study are all of the middle variety, but have minor regional distinctions. The accent used for the target words, as well as the segmental information of the carrier sentences is Western Sydney AusE. The suprasegmental information for the native carrier sentences is Northern Sydney AusE (see *STRAIGHT Manipulations*, below). However, for simplicity, the accent as a whole will be referred to as AusE.

**9.1.2.4 Recordings.** Recordings were made in a sound attenuated room. Participants were first asked to tell, in their native language, a 1-minute story from their childhood to encourage their native language mode. They were then shown the target words and sentences on a monitor connected to an LG laptop. Opa 1.0 software, developed at the Marcs Institute (Antoniou, Best, Tyler, & Kroos, 2010), was used to present the target words and sentences in a random order. Opa has the added benefit of allowing the administrator to make online quality judgements regarding the stimuli, such that if an utterance is not acceptable it is not counted, and the target or sentence can be placed back into the random queue to be read again. Participants were instructed to say the words and sentences as if they were speaking to a toddler, and spoke into a Shure SM10A headworn microphone, using an Edirol UA-25 sound card, which fed into CoolEdit 2000 (Syntrillium Software) on a PC laptop computer. Recordings were made into one channel, at 44,000 Hz. This channel was then doubled to create a stereo monophonic recording.

**9.1.2.5 STRAIGHT Manipulations.** Test sentences were created using STRAIGHT resynthesis of the selected natural speech recordings. STRAIGHT is a
MATLAB program developed for high quality speech morphing (Kawahara et al., 2008; Kawahara, Masuda-Katsuse, & de Cheveigné, 1999). STRAIGHT has been used successfully in speech research as a source of strikingly natural-sounding synthetic speech stimuli (Matsui & Kawahara, 2003; Yonezawa, Suzuki, Mase, & Kogure, 2005), and is particularly well-suited for applications involving manipulation of F0 features for auditory morphing. STRAIGHT is in essence a VOCODER, taking auditory input and converting it into three time-based maps: a spectrogram, an aperiodicity map, and an F0 map. These can be manipulated independently, making it easy to morph from one sound to another based on these features prior to resynthesis. TANDEM-STRAIGHT, used here, is a more recent implementation of STRAIGHT (Kawahara et al., 2008), that uses an improved, temporally-stable method of calculating the F0 trajectory.

To test children's perception of non-native suprasegmental information separately from non-native segmental information, AusE and French-accented English recordings of the same sentences were synthetically combined to create one sentence, composed of the AusE segmental information (Western Sydney accent) and the suprasegmental information from French-accented English (henceforth referred to as AusE-FrE sentences). Prior to manipulation, the amplitudes of sound files were each normalized in CoolEdit 2000. The sound files were imported into STRAIGHT and converted to STRAIGHT Objects, which comprised spectrum, frequency, aperiodicity, F0, and time arrays. Corresponding segmental features in AusE and FrE sentences were marked using Praat (Boersma & Weenink, 2008), and these served as temporal anchors prior to morphing. A continuous morphing space was then generated from the two stimuli. A morphing value of 1.0 corresponded to
the original unaltered FrE token, while a morphing value of 0.0 corresponded to the original unaltered AusE token. In this way, the values for the spectrum, frequency, and aperiodicity arrays were set to 0.0 (AusE), while the time and F0 arrays were set to 1.0 (FrE). The target utterance was then resynthesized into a morphed speech file. After manipulation, the morphed target items were noise reduced using CoolEdit 2000, and scaled to 65 dB. AusE control sentence stimuli were created likewise, using the same AusE speaker from the AusE-FrE stimuli to provide the segmental information, but with suprasegmental information coming from recordings of another AusE speaker (Northern Sydney accent). The AusE target word recordings to be used in the familiarization were not manipulated using STRAIGHT, but were noise reduced and scaled to 65 dB.

9.1.3 Procedure. In this serial conditioned fixation preference procedure, participants were seated on their caregiver’s lap in the middle of a room in front of a central video computer monitor. A video camera positioned to the right of the monitor both recorded the session and provided a live stream to the experimenter in the adjacent control room. Each trial began with a blinking checkerboard on the monitor. Once the participant oriented to the screen for a minimum of 400 ms, the checkerboard stopped blinking, and speech played from two speakers hidden behind the monitor (initiated by a keyboard press by the experimenter). A new trial would start once the participant turned away for at least 2000 ms, at which point the checkerboard would resume blinking. If a participant turned away for less than 2000 ms, the same trial continued once the participant reoriented to the monitor for at least 100 ms, with the time spent looking away not counting towards the total looking time of that trial. A trial would also time out once 1400 ms of looking time
accumulated, at which point the checkerboard would resume blinking for 1850 ms before the next trial commenced, in order to maintain attention of the participant. For the duration of the study, caretakers wore Macally MTUNE headphones that played a mixture of music and test stimuli, keeping them deaf to experimental conditions. Likewise, the experimenter was kept deaf to the test stimuli played to the participant.

**9.1.4 Design and counterbalancing.** All participants were familiarized to the isolated target words produced by the AusE speaker. Participants were familiarized to two pairings (four words total) of target words presented at alternating trials, until they accumulated 30 s of familiarization (looking time) to each pair. Once children received 30 s of familiarization time to one pair, it was removed from the stimuli set and no longer played during familiarization, such that familiarization time for the two pairs was equal. Half of the participants were familiarized to the target word pairs ORANGE/PRESENT and PIZZA/PICTURE, and the other half were familiarized to the word pairs GARBAGE/RAISIN and TIGER/HAMMER. Participants were then tested with two blocks of four test trials, presented either in AusE for half of the participants, or in FrE for the other half of participants. Each trial consisted of the sentences for one of the four word pairings (the two familiarized pairs, and the two unfamiliarized pairs), such that each four-trial test block included all four trial types. The sentences that played for the duration of the trial (i.e., while the participant was oriented towards the monitor) were randomly selected without replacement from the designated sentence set for that trial, so that the child would not hear the same sentences twice within the same trial.
9.2 Results

Children's total looking times were compared in a 2 x 2 repeated measures ANOVA with familiarity (sentences containing familiarized versus unfamiliarized words) as a within-subjects factor, and condition (native accent suprasegmental features versus non-native suprasegmental features) as a between-subjects factor (Figure 7). No effects were found. To look more closely at any differences in looking times between sentences containing familiarized versus unfamiliarized words, we calculated the proportion of looking time for sentences containing familiarized words for each participant by dividing the total looking time to sentences containing familiarized words by the total looking time spent looking to sentences containing familiarized and unfamiliarized words. The proportion of looking time to sentences containing familiarized words was compared to chance for each accent group. One-way ANOVA analysis comparing proportion looking to the sentences containing familiarized words by accent condition was not significant, but further t test analysis revealed a trend towards a novelty preference was found only for the non-native FrE accent, \( t(11) = -1.800, p = .099 \) (see Figure 8), while no difference was observed in the native AusE accent.
Figure 7. Total looking time to sentences containing familiarized versus unfamiliarized words in the native accent suprasegmentals (AusE) and non-native suprasegmentals (FrE) conditions. Children's total looking times did not differ across accent or familiarization of words embedded in sentences. Error bars represent standard error.
9.3 Discussion

Contrary to our expectations, the participants did not discriminate sentences containing familiarized words from those containing unfamiliarized words in the native accent. The expectation that they would show such an effect was based on outcomes from previous studies implementing this procedure, which have demonstrated that at 12 months (e.g., Schmale et al., 2010), and as early as 7.5 months (Jusczyk & Aslin, 1995) children discriminate such stimuli in their native accent, and prefer listening to the sentences containing familiarized words.

Figure 8. Percentage total looking time for sentences containing familiarized versus unfamiliarized words in the native suprasegmentals (AusE) and non-native suprasegmentals (FrE) conditions. Participants showed no reliable looking preference for sentences in the accent that displayed native suprasegmental information, but showed a trend towards a novelty preference for sentences containing unfamiliarized words in the accent containing non-native suprasegmental information. Error bars represent standard error. Values within each accent group are reciprocals, but have been included for clarity.
However, in order to provide sufficient challenge to the robustness of children's ability to overcome suprasegmental variation, our study had modified the procedure to familiarize the children to four words, whereas previous studies have used only two familiarized words. Further, the exposure time to familiarization stimuli in these studies has generally been 30 s per word. In this study the exposure time to each pair of words was 30 s familiarization time, but with two pairs of familiarized words, this effectively halved the familiarization time per word as compared to previous studies (Schmale et al., 2010; e.g., Schmale & Seidl, 2009). It might seem, then, that the added cognitive demand may have overcome their ability to perform the task at all. However, this conclusion is challenged by children's trend towards discrimination in the non-native accent test trials. Instead, we propose that the added demands of the larger stimulus set may have affected which information the participants attended to during the task.

It has been proposed that task demands can interfere with children's access to linguistic detail, accounting for performance differences between children of a given age across different tasks, as well as performance of children of different ages within the same task (see Fennell & Werker, 2003). For example, in a switch task, in which children were taught a new word via habituation to the pairing of a novel auditory word (e.g., BIH) and novel visual referent, 14-month-olds failed to reliably notice when the auditory word incorporated a single, critical phonetic change from the taught word (e.g., DIH), indicating that 14-month-olds did not encode, or were not attending to, the fine phonetic detail of the newly learned word (Stager & Werker, 1997). But 14-month-olds could succeed in such a task when the cognitive demand of the task was reduced, either by presenting words containing gross rather
than minimal differences (e.g., LIFF vs. NEEM; Werker & Stager, 2000), or by testing discrimination of known words (e.g., BALL vs. DOLL), thus not requiring word to be learned through the course of the experiment (Fennell & Werker, 2003).

This suggestion that task demands may block attention to linguistic detail is extended in the PRIMIR model of speech processing, which proposes that such demands interact with other factors (such as age, biases) to influence which aspects of linguistic detail are attended to, so that attention is diverted, rather than blocked (Werker & Curtin, 2005). In this way, certain tasks can elicit attention to detail that would otherwise be expected, at that developmental stage, to be overridden by attention to another source of linguistic detail.

In the developmental course of word segmentation, we have seen that at 9 months, children appear to segment words from running speech based on their adherence to the predominant metrical pattern in their language (Mattys et al., 1999; Thiessen & Saffran, 2003). By 10.5 months this dependence on the predominant metrical pattern for detection of word boundaries falls, so that children are able to segment words with a variety of metrical patterns, which may indicate that they have reached a stage where they are able to integrate multiple cues, including segmental structure, to better determine word boundaries (Jusczyk, Houston, et al., 1999).

Returning to the present study, we propose that our expanded stimulus set increased the cognitive demand of the study. In the native accent condition, this increased demand resulted in children falling back on their use of metrical patterning to discern possible word boundaries from the sentences. While the word boundaries outlined by the metrical patterning were congruent with those of the
target words, this shifted their attention away from the segmental detail of the sentences. This change in attentional focus failed to provide them with the segmental information they needed to distinguish the familiarized words from the unfamiliarized words in the native accent condition, and thus children showed no discrimination of sentences containing familiarized words from those containing unfamiliarized words.

For the non-native accent, however, the test sentences were manipulated so that the expected, native metrical patterning would be disrupted. French-accented English was specifically chosen as French contains fixed stress (Dupoux et al., 1997), with words tending towards iambic stress, whereas in English, multisyllabic words predominantly carry trochaic stress. In effect, the French metrical patterning in the non-native accent condition rendered the suprasegmental structure of the sentences incoherent from an English-learner's perspective. We propose that children fell back on attending to the metrical patterning. In the native accent, attention to these cues drew attention away from segmental detail. However, in the non-native prosodic accent, the non-native metrical patterns were incomprehensible to these English-learning children, thus leaving their attention focused instead on the segmental information. As the segments in this non-native metrical condition were native, the children were perhaps more able to discriminate the sentences containing familiarized words from those containing unfamiliarized words. This interpretation is consistent with evidence that segmentation without reference to suprasegmental information can be accomplished by children of this age, as can be seen in artificial language learning experiments, where speech is presented in a monotone fashion (e.g., Saffran, Aslin, et al., 1996; Saffran & Wilson, 2003).
The failure of the 12-month-olds to segment the familiarized words from the sentences in their native accent implies that their word segmentation abilities as demonstrated in previous studies (e.g., Jusczyk & Aslin, 1995; Jusczyk, Houston, et al., 1999), at younger ages and under reduced stimulus loads, may not be so fully robust as to occur outside of a simple experimental context even by 12 months. Compared to similar studies, the current study familiarized children to four words (as opposed to two), and gave children 15 s familiarization time per word (as opposed to 30 s). Although sentences were composed entirely of early vocabulary words, this familiarity with surrounding words still did not result in children's ability to segment familiarized words from the native accent sentences. Thus while word segmentation is regularly observed in children of these ages in a laboratory setting, if this same skill cannot persist in the face of moderately increased task difficulty such as was used here, this implies that it is less likely to apply in a meaningful way outside of a very limited experimental setting.

Children's tendency to discriminate sentences containing familiarized words from those containing unfamiliarized words in an accent containing non-native suprasegmental features, despite their complete inability to do so with sentences that do have native suprasegmental features, supports the idea that children's performance in a given task may be affected by task demands and changes in attentional focus. Given the variety of situations and linguistic demands placed on children outside of the laboratory, a realistic approach to describing development should involve focusing on the point at which a skill or ability becomes robust, and persists despite many task demands and factors even in laboratory tasks.
Still, that 12-month-olds demonstrated a tendency to segment words from a resynthesized accent containing non-native suprasegmental features, but native segmental information, indicates that they may beat a stage where they are not only able to segment words based on the segmental structure, but may also be able to do so when the suprasegmental information is less- or non-informative. By our proposed explanation, 12-month-olds' tendency to discriminate sentences containing familiarized words from sentences containing unfamiliarized words in the non-native suprasegmental condition may have been due to their ability to ignore the non-informative suprasegmental features. Thus, rather than resolving the non-native suprasegmental features through attention to the higher-order phonological structure of a word, it may have instead been ignored, allowing attention to the segmental information of the familiar versus unfamiliar target words. In mature word segmentation, suprasegmental features are believed to take on a supplementary role (e.g., Norris & McQueen, 2008) – however, a distinction is made between the process here and mature segmentation of non-native accents, in that in this study, native segments were used. Still, if non-informative suprasegmental information is ignored in mature spoken word recognition, the lack of supportive prosodic cues to word boundaries could account in part for the increased cognitive processing load found in adult recognition of words in non-native foreign accented speech (Snijders et al., 2007). We believe the results from the present experiment suggest that by 12 months children may still fall back on the MSS for word segmentation in cognitively demanding situations, but may still be able to segment words on the basis of the segmental structure, and may also be able to direct their attention away from metrical patterns when such information is non-informative. We further
believe that this tendency to segment words across suprasegmental variation is not necessarily tied to the development of phonologically specified word forms, and may instead be more reliant on development of a sizeable receptive lexicon. While further investigation is needed to clarify this claim, the present results may reflect the beginnings of a developmental transition from immature word segmentation that over-relies on suprasegmental features when the task is demanding, to more mature spoken word recognition that can focus on segmental features regardless of the suprasegmental features even when the task is demanding.
10 Experiments 3 and 4

Development of phonological specification of word forms: 15-month-olds' sensitivity to vowel variation, and 19-month-olds' sensitivity to consonant variation from the native accent

Successful speech perception involves rapid segmentation of individual words from a connected speech stream that does not reliably denote word boundaries with pauses or other acoustic discontinuities. In adults, word segmentation occurs in tandem with word recognition. In models such as Shortlist B, the incoming segmental sequence is compared with the existing lexicon, with word recognition resulting from lexical competition of words on the basis of their consistency with the incoming segments in the speech stream (e.g., Norris & McQueen, 2008). Successful word segmentation in this manner thus requires a large lexicon. However, in order to build such a large lexicon, infants must first discover word boundaries by other means, before learning word structures and attaching meaning to them.
Early word segmentation by infants makes use of transitional probabilities (e.g., Goodsitt et al., 1993; Jusczyk et al., 1994; Mattys & Jusczyk, 2001a; Mattys et al., 1999; Saffran, Aslin, et al., 1996), phonotactics (e.g., Mattys & Jusczyk, 2001a, 2001b; Mattys et al., 1999), metrical patterns (e.g., Cutler, 1994; Cutler & Carter, 1987; Jusczyk, Houston, et al., 1999), and allophonic variations (e.g., Jusczyk, Hohne, et al., 1999) to discern word boundaries so that the children can separate out the words and begin to associate these words with concepts. For example, in American English /k/ is realized as the allophone [kʰ] word-initially, but as [k] following /s/ or /ʃ/ or in stress- or word-final position (e.g., ACTIVE [ˈaktɪv] or TACK [tæk]). Sensitivity to these systematic properties can serve as a reliable cue for word boundaries. In this way children slowly develop a vocabulary, which, like adults, they can then exploit to better inform the location of word boundaries (see Brent, 1999a, 1999b; Brent & Cartwright, 1999; Dahan & Brent, 1999) until their lexicon is large enough to begin to segment words in a mature manner as adults do, via lexical competition of words on the basis of their consistency with the incoming segments in the speech stream (e.g., Norris & McQueen, 2008). Thus a large portion of the cues infants implement for early word segmentation, such as allophonic variation and phonotactics, stem from the segmental information in speech, with word segmentation ultimately advancing toward a primary reliance on segmental information.

Along with the challenge of building their lexicon, children are faced with another hurdle as they progress toward mature word segmentation. Despite adults' primary reliance on the segmental structure of speech for word segmentation, they are able to segment words across considerable variation from this information.
Pronunciations of the same utterance by two people will vary from each other phonetically, resulting from individual differences in vocal tract characteristics through to the systematic variation seen between different regional accents of the native language. Yet, from experience we know that adults are readily able to understand speech across speakers, and research has shown that adults are able to process speech across unfamiliar regional accents with only slight delays in processing as compared to speech in the native regional accent (Floccia et al., 2006). Recognition of speech across such phonetic variation in segmental realizations requires perceivers to overcome variations in speaker pronunciations to resolve the underlying phonological structure of a word – that is, the word intended by the speaker. Specifically, perceivers must grasp when phonetic variations, be they sub-phonemic or phonemic, maintain a word's underlying phonological structure, that is, display phonological constancy, versus when they do not, i.e., when the two utterances display phonological distinctiveness. For example, phonological constancy is reflected in the ability of Australian English (AusE) perceivers of the American English (AmE) NICE ([naɪs]) to recognize the underlying phonological word as NICE, despite the vowel change from the AusE pronunciation, [nɑːs]. Conversely, phonological distinctiveness is reflected in the ability of AusE perceivers to recognize that another vowel change, from [naːs] to [niːs] delineates two separate phonological words, NICE and NIECE, respectively. Thus, word recognition must eventually reach a stage where it is flexible enough to accept speaker variation, but exacting enough to accept only those variations that preserve the identity of the word.
By 10.5 months children are able to segment words containing phonetic variation occurring across different voices and affects (Houston & Jusczyk, 2000; Singh, Morgan, & White, 2004), indicating that by this age children are no longer sensitive to variation within their native regional accent that does not alter a word's phonological structure. However, it is not until 18-20 months that children demonstrate a reliable grasp of both phonological distinctiveness (e.g., Swingley, 2003; Swingley & Aslin, 2000; Werker et al., 2002) and phonological constancy (Best et al., 2009i; Dietrich et al., 2007; Mulak, Best, Tyler, Kitamura, & Irwin, resubmitted; Chapter 8) capable of resolving greater variation. Children's grasp of phonological distinctiveness is apparent from their ability to reliably detect when a phonetic change alters a word's phonological structure for familiar (e.g., Swingley, 2003) as well as newly learned (e.g., Swingley & Aslin, 2000; Werker et al., 2002) words - for example, they can discriminate the familiar word BABY from the nonword VABY, or newly learned word BIH from DIH. Likewise, at this age children's grasp of phonological constancy is reflected in their ability to understand pronunciations of familiar words across regional accents, demonstrated by 19-month-olds' ability to identify visual referents of words pronounced in an unfamiliar regional accent (Mulak et al., in preparation; Chapter 8) and can be inferred through their inability to discriminate a newly learned word from a minimal pair stimulus that is not lexically-contrastive in their language (Dietrich et al., 2007).

Recently, Best et al. (2009) proposed a perceptual attunement account, converging with contemporary accounts that posit that phonologically specified word forms emerge at 18-20 months (e.g., Swingley, 2008; Thiessen, 2007; Werker & Curtin, 2005). By this account, a perceptual shift in attention from phonetic detail...
to abstract phonological information around the time of the vocabulary spurt (17-18 months) results in a shift in the specificity of word forms (see also Nazzi & Bertoncini, 2003). As early word learners, children attend to the phonetic patterns in their language environment as a means to recognize their first words, but as they mature they become attuned instead to the higher-order, more abstract phonological features of words (see Best, 1994b, 1995). Importantly, the perceptual attunement account proposes that exposure to systematic variation in the language allows children to derive a word's underlying phonological structure around the time of the vocabulary spurt. Around that time, attention shifts to the phonological structure of words rather than their specific phonetic forms, allowing children to understand pronunciations they have never before encountered (Best et al., 2009).

However, while accounts of the development of phonologically specified word forms have supposed that specificity for consonants and vowels occurs together, there is evidence to suggest this may not be the case. In adults, while consonants tend to be perceived categorically (Liberman et al., 1967), vowels tend to be perceived in a more continuous manner (Beddor & Strange, 1982; Fry et al., 1962; Polka, 1995; Stevens et al., 1969). It has been proposed that consonants play a more prominent lexical role than vowels in speech perception (Nespor, Peña, & Mehler, 2003). In adults, consonants play a more salient role than vowels in lexical access (Cutler, Sebastián-Gallés, Soler-Vilageliu, & van Ooijen, 2000; New, Araújo, Bour, & Nazzi, 2008; New, Araújo, & Nazzi, 2008), and in determination of word boundaries via an increased tendency to track transitional probabilities across consonants rather than vowels in artificial language learning tasks (Toro, Nespor, Mehler, & Bonatti, 2008). In children, a consonant bias is also found in
word-learning tasks. In a name-based categorization task, children were taught two novel names corresponding to two novel objects that were minimal pairs either by consonants or vowels. They were then taught that a third object shares the name of one of the other objects, and asked to indicate the object with the matching name. Sixteen- (Havy & Nazzi, 2009), 20- (Nazzi, 2005; Nazzi & New, 2007), and 30-month-olds (Nazzi, Floccia, Moquet, & Butler, 2009) were able to indicate the object with the matching name when the names of the two original objects were minimal pairs for consonants, but not when they were minimal pairs for vowels.

Furthermore, research has also revealed divergent developmental courses for the specificity of consonants and vowels in words. In a preferential looking task, 18-month-olds discriminate correct pronunciations of familiar words (e.g., BALL) from mispronunciations containing a consonant substitution (e.g., GALL) or a vowel substitution (e.g., BULE) marked by reduced looking to the corresponding image compared to the correct pronunciation condition. However, 15-month-olds discriminated only mispronunciations containing a consonant difference, and treated vowel mispronunciations as correct pronunciations (Mani & Plunkett, 2007).

While the bulk of research has looked at children's handling of segmental variation for individual words, these studies cannot inform about the effect of such variation on word segmentation. That is, they cannot inform whether children have developed mature, robust word segmentation strategies as are seen in adults. However, a recent study (Schmale et al., 2010) looked at children's segmentation of words across regional accents, as pronunciations of the same word across regional accents may contain systematic phonemic and/or sub-phonemic phonetic variation, but maintain the same underlying phonological structure. Nine- and 12-month-old
American English (AmE) learners were familiarized to words either in their native AmE, or in Canadian English (CanE), and then tested with sentences in the alternate accent, that contained familiarized or unfamiliarized words. The stimulus set across the regional accents differed primarily with respect to the vowels /ae/ and /ɪ/, which are lowered and more backed in CanE than in AmE; otherwise these two accents are quite similar. It was found that 12-, but not 9-month-olds, preferred sentences that contained familiarized words versus unfamiliarized words, suggesting that at 12 months, children are able to segment words in the face of phonetic variation to native vowels.

While Schmale et al. (2010) demonstrated that at 12 months children are able to segment words in a regional accent containing phonetic variation to a subset of vowels, children's ability to segment words containing cross-regional accent variation to consonants has not yet been examined. Moreover, the two English accents in their study were quite similar segmentally, except for the modest differences in the two vowels. Thus, we compared children's ability to segment words across two more notably different regional accents, one in which all words were selected to differ from participants' native Australian English (AusE) in their vowels (Experiment 3; Tyneside English [TE], spoken in Newcastle, England), and another in which all words were selected to differ from AusE in their consonants (Experiment 4; southeast London English [seLE], spoken in London, England). Children were familiarized to familiar words spoken in their native AusE, and then tested on their ability to segment the words from passages spoken in the non-native regional accent using a serial preference procedure. To isolate children's ability to overcome segmental variation from interference in suprasegmental variation, the
non-native accent test sentences were resynthesized so that they contained non-native segmental information, but native-accented suprasegmental information.

Furthermore, we also examined children's ability to handle differing degrees of phonetic variation. While pronunciations of a given word across two regional accents share the same phonological structure, such variation can differ in its degree of deviation from the perceiver's native pronunciation. Variation can either comprise a sub-phonemic change, that is, a Category-Goodness variation (CG), or they can incorporate a phonemic change, that is, a Category-Shifting variation (CS) to a vowel or consonant within the word. These variations and their predicted assimilations or non-assimilations are based on the contrast assimilations put forth by the Perceptual Assimilation Model (PAM; Best, 1994b, 1995). By this account, a CG assimilation occurs when a given consonant or vowel produced by a speaker of another regional accent is perceived by a native AusE speaker as the same phoneme in AusE, though perhaps an odd example of it. For example, the realization of initial or medial /r/ in seLE [ɹʷ] or [ʋ], differs from the AusE realization [ɹ], but would still be perceived as /r/ by an AusE speaker. These CG-type variations, since they still correspond to the same phonemic category for the listener, are predicted to be more easily overcome than CS assimilations, which occur when the phonetically varied consonant or vowel is perceived as a different phoneme by a native AusE speaker. For example, the seLE realization of medial /ð/ as [v], differs from the AusE realization as [ð], and would be perceived as /v/ by an AusE speaker. These predictions are supported by myriad evidence when dealing with consonant perception in both adults (e.g., Best & Hallé, 2007; Best, McRoberts, & Goodell, 2001; Best, McRoberts, & Sithole, 1988) and children (e.g., Best,
McRoberts, LaFleur, & Silver-Isenstadt, 1995). However, vowel perception by adults appears to be less categorical (e.g., Beddor & Strange, 1982; Stevens et al., 1969), and thus the predictions set forth by PAM may be less applicable to vowel perception.

### 10.1 Experiment 3

To determine when children can segment words from connected speech containing systematic variations in their vowels, thus informing when children possess a robust phonological knowledge for vowels, we examined 12- and 15-month-olds' ability to segment words across a regional accent that contained marked variation in its vowels as compared to the native regional accent, and compared this with their ability to segment words from sentences in the native regional accent (AusE). Tyneside English (TE), spoken in Newcastle, England was used as it contains marked vowel differences from the participants' native regional accent, AusE, but its consonant realizations are quite similar to those of AusE. In a serial preference procedure, children were familiarized to four words spoken in their native regional accent, and tested on sentences that either contained or lacked the familiarized words, which were either spoken by another speaker of AusE, or by a native speaker of TE. In the TE test condition, target words contained either a single CG (sub-phonemic), or a single CS (phonemic) vowel variation from the AusE pronunciation. To isolate children's ability to overcome segmental variation from interference resulting from suprasegmental variation, the accented sentences were resynthesized so that they contained non-native segmental information, but native suprasegmental information.
Expanding on our previous findings (Chapter 9), participants were familiarized to twice as many words as in Schmale et al., (2010), that is, four words instead of two. Successful segmentation in this setting of increased demand would imply a robustness in segmentation ability that is more likely to correspond to success in a natural setting.

Based on findings that 12-month-olds are able to segment words across regional accents in which the stimuli contain variation to two vowels (Schmale et al., 2010), we looked at 12-month-olds' ability to segment words across TE. Further, while children do not appear to have developed fully specified phonological word forms until around 18-20 months (Best et al., 2009h; Swingley, 2003; Swingley & Aslin, 2000; Werker et al., 2002 see also Experiment 1), some research suggests that specificity for vowels and consonants may follow different developmental courses (e.g., Mani & Plunkett, 2007), and thus we suspected that phonological specification for vowels (and therefore success in this task) may develop by 15 months.

Thus, we predicted a difference in performance between 12- and 15-month-olds. While we expected both 12- and 15-month-olds would be able to segment familiarized words from sentences in their native regional accent, we expected the vocalic variation would overcome 12-month-olds' ability to perform the task in the non-native segments condition. Thus we predicted a greater distinction between familiarized and unfamiliarized words to be made by 15-month-olds than by 12-month-olds in the non-native accent condition, and specifically that only 15-month-olds would show successful segmentation of vowel-differing words from the accent containing non-native segmental information.
10.1.1 Method.

10.1.1.1 Participants. Participants were 22 11.5- to 12.6-month-olds, 
\((M = 12 \text{ months}, 1 \text{ day}; \ SD = 9 \text{ days}; 9 \text{ females}; 12 \text{ in native condition})\) and 23
14.7- to 15.6-month-olds, \((M = 15 \text{ months}, 5 \text{ days}; SD = 9 \text{ days}; 5 \text{ females}; 12 \text{ in}
native condition)\) from AusE-speaking households in Sydney, Australia, whose
exposure to other languages or non-AusE accents ranged from 0 to no more than 4
hours per week \((M = 0.3 \text{ hrs/week})\). Data from an additional 26 toddlers were
collected but not included in the analysis due to fussiness \((n = 25)\) or parental
interference \((n = 1)\).

10.1.1.2 Stimuli and Materials.

10.1.1.2.1 Target words. Eight disyllabic target words appearing in early-
toddler vocabularies were selected from the Australian English vocabulary
inventory (OZI; Schwarz & Burnham, 2006). All target words had a trochaic
(strong-weak) stress pattern, and were selected according to their expected vowel
assimilation patterns (CG or CS) between the non-native and the native regional
accents. Target words were limited to those containing only one variant vowel, with
all other vowels (and all consonants) in the words native-like. A summary of the
target words, and their assimilations can be found in Table 3. There were three
unique CG and three unique CS vowels used. One CG and one CS vowel each were
repeated in order to create a full set of eight words, comprising four words with CG
assimilation vowels between the non-native TE accent and the native AusE accent,
and four words with CS assimilation vowels between the two groups. Word
frequencies were balanced across groups, with the average frequency (based on a
15-month-olds' expressive vocabulary) set at 31.5% \((SD = 6.8\%)\).
Table 3

Target words for Experiment 3, their assimilation type, target vowels, AusE and TE broad transcriptions, and AusE assimilations of target words for CS only (CG variants are perceived as categorically the same word as in the naive pronunciation).

<table>
<thead>
<tr>
<th>Target Word</th>
<th>Assimilation type</th>
<th>Target phoneme</th>
<th>Transcription</th>
<th>AusE assimilation</th>
</tr>
</thead>
<tbody>
<tr>
<td>chicken</td>
<td>CG</td>
<td>/ɪ/</td>
<td>/tʃɪkən/</td>
<td>/tʃɪkən/</td>
</tr>
<tr>
<td>present</td>
<td>CG</td>
<td>/e/</td>
<td>/prɛzənt/</td>
<td>/prɛzənt/</td>
</tr>
<tr>
<td>pizza</td>
<td>CG</td>
<td>/i/</td>
<td>/pˈɪztə/</td>
<td>/pɪztə/</td>
</tr>
<tr>
<td>picture</td>
<td>CG</td>
<td>/i/</td>
<td>/pɪktʃə/</td>
<td>/pɪktʃə/</td>
</tr>
<tr>
<td>money</td>
<td>CS</td>
<td>/ʌ/</td>
<td>/mʊnɪ/</td>
<td>/mʊnɪ/</td>
</tr>
<tr>
<td>lady</td>
<td>CS</td>
<td>/eɪ/</td>
<td>/l æɪdɨ/</td>
<td>/l æɪdɨ/</td>
</tr>
<tr>
<td>garbage</td>
<td>CS</td>
<td>/ɑr/</td>
<td>/g ɐːbəʤ/</td>
<td>/g ɒːbəʤ/</td>
</tr>
<tr>
<td>garden</td>
<td>CS</td>
<td>/ɑr/</td>
<td>/g ɐːdən/</td>
<td>/g ɒːdən/</td>
</tr>
</tbody>
</table>

Note. AusE = Australian English; TE = Tyneside English; CG = Category Goodness; CS = Category Shifting. Shading indicates paired target words. Characters in bold are phonemic rather than phonemic.

10.1.1.2.2 Carrier Sentences. Carrier sentences were also constructed from words available on the OZI. However, as the OZI does not contain articles, pronouns, or auxiliary verbs, the 100 most common English words (as based on the Oxford English Corpus)⁶ were also allowed in sentence construction. Care was taken to limit contextual information for the target word, by pairing each CG target word paired with a CS target word, and for each of these pairings, creating six nonsense sentences into which either target word could be placed. The sentences contained 8-10 syllables (\(M = 8.7\)), and were controlled so that a carrier sentence words did not differ in frequency from the target word by more or less than 20% (\(M = \lvert 6.7\rvert\%\)), and had an average frequency of 33.0% (\(SD = 8.7\%\)). As frequency information was not available for articles, pronouns, and auxiliary verbs,

⁶ http://oxforddictionaries.com/page/oec
those words were not included in sentence frequency calculations. Additionally, while some sentence words did contain CG assimilations, none contained any of the assimilations found in any of the target words. The target word was either located at the beginning of the sentence (with word onset at the second syllable), middle of the sentence, or as the final word of the sentence, an equal number of times. As well, for each of these two sentences, the target word was once preceded by a strong syllable, and once by a weak syllable. For a full list of carrier sentences used in the experiment, see Table 4.

Table 4

*Carrier sentences used in Experiment 3*

<table>
<thead>
<tr>
<th>Target Words</th>
<th>Carrier Sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td>chicken, garbage</td>
<td>Bad [target] was food for the bear</td>
</tr>
<tr>
<td></td>
<td>The [target] flew from the blue bush</td>
</tr>
<tr>
<td></td>
<td>Four dogs share [target] on a roof.</td>
</tr>
<tr>
<td></td>
<td>A broom could push the [target] on the porch.</td>
</tr>
<tr>
<td></td>
<td>The sad kangaroo could want [target].</td>
</tr>
<tr>
<td></td>
<td>A horse and lamb looked for the [target].</td>
</tr>
<tr>
<td>pizza, money</td>
<td>Good [target] was on the new shoes.</td>
</tr>
<tr>
<td></td>
<td>The [target] was on the chair for you.</td>
</tr>
<tr>
<td></td>
<td>A boy wants [target] for a dog.</td>
</tr>
<tr>
<td></td>
<td>A man had [target] and corn for Nanna.</td>
</tr>
<tr>
<td></td>
<td>Pour blue sauce on the bear's [target].</td>
</tr>
<tr>
<td></td>
<td>Juice splashed and flew on the [target].</td>
</tr>
<tr>
<td>present, lady</td>
<td>Bear's [target] stands on the new stairs.</td>
</tr>
<tr>
<td></td>
<td>The [target] was as long as a broom.</td>
</tr>
<tr>
<td></td>
<td>The man and blue [target] flew on a goose.</td>
</tr>
<tr>
<td></td>
<td>Nanna looked for the [target] from the zoo.</td>
</tr>
<tr>
<td></td>
<td>On the porch there was a good [target].</td>
</tr>
<tr>
<td></td>
<td>A dog ran and scratched a [target].</td>
</tr>
<tr>
<td>picture, garden</td>
<td>Four [target] rooms stand on the roof.</td>
</tr>
<tr>
<td></td>
<td>Their [target] looks good because of the bush.</td>
</tr>
<tr>
<td></td>
<td>You would have a new [target] of corn.</td>
</tr>
<tr>
<td></td>
<td>Thank you for the [target] and shorts.</td>
</tr>
<tr>
<td></td>
<td>The mad deer ran on the frog's [target].</td>
</tr>
<tr>
<td></td>
<td>Nanna splashed sauce on the [target].</td>
</tr>
</tbody>
</table>
10.1.1.2.3 Speakers. Target words were produced by a female native speaker of AusE from Western Sydney, Australia. The sentences were produced by two additional female native speakers of AusE, who were from Western Sydney and Northern Sydney, Australia, respectively. The sentences were also produced by a female speaker of TE, from Newcastle, England. All speakers had grown up in and currently resided in their hometowns, and were of similar age ($M = 24$ years, $3$ months; $SD = 3$ years, $11$ months) and were selected from a total of four recorded AusE and four TE speakers based on their desired pronunciation of the vowels used in our assimilation pairs.

It should be noted that varieties of AusE are not completely homogeneous. For instance, high, middle, and broad varieties are widely recognized. Increasingly, distinctions are being made between regional accents of Australian English as well (see Cox & Palethorpe, 2001). However, compared to the grosser variations found in British English or AmE varieties, the variations found in AusE are generally accepted to be quite minor, particularly in regard to suprasegmental variation. In this experiment, the accent used for the target words is middle Western Sydney AusE. For the native carrier sentences, the segmental information is also middle Western Sydney AusE, and the suprasegmental information is middle Northern Sydney AusE (see STRAIGHT Manipulations, below). However, for simplicity, the accent as a whole will be referred to as AusE.

10.1.1.2.4 Recordings. Recordings were made in a sound attenuated room. Participants were shown the target words and sentences on a monitor connected to an LG laptop. Opa 1.0 software, developed at the Marcs Institute (Antoniou et al., 2010), was used to present the target words and sentences in a random order. Opa
has the added benefit of allowing the administrator to make online quality 
judgments regarding the stimuli, such that if an utterance is not acceptable it is not 
counted, and the target item can be placed back into the random queue to be read 
again. Participants were instructed to say the words and sentences as if they were 
speaking to a toddler, and spoke into a Shure SM10A headworn microphone, using 
an Edirol UA-25 sound card that fed into CoolEdit 2000 (Syntrillium Software) on 
a PC laptop. Recordings were made into one channel, at 44,000 Hz. This channel 
was then doubled to create a stereo monophonic recording.

10.1.1.2.5 STRAIGHT Manipulations. Test sentences were created using 
STRAIGHT resynthesis of the selected natural speech recordings. STRAIGHT is a 
MATLAB program developed for high quality speech morphing (Kawahara et al., 
2008, 1999). STRAIGHT has been used successfully in speech research as a source 
of strikingly natural-sounding synthetic speech stimuli (Matsui & Kawahara, 2003; 
Yonezawa et al., 2005), and is particularly well-suited for applications involving 
manipulation of F0 features for auditory morphing. STRAIGHT is in essence a 
VOCODER, taking auditory input and converting it into three time-based maps: a 
spectrogram, and aperiodicity map, and an F0 map. These can be manipulated 
dependently, making it easy to morph from one sound to another based on these 
feature prior to resynthesis. TANDEM-STRAIGHT, used here, is a more recent 
implementation of STRAIGHT (Kawahara et al., 2008), that uses an improved, 
temporally-stable method of calculating the F0 trajectory.

To test children's perception of non-native segments without confounding 
non-native suprasegmental information, TE and AusE recordings of the same 
sentences were synthetically combined to create one sentence, composed of the TE
segmental information and the AusE suprasegmental information (Northern Sydney accent). Prior to manipulation, the amplitudes of sound files were each normalized in CoolEdit 2000. Sound files were then imported into STRAIGHT and converted to STRAIGHT Objects, which comprised spectrum, frequency, aperiodicity, F0, and time arrays. Corresponding segmental features in AusE and TE sentences were marked using Praat (Boersma & Weenink, 2008), and these served as temporal anchors prior to morphing. A continuous morphing space was then generated from the two stimuli. A morphing value of 1.0 corresponded to the original unaltered AusE token, while a morphing value of 0.0 corresponded to the original unaltered TE token. Values for the spectrum, frequency, and aperiodicity arrays were set to 0.0 (TE), while the time and F0 arrays were set to 1.0 (AusE). The target utterance was then synthesized into a morphed speech file. After manipulation, they were noise reduced using CoolEdit 2000, and scaled to 65 dB. AusE control sentence stimuli were created likewise, using the same AusE speaker from the TE-AusE stimuli to provide the suprasegmental information, but with segmental information coming from recordings of another AusE speaker (Western Sydney accent). Target word recordings were not manipulated using STRAIGHT, but were noise reduced and scaled to 65 dB.

10.1.2 Procedure. In this serial preference procedure, participants were seated on their caretaker’s lap in the middle of a room in front of a central monitor. A video camera positioned to the right of the monitor both recorded the session and provided a live stream to the experimenter in the adjacent control room. Each trial began with a blinking checkerboard on the monitor. Once the participant oriented to the screen for a minimum of 400 ms, the checkerboard stopped blinking, and
speech played (initiated by a keyboard press by the experimenter) from two
speakers centered behind the monitor. A new trial would start once the participant
turned away for at least 2000 ms, at which point the checkerboard would resume
blinking. If a participant turned away for less than 2000 ms, the same trial
continued once the participant reoriented to the monitor for 100 ms, with the time
spent looking away not counting towards the total looking time. A trial would also
time out once 1400 ms of looking time accumulated, at which point the
checkerboard would resume blinking for 1850 ms before the next trial commenced,
in order to maintain attention of the participant. For the duration of the study,
caretakers wore Macally MTUNE headphones that played a mixture of music and
speech, keeping them deaf to experimental conditions. Likewise, the experimenter
was kept deaf to the stimuli the participant heard.

10.1.3 Design and counterbalancing. All participants were familiarized to
the target words produced by the AusE speaker. Participants were familiarized to
two pairings of target words (four words total) until they accumulated 30 s of
familiarization (looking time) to each pair. Once children received 30 s of
familiarization time to one pair, it was removed from the stimuli set and no longer
played during familiarization, such that familiarization time for the two pairs was
equal. Half of the participants were familiarized to the pairs CHICKEN/PRESENT
and PIZZA/PICTURE, and the other half were familiarized to the pairs
LADY/MONEY and GARBAGE/GARDEN. Participants were then tested with two
blocks of four trials, counterbalanced across the native and non-native accent
conditions. Each trial consisted of the sentences for one of the four word pairings
(the two familiarized pairs, and the two unfamiliarized pairs). The sentences that
played for the duration of the trial (i.e., while the participant was oriented towards the monitor) were randomly selected without replacement, so that the child would not hear the same sentences twice within the same trial.

10.1.4 Results. Total looking time to trials containing familiarized and unfamiliarized words were compared in a 2 x 2 x 2 repeated-measures ANOVA on age group (12 months, 15 months), regional accent (native [AusE], non-native [TE]), and familiarity of target words in test sentences (familiarized, unfamiliarized), with age and regional accent condition as between subject factors, and familiarity as a within subject factor. A main effect of accent was found \( F(1, 41) = 6.423, p = .015, \eta^2_p = .135 \), reflecting that participants in the non-native accent condition looked longer during the task compared to participants in the native accent condition. A trend regarding a main effect of age group was found \( F(1, 41) = 2.963, p = .093 \), reflecting a tendency for 12-month-olds to look longer during the task compared to 15-month-olds. No other main effects were found. A trend towards an interaction between age and familiarity was found, \( F(1, 41) = 3.426, p = .071 \), reflecting a tendency for 12-month-olds to look more at sentences containing unfamiliarized words, and 15-month-olds to look more at sentences containing familiarized words across accents. A trend towards an interaction was also found between accent and familiarity, \( F(1, 41) = 3.707, p = .061 \), reflecting a tendency for participants to look more to sentences containing familiarized words in the native regional accent, and look more to sentences containing unfamiliarized words in the non-native regional accent overall. No other interactions were found.
An additional 2 x 2 x 2 repeated-measures ANOVA comparing age group, familiarity, and assimilation type (CG, CS) was performed in the non-native regional accent. The native accent was not included in this analysis as CG and CS variations are in relation to native accent productions, and so these variations cannot exist in the native regional accent. A trend towards a main effect of assimilation type was found, $F(1, 19) = 4.160, p = .056$, reflecting that sentences containing CG variations tended to be looked at longer than sentences containing CS variation overall. An interaction between age group and assimilation type was found, $F(1, 19) = 7.245, p = .014, \eta^2_p = .276$, in that 15-month-olds looked longer to sentences containing CG variations than CS variations, while 12-month-olds show a slight preference in the opposite direction. Further, a trend towards an interaction was found for age group and familiarity, $F(1, 19) = 4.003, p = .060$, reflecting a tendency for 15-month-olds, but not 12-month-olds, to look more to sentences containing familiarized words versus unfamiliarized words in the non-native regional accent. A three-way interaction between age group, familiarity, and assimilation type was also found, $F(1, 19) = 4.765, p = .042, \eta^2_p = .201$. Fifteen-month-olds looked more to CG contrasts than to CS contrasts, and looked more to sentences containing familiarized vs. unfamiliarized words, whereas 12-month-olds did not show these preferences. Together, these results suggest that 12- and 15-month-olds treat the segmentally varied stimuli differently.

To specifically clarify whether 12- and 15-month-olds discriminated sentences containing familiarized words from sentences containing unfamiliarized words across the regional accents, total looking times for sentences containing familiarized versus unfamiliarized words for each age group and accent condition
were compared in paired-samples $t$ tests. For the 12-month-olds, no difference
between looking to sentences with familiar words versus unfamiliar words was
found in the native or non-native regional accent conditions. For the 15-month-olds,
no discrimination was found in the native regional accent condition, but a
preference was found for sentences containing familiarized words in the non-native
regional accent condition, $t(10) = 2.916, p = .015, 95\%$ CI (2256.964, 16879.582;
Figures 9 and 10). To clarify the interaction between age group, familiarity, and
assimilation type found for participants in the non-native regional accent condition,
separate paired-samples $t$ tests were conducted to examine participants' performance
in the non-native regional accent across words containing CG and CS variations.
Again, no differences were observed in these analyses for the 12-month-olds, but
15-month-olds demonstrated a familiarity preference for sentences containing CS
variations, $t(10) = 3.233, p = .009, 95\%$ CI (2373.366, 12901.907), but not for
CG variations (Figures 11 and 12).
Figure 9. Total looking time to sentences containing familiarized versus unfamiliarized words in the native segments condition (AusE) and non-native segments condition (TE), which contained vocalic variation compared to AusE. Twelve-month-olds did not show discrimination in either regional accent. Fifteen-month-olds did not show discrimination in the native AusE condition, but did show increased looking to sentences containing familiarized words versus sentences containing unfamiliarized words in the non-native segments condition. Error bars represent standard error.
Figure 10. Difference in looking time to sentences containing familiarized versus unfamiliarized words in the native segments condition (AusE) and non-native segments condition (TE), which contained vocalic variation compared to AusE. Twelve-month-olds did not show a difference in looking time between sentences containing familiarized and unfamiliarized words in either condition. Fifteen-month-olds did not show a difference in looking time in the native segments condition, but did show a difference in the non-native segments condition via increased looking to sentences containing familiarized words versus sentences containing unfamiliarized words. Error bars represent standard error.
Figure 11. Total looking time in the non-native segments condition (TE) across sentences containing familiarized versus unfamiliarized words that contain either a Category Goodness or Category Shifting vocalic variation compared to native segments. Twelve-month-olds did not show discrimination in either condition, but 15-month-olds demonstrated a familiarity preference for sentences containing familiarized words when target words contained a Category Shifting vowel variation. Error bars represent standard error.
As predicted, 15-month-olds’ performance in the task differed from that of the 12-month-olds’. Specifically, 15-month-olds showed discrimination of sentences containing familiarized words from sentences containing unfamiliarized words in the non-native segments condition, whereas 12-month-olds showed no such discrimination. Further, discrimination was specifically to target words containing CS variations to vowels. This sensitivity to CS variations to vowels is consistent with the notion that phonological specificity for vowels occurs at around 15 months.

Surprisingly, 12- and 15-month-olds both failed to show discrimination in the native regional accent condition. This was unexpected, given that 12-month-olds have shown discrimination of familiarized stimuli in similar tasks involving half as
many stimuli as were presented here (e.g., Schmale et al., 2010). Doubling the stimuli with respect to previous studies served two purposes. The first was to test the robustness of children's ability to segment words from running speech at that age, both with and without segmental variation. The results suggest that while children are able to segment words from continuous speech in a cognitively simplified context, this ability is not robust. The second reason for increasing the stimulus set here was in order to make the task appropriate for testing older children. To our knowledge, this task has not been performed with children older than 13 months (Schmale & Seidl, 2009), and thus the increase in stimuli was employed to keep the task adequately stimulating for older participants. That 12-month-olds failed to show discrimination across conditions suggests that the increased stimulus load may have made the task too cognitively demanding, and failure to show discrimination in the native regional accent corresponds with 12-month-olds' failure to discriminate in the native regional accent under an almost identical stimulus load (Chapter 9). However, the fact that 15-month-olds did show discrimination in the non-native regional accent, which is purported to be the more challenging condition, suggests that their non-discrimination in the native regional accent was perhaps due to a lower level of interest in the native-accent condition, rather than an inability to handle cognitive demands imposed by the task.

10.2 Experiment 4

In this experiment, we examined children's ability to segment words across a synthesized regional accent that contained marked variation in its consonants compared to the native regional accent, but within the context of native
suprasegmental information. Southeast London English (seLE), spoken in London, England was used as it contained marked consonant but not vowel differences from the participants' native AusE.

Using the same procedure as in Experiment 3, we included the same age groups from Experiment 3, 12- and 15-month-olds, and included 19-month-olds as well. While the results from Experiment 3 suggest that word forms at 15 months contain phonological specification for vowels, we expect that fully specified phonological word forms that include phonologically specified consonants as well, do not develop until around 19 months, as evidenced by children's ability to identify regionally accented words containing segmental variation to both consonants and vowels, at 19 months, but not at 15 months (Chapter 8). Thus we predicted a difference in performance between 19-month-olds compared with 12- and 15-month-olds. Specifically, we predicted that 19-month-olds would show a greater distinction between familiar and unfamiliar words in the non-native regional accent when compared to the 12- and 15-month-olds, and that only 19-month-olds would show a significant difference between familiar and unfamiliar words in this condition.

10.2.1 Method.

10.2.1.1 Participants. Participants were 25 11.6- to 12.6-month-olds ($M = 12$ months, 2 days; $SD = 9$ days; 9 females; 12 in native condition), 21 14-7- to 15.7-month-olds ($M = 15$ months, 3 days; $SD = 9$ days; 8 females; 10 in native condition), and 21 18.5- to 19.6-month-olds ($M = 19$ months, 3 days; $SD = 10$ days; 9 females; 11 in native condition), from AusE-speaking households in Sydney, Australia, whose amount of exposure to non-native languages or non-
AusE accents ranged from 0 to 2.5 hours per week ($M = 0.2$ hrs/week). Data from an additional 27 toddlers were collected but not included in the analysis due to fussiness ($n = 26$) or parental interference ($n = 1$).

10.2.1.2 Stimuli and Materials.

10.2.1.2.1 Target words. Target word selection took the same approach as in Experiment 3, with the exception that target words were selected according to their expected consonant assimilation patterns between the two regional accents (CG or CS). As in Experiment 3, three unique CG and three unique CS vowels used, and so one CG and CS vowel was repeated across the eight target words. The target words were divided into two groups of four based on whether the assimilation consonant they contained was CG or CS. Word frequency across the groups were controlled, so that across both groups the average frequency (based on a 15-month-old's expressive vocabulary) was 30.8% ($SD = 10.8\%$). A summary of the target words, and their assimilations can be found in Table 5.
Table 5

Target words for Experiment 4, their assimilation types, target consonants, AusE and seLE broad transcriptions, and AusE assimilations of target consonants for CS only (CG variants are perceived as categorically the same word as in the native pronunciation).

<table>
<thead>
<tr>
<th>Target Word</th>
<th>Assimilation type</th>
<th>Target phoneme</th>
<th>Transcription</th>
<th>AusE</th>
<th>seLE</th>
<th>AusE assimilation</th>
</tr>
</thead>
<tbody>
<tr>
<td>tiger</td>
<td>CG</td>
<td>initial /t/</td>
<td>/tʰaɪɡə/</td>
<td>/tʰaɪɡə/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>raisin</td>
<td>CG</td>
<td>initial, medial /t/</td>
<td>/ɹaɪzən/ or /واسزان/</td>
<td>/ɹʷaɪzən/ or /واسزان/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>orange</td>
<td>CG</td>
<td>initial, medial /r/</td>
<td>/ɒɹənʤ/ or /اورونج/</td>
<td>/ɒ ɹʷənʤ/ or /اورونج/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tractor</td>
<td>CG</td>
<td>initial /tr/</td>
<td>/tʰɹæktə/</td>
<td>/tʰʋæktə/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>bottom</td>
<td>CS</td>
<td>medial /t/</td>
<td>/bɒɾəm/ or /باشروم/</td>
<td>/bɒʔəm/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>brother</td>
<td>CS</td>
<td>medial /ð/</td>
<td>/brʌðə/ or /براشر/</td>
<td>/brʌ və/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>photo</td>
<td>CS</td>
<td>medial /t/</td>
<td>/fouroʊ/ or /فووروح/</td>
<td>/fouʔou/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>hammer</td>
<td>CS</td>
<td>initial /h/</td>
<td>/hæmə/ or /هاشر/</td>
<td>/æmə/</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. AusE = Australian English; seLE = southeast London English; CG = Category Goodness; CS = Category Shifting. Shading indicates paired target words. Characters in bold are phonetic rather than phonemic.

10.2.1.2.2 Carrier sentences. Carrier sentences were constructed in a similar manner to Experiment 3. The sentences contained 8-10 syllables (\(M = 8.8\)), and were controlled so that a carrier sentence did not differ in frequency from its target word by more or less than 20% (\(M = |8.5|\%\)), and had an average frequency of 31.3\%, (\(SD = 7.3\%\)). For a full list of carrier sentences used in the experiment, see Table 6.
Table 6

Carrier sentences used in Experiment 4

<table>
<thead>
<tr>
<th>Target Words</th>
<th>Carrier Sentences</th>
</tr>
</thead>
</table>
| tiger, brother | Pop's [target] sat on some white plates.  
My [target] was lost on a boat.  
You've got one [target] and a wet cat.  
Someone put a [target] on a cot.  
Please get me a cute new [target].  
Don't put Vegemite on my [target]. |
| orange, photo | Five [target] shops were on my street.  
My [target] was put on my plate.  
You can get a nice [target] at some shop.  
I got my [target] from a policeman.  
Wipe Vegemite off Pop's [target].  
You got wet soup on my [target]. |
| tractor, hammer | No [target] can do what my friend's does.  
My [target] was broken by a boy's bat.  
What a nice [target] my nurse gave me.  
I can use a [target] or a boat.  
I will get you just one [target].  
A policeman does not want my [target]. |
| raisin, bottom | One [target] sat on Nanna’s street.  
My [target] was put on a cot.  
Soup splashed and got Pop's [target] so wet.  
On my [target] was a cute butterfly.  
Come and look at my cat’s white [target].  
At lunch I sat on my [target]. |

10.2.1.2.3 Speakers. The AusE target words and AusE control sentences were produced by the same speakers as in Experiment 3. For the test sentences, the same AusE speaker as in Experiment 3 was used to provide the suprasegmental information. This was paired with recordings by a female speaker of seLE, from London, England. All speakers had grown up in and currently resided in their hometowns, and were of similar age ($M = 22$ years, 8 months, $SD = 2$ years, 11 months) and were selected from a total of four AusE and five seLE speakers based on their desired pronunciation of the consonants used in our assimilation pairs.
10.2.1.2.4 Recordings. Recordings were made as in Experiment 3.

10.2.1.2.5 STRAIGHT Manipulations. STRAIGHT Manipulations followed the same procedure as in Experiment 3, with the same AusE speaker from the previous experiment providing the suprasegmental (F0, timing) information (which is also the same speaker providing the suprasegmental information in the AusE control stimuli), and the seLE speaker providing the segmental information (spectrum, frequency, aperiodicity).

10.2.2 Procedure. As in Experiment 3.

10.2.3 Design and counterbalancing. As in Experiment 3, participants were familiarized to the target words produced by the AusE speaker. Participants were familiarized to two pairings (four words total) of target words until they accumulated 30 s of familiarization (looking time) to each pair. Once children received 30 s of familiarization time to one pair, it was removed from the stimuli set and no longer played during familiarization, such that familiarization time for the two pairs was equal. Half of the participants were familiarized to the pairs TIGER/RAISIN and PHOTO/HAMMER, and the other half were familiarized to pairs ORANGE/TRACTOR and BOTTOM/BROTHER. Participants were then tested with two blocks of four trials presented either in AusE for half of the participants, or in seLE for the other half of participants. Each trial consisted of the sentences for one of the four word pairings (the two familiarized pairs, and the two unfamiliarized pairs), such that each four-trial test block included all four trial types. The sentences that played for the duration of the trial (i.e., while the participant was oriented towards the monitor) were randomly selected without
replacement, so that the child would not hear the same sentences twice within the
same trial.

10.2.4 Results. Total looking time to trials containing familiarized and
unfamiliarized words were compared in a 3 x 2 x 2 ANOVA comparing age group
(12, 15, and 19 months), regional accent (native [AusE], non-native [seLE]), and
familiarity of target words in the test sentences (familiarized, unfamiliarized), with
age group and regional accent as between-subject factors, and familiarity as a
within-subjects factor. Planned comparisons were made for effects of accent,
familiarity, and age group. Nineteen-month-olds were compared to 12- and 15-
month-olds, and to 15-month-olds alone, as the 19-month-olds, having fully
specified phonological word forms, were expected to respond differently to the non-
native regional accent stimuli containing variation to consonants, compared to the
other age groups that have not yet developed fully specified phonological word
forms. Interactions between age group, regional accent, and familiarity were also
tested. However, no effects were found, though there was a trend toward a main
effect of familiarity, \( F(1, 61) = 3.214, p = .078 \). Similar to Experiment 3,
participants in the non-native accent condition tended to look longer during the task
compared to participants in the native accent condition.

An additional set of comparisons were made examining performance in the
non-native regional accent condition, comparing familiarity for CG and CS across
age groups. The native accent was not included in this analysis as CG and CS
variations are in relation to native accent productions, and so these variations
cannot exist in the native regional accent. Planned orthogonal comparisons were
made for effects of age group, assimilation type (CG, CS), and familiarity. As
before, 19-month-olds were compared to 15- and 12-month-olds, and to 15-month-olds alone. No main effects were observed. However, a three-way interaction was observed between age group (19-month-olds versus 12- and 15-month-olds), assimilation type, and familiarity, $F(1, 31) = 4.223, p = .048, 95\% \text{ CI} (0.004, 0.997)$, but not between age group (19-month-olds versus 15-month-olds), assimilation type, and familiarity. No other interactions were significant. These results are consistent with the interpretation that 19-month-olds showed a larger difference between familiarized and unfamiliarized words in the CS condition than in the CG condition, compared to the younger participants.

To clarify whether 12-, 15-, and 19-month-olds discriminated sentences containing familiarized words from sentences containing unfamiliarized words across the regional accents, we separately analyzed total looking times for sentences containing familiarized versus unfamiliarized in paired-samples $t$ tests. For the 12-month-olds, a familiarity preference was found in the native regional accent condition, $t(11) = 2.555, p = .027, 95\% \text{ CI} (1240.480, 16670.687)$, but no effect was observed in the non-native regional accent condition. For the 15-month-olds, no effect was found in either the native or the non-native regional accent condition. For the 19-month-olds, discrimination also was not found in the native regional accent or the non-native regional accent (Figures 13 and 14).

Further paired-samples $t$ tests in the non-native regional accent condition were carried out to examine participants' performance in the non-native regional accent across words containing CG and CS variations. Twelve-month-olds did not discriminate sentences containing familiarized versus unfamiliarized words containing either CG or CS variations (Figures 15 and 16). Fifteen-month-olds also
failed to show such discrimination. However, while 19-month-olds did not show a familiarity preference for sentences containing CG variation, they did show a familiarity preference for sentences containing familiarized words that contain CS consonants variations, \( t(9) = 2.972, p = .016, 95\% \text{ CI} (1009.157, 7440.843; \) Figures 15 and 16).

![Figure 13. Total looking time to sentences containing familiarized versus unfamiliarized words in the native segments condition (AusE) and non-native segments condition (seLE), which contained variations to consonants compared to AusE. Only 12-month-olds showed any discrimination via increased looking to sentences containing familiarized words versus sentences containing unfamiliarized words in the native segments condition. Error bars represent standard error.](image-url)
Figure 14. Difference in looking time to sentences containing familiarized versus unfamiliarized words in the native segments condition (AusE) and non-native segments condition (seLE), which contained variations to consonants compared to AusE. Only 12-month-olds showed any discrimination via increased looking to sentences containing familiarized words versus sentences containing unfamiliarized words, in the native segments condition. Error bars represent standard error.
Figure 15. Total looking time in the non-native segments condition (seLE) across sentences containing familiarized versus unfamiliarized words that contain either a Category Goodness or Category Shifting consonant variation compared to native segments. Neither 12- nor 15-month-olds showed discrimination in either condition. Nineteen-month-olds did not show discrimination when target words contained a Category Goodness variation, but did show a familiarity preference for sentences containing familiarized words when target words contained a Category Shifting variation. Error bars represent standard error.
10.2.5 Discussion. As predicted, 19-month-olds treated the stimuli differently compared to 12- and 15-month-olds combined, showing a greater discrimination of familiarized versus unfamiliarized stimuli in the non-native regional accent containing consonant variation. Further, 19-month-olds were the only age group to show discrimination in the non-native regional accent condition, demonstrating a preference for sentences containing familiarized words having CS variation to consonants, but not for CG variation. The 19-month-olds' performance follows an almost identical pattern to that of the 15-month-olds in Experiment 3 for vowel variation, in which 15-month-olds showed a preference for words containing familiarized words having CS variation, but not CG variation. The only difference between the two findings is that the 15-month-olds also showed discrimination in

Figure 16. Difference in looking time in the non-native segments condition (seLE) across sentences containing familiarized versus unfamiliarized words that contain either a CG or CS consonant variation compared to native segments. Neither 12- nor 15-month-olds showed a difference in looking time in either condition, nor did 19-month-olds show a difference in looking time when words contained a CG variation. However, 19-month-olds did show a difference in looking time when words contained a CS consonant variation, looking longer to sentences containing familiarized words. Error bars represent standard error.
the non-native regional accent condition overall. Still, this suggests a developmental progression of phonological specificity to words in which specificity for vowels occurs by 15 months, but specification for consonants occurs later, by 19 months.

In regard to the native accent condition, only 12-month-olds discriminated sentences containing familiarized versus unfamiliarized words. This was surprising in the light of findings from Experiment 3 in which 12-month-olds failed to show discrimination. This challenges the conclusion brought forth in Experiment 3, that the number of familiarization stimuli overcame 12-month-olds' ability to perform the task. However, this demonstration that 12-month-olds are able to perform the task provides further suggestive support for the argument that 15-month-olds' non-discrimination in the native regional accent in Experiment 3 was a result of the task being insufficiently stimulating, rather than due to an inability to perform it.

10.3 General Discussion

10.3.1 Summary of results. Experiment 3 compared 12- and 15-month-olds' ability to segment familiarized words from sentences in their native regional accent (AusE), as well as in a non-native regional accent (TE) containing CG and CS variations to vocalic segments, with native suprasegmental information superimposed. This was measured via their discrimination of sentences containing familiarized words from sentences containing unfamiliarized words in a preference test. Performance between 12- and 15-month-olds differed, with 12-month-olds failing to discriminate sentences containing familiarized words versus sentences containing unfamiliarized words in either the native regional accent (AusE), or the non-native regional accent (TE). Fifteen-month-olds mirrored the 12-month-olds in
their failure to show discrimination in the native regional accent, but did
discriminate sentences containing familiarized words versus unfamiliarized words
in the non-native regional accent. Further analysis revealed that this discrimination
was driven by a familiarity preference when the target words contained CS vocalic
differences, but not CG vocalic differences.

Experiment 4 tested 12-, 15-, and 19-month-olds' ability to segment
sentences in their native regional accent, and in a non-native regional accent (seLE)
that contained CG or CS variations to consonants. Nineteen-month-olds'
performance differed from the combined performance of 12- and 15-month-olds.
Similar to the 15-month-olds' performance in Experiment 3, 19-month-olds
demonstrated a familiarity preference for target words containing CS consonant
differences, but not CG consonant variations. In the native regional accent, only 12-
month-olds showed a familiarity preference for sentences containing familiarized
words in the native regional accent, but not in the non-native regional accent.

10.3.2 12-month-olds. Twelve-month-olds discriminated sentences
containing familiarized words from sentences containing unfamiliarized words in
the native regional accent in Experiment 4, but not in Experiment 3. Previous
studies implementing this procedure have demonstrated that children as young as
7.5 months (Jusczyk & Aslin, 1995), and as old as 13 months (Schmale & Seidl,
2009) demonstrate such discrimination in their native regional accent. As there are
no substantive differences between the stimuli used in Experiments 3 and 4, it is
difficult to explain the disparity in experimental findings. However, in Experiment
2, which used a similar stimulus set, 12-month-olds also failed to show
discrimination in the native accent condition. Thus, it may be simply that the more
difficult task (compared to previous studies) used here has resulted in an effect that may not be consistently observable, or alternatively, the more restricted set of vowels used in the TE stimuli in Experiment 3, resulted in a greater task difficulty. It's possible that at 12 months vowel differences are captured by differing prosodic features, and thus the restricted vowel differences within the native prosodic structure may have been relatively uninformative. This links with Experiment 2 where it was proposed that non-informative prosodic information interfered with segmentation. However, this interpretation would need to be tested in a separate series of experiments.

10.3.3 15- and 19-month-olds. While neither the 15- nor the 19-month-olds demonstrated discrimination in the native regional accent, 15-month-olds showed a sensitivity to non-native accented CS variations to vowels, and 19-month-olds showed a sensitivity to non-native accented CS variations to consonants. That the 15- and 19-month-olds showed discrimination in the more challenging condition indicates that their failure to show discrimination in the native regional accent cannot have been due to it being too cognitively challenging. This is even more apparent by the fact that the younger 12-month-olds discriminated in the native regional accent in Experiment 4. To our knowledge, the preference procedure as implemented here in a word segmentation context has not been used on children as old as 15 and 19 months, and is instead generally used on younger children, ranging in age from 7.5-13 months. Thus our concern was that the task might not be adequately engaging for our older participants, which led us to double the target stimulus set with respect to similar implementations of the procedure on younger
children. However, it would appear that even with the increased stimulus set, the task was still not sufficiently interesting or engaging for our older participants.

The results from the non-native regional accent, however, demonstrate that 15-month-olds have a grasp of phonological constancy for vowels, and only 19-month-olds have a grasp of phonological constancy for consonants. While they are sensitive to segmental changes, they also recognize that despite the variation, the variant pronunciations are still those of familiarized words. Further, 15- and 19-month-olds' non-discrimination in the native regional accent, but discrimination in the non-native regional accent for CS variations suggests that their attention was drawn specifically to CS variations (to vowels and consonants, respectively). This attentional bias may signal a change in perceptual attention, reflective of the children's emerging phonological knowledge.

The developmental disparity between 15- and 19-month-olds suggests that the development of phonological constancy has a different progression for consonants and vowels, with phonological constancy for vowels emerging sooner than that for consonants. This proposal may clarify previous findings on children's ability to recognize words across regional accents. Mulak et al. (under review) found that 19-month-olds could identify words in both their native and a non-native regional accent that simultaneously contained consonant and vowel variation, but 15-month-olds could do so only in their native regional accent. If 15-month-olds have phonological constancy only for vowels, then it may be this that prevented their getting past the consonant variation also included in JaME, while 19-month-olds having phonological constancy for both vowels and consonants, were able to handle both forms of variation.
In regard to assimilation types, we can say definitively that 15-month-olds were able to segment words having vocalic CS differences, and 19-month-olds could do so for words having CS variations to consonants. Based on the assumptions of PAM (Best, 1994b, 1995), CG variations to a phoneme assimilate to the same category, while CS variations assimilate to a different category. In this context, the CG variants should be more easily heard as native-like than the CS variants. This pattern was observed here, in that CG variations were not discriminated, just as the native regional accent was not discriminated, while CS variations resulted in a change in perceptual pattern.

Overall, the results from this study support the general view that phonologically specified word forms appear to be in place by 19 months, but are emerging at 15 months, with the younger children sensitive to novel variation in vowels, and the older children also being sensitive to novel variation in consonants, having already mastered vocalic variation. Given that the methods employed here may have been too simple to hold the attention of 15- and 19-month-old children, it may be advisable to use a more engaging method such as the eye tracking task used in Mulak et al. (in preparation; Chapter 8) to test this possible interpretation further. Based on these findings, stimuli should be chosen with phonologically constant variations split in the same manner as here (CG vs. CS, vowels vs. consonants, with native suprasegmental information), and it is expected that 15- and 19-month-olds will handle CG variations to both vowels and consonants when recognizing known words, but 15-month-olds will struggle with CS consonant variations to the pronunciations of known words.
11 General Discussion

11.1 Summary of Findings

Experiment 1 asked when children develop phonologically specified word forms, via their ability to identify words containing cross-regional accent variation. The experiment examined 15- and 19-month-olds' ability to identify words in Jamaican Mesolect English, a non-native regional accent that contains variation in consonants, vowels, and prosody in comparison with the native regional accent. In a visual choice eye tracking task, two images were presented on-screen, and recordings in either their native or the non-native regional accent asked children to look at one of the named pictures. While both 15- and 19-month-olds looked more to the named than unnamed pictures in their native regional accent, only 19-month-olds did so also in the non-native regional accent, indicating phonological constancy has emerged by 19, but not 15 months. Expressive vocabulary positively predicted
15-month-olds' performance in the non-native regional accent, suggesting phonological development may be closely linked with vocabulary size.

Experiments 2-4 looked at the emergence of phonologically specified word forms and their effect on recognition of words in continuous speech, and in particular, segmentation of words spoken with a non-native accent. Prosodic, vocalic, and consonantal variation were examined separately in order to determine whether components of word segmentation and phonological specificity have varying developmental progressions. In each experiment, participants were familiarized to four words in AusE and then tested on their ability in a serial looking preference task to discriminate sentences containing the familiarized words from those containing unfamiliarized words. The test sentences were presented either in a synthesized accent containing native (AusE) segments and suprasegmental information, or in a synthesized accent, in which suprasegmental information from a foreign accent was synthesized onto native segmental information (Experiment 2), or in accents in which segmental information from an unfamiliar regional accent was synthesized with native suprasegmental information (Experiments 3 and 4).

Experiment 2 examined 12-month-olds' ability to recognize words from continuous speech containing only suprasegmental (or prosodic) variation from their native accent, by synthesizing suprasegmental features from French-accented English onto AusE segmental features. In the non-native suprasegmental features condition, the 12-month-olds tended to look longer to sentences containing unfamiliarized words, but did not show discrimination in the native suprasegmental features condition. It was proposed that their tendency to discriminate sentences in
the non-native suprasegmental features condition, but not in the native accent may have been due to the cognitive demand of the experiment resulting in participants falling back on the MSS (Cutler & Norris, 1988) for discovering word boundaries (see Section 9.3). In the native regional accent, attention to the metrical boundaries may not have provided enough detail as to the identity of words, as attention was perhaps drawn away from the segmental features, and thus there were was no discrimination. In the non-native suprasegmental information condition, the unusual metrical cues perhaps could not be interpreted, and thus infants ignored them, returning focus to the (native) segmental cues, and thus supporting discrimination of novel from familiar words.

Experiment 3 explored children's ability to recognize words from continuous speech containing only vocalic variation from the AusE. Words in Tyneside English (TE) were selected such that they only contained CG or CS variations in vowels compared to AusE realizations. Native AusE suprasegmental cues were then synthesized onto the TE segments, and performance in this accent was compared to the native segments condition, in which segments from one speaker of AusE were synthesized with the suprasegmental features of another AusE speaker. Performance between the 12- and 15-month-olds differed in that only 15-month-olds showed a familiarity preference in the non-native vowels condition, and specifically only for the CS vowel condition of the non-native regional accent. This reflects their ability to segment words from running speech across variation to vowels, and their emerging phonological knowledge for vowels. Neither group showed discrimination in the native accent condition.
Experiment 4 looked at children's ability to recognize words from continuous speech containing only CG or CS variation to consonants in another non-native regional accent. Words in southeast London English (seLE) were synthesized with native AusE suprasegmental features. Nineteen-month-olds' performance differed from that of 12- and 15-month-olds. Only 19-month-olds showed a familiarity preference for the non-native consonants condition, and specifically only for the CS consonants condition, suggesting emerging phonological knowledge for consonants at that age, and mastery of phonological knowledge for vowels. In the native accent condition, only 12-month-olds showed a familiarity preference for sentences containing familiarized words versus sentences containing unfamiliarized words.

11.2 Task Differences

Comparing among the results from Experiments 1-4, as well as comparisons to relevant findings from the literature, we see several notable patterns of results. In Experiment 1, 15- and 19-month-olds were able to perform the task in the native regional accent, whereas this was not the case for Experiments 3 (vowel variation) or 4 (consonant variation), which also tested children at those ages. Similarly, previous studies implementing a similar listening preference task as in Experiments 2-4 have found children as young as 7.5 months are able to perform the task in their native regional accent (Jusczyk & Aslin, 1995), whereas that was not found in any of our experiments implementing a listening preference task, apart from the 12-month-olds in Experiment 4. Finally, in Experiment 2, the trend towards discrimination by the 12-month-olds was observed as a novelty preference, with
children tending to look more to sentences containing unfamiliarized words versus those containing familiarized words. In Experiments 3 and 4, where discrimination by this age group was found for native speech, it was in the form of a familiarity preference. The sections below discuss differences between the two tasks used in this thesis (visual choice and serial listening preference), and also discuss differences between the listening preference task and previous approaches, in order to address possible reasons behind these disparate findings.

11.2.1 Visual choice task versus listening preference task. Experiment 1 implemented a visual choice eye tracking paradigm. In this task, children are presented with two visual objects on a screen – a target image, and a distractor image. The auditory token for the object corresponding to the target image is then played, and during this, children's eye gaze to each of the items is recorded. Recognition of the item is inferred from a greater proportion of gaze directed at the target image, compared to the distractor image.

Experiments 2 through 4 implemented a serial listening preference procedure. When used to test word segmentation, this task first involves a familiarization phase, followed by a test phase. In the familiarization phase, participants are exposed to repetitions of target words for a specified duration of time. Once familiarization has elapsed, children are played alternating or randomized trials of sentences that either contain the familiarized words or do not, with trial length contingent upon the duration that the participant fixes on a central screen.

In Experiment 1, which implemented the visual choice task, both 15- and 19-month-olds were able to perform the task in the native regional accent, as would
be expected, whereas this was not seen in any of the subsequent experiments which implemented the listening preference task. The visual choice task presents a visual depiction of the named word, and in this case further context was provided by placing target words at the end of carrier sentences, and providing task reinforcement via animation of the target image and a reinforcing sentence at the end of each trial. The listening preference task, by contrast, does not include any context cues that could improve recognition of the word – there are no images, and sentences were constructed to ensure that they did not contain any semantic information that could aid recognition of the word. As well, the task does not offer any reinforcement. Thus at first glance, it may appear that 15- and 19-month-olds’ discrimination in the native regional accent in Experiment 1, but not in Experiments 3 and 4 is likely due to differences in available context between the two tasks. However, in Experiments 3 and 4, 15- and 19-month-olds were respectively able to perform the task in the non-native segment conditions, which are expected to be more challenging than the native segment conditions. To our knowledge, the listening preference task has only been used to test word segmentation in children as old as 13 months (Schmale & Seidl, 2009). While efforts were made to make the task suitable for older children by increasing the number of familiarization and test stimuli, it appears that the task, at least in the native conditions, was not sufficiently engaging for our older participants, and thus the more likely explanation for their nondiscrimination in the native segments conditions was due to relative inattention. If this is the case, then when 15- and 19-month-old listeners in this task do show discrimination, this may be interpreted as their attention having been drawn to the variant pronunciations.
11.2.2 Differences in task demands in the listening preference task

compared to the literature. The results obtained here differ not only across tasks, but also within listening preference tasks as well, when compared to the existing literature. The serial preference procedure used here is adapted from the Headturn Preference Procedure (Jusczyk & Aslin, 1995), where left and right situated monitors (or lights, in its original implementation) accompany a single central monitor. To begin each test trial, the central monitor blinks, and once head position is centered, a peripheral monitor blinks to elicit a headturn to that monitor, at which point blinking stops and sentences containing either familiarized or unfamiliarized words play in an alternating or randomized trial order, until the participant turns away from the peripheral monitor. This is in contrast to the serial preference procedure, in which there is only a centrally positioned monitor that blinks to elicit looking to the monitor, at which point blinking stops and the stimuli play either in an alternating or randomized trial order until the participant turns away from the screen. Findings from both types of listening preference procedure are often compared, and it is not believed that the slight variations between methods leads to differences in the interpretability of results. It has been found that children as young as 7.5 months show a familiarity preference with the head-turn preference procedure in their native regional accent (Jusczyk & Aslin, 1995). Similarly, studies looking at children's ability to recognize words in continuous speech across accents have found that 12-month-olds show a familiarity preference when familiarized to words either in their native regional accent, or a non-native regional accent, and subsequently tested on passages in the alternate accent (Schmale et al., 2010).
Similar results were found for 13-month-olds when tested in the same procedure with a foreign accent, rather than a regional accent (Schmale & Seidl, 2009).

However, as seen in Experiment 2, 12-month-olds did not show reliable discrimination in the native suprasegmental features (with native segments) condition. Neither did 12-month-olds in the non-native segments (with native suprasegmental features) conditions in Experiments 3 and 4. What can account then for the discrepant performance between the participants in the current thesis and the results of previous studies? Aside from differences in participant criteria compared to previous studies (see Chapter 9), the current studies familiarized children to twice as many stimuli (four words) compared to previous studies (two words), and tested them on eight words as compared with four words in previous studies. As well, children received half of the amount of familiarization time per word. Thus the method used here may be considered substantially more cognitively demanding than previous experiments.

Performance varied across the studies as well. Twelve-month-olds did not show discrimination in the native accent condition in Experiment 3, but did in Experiment 4. As there are no substantive differences between the stimuli used across those experiments, it is difficult to account for the disparity. It may simply be that the more difficult task used here relative to previous studies has resulted in an effect that may not be consistently observable.

Twelve-month-olds also failed to show discrimination in the native regional accent in Experiment 2. While the stimuli for Experiments 3 and 4 were controlled in the exact same manner, the stimuli used in Experiment 2 were controlled in a different way. In Experiments 3 and 4, stimuli were arranged so that each
participant heard a given carrier sentence up to four times within an experiment; twice containing a familiarized word, and twice containing an unfamiliarized word. However, due to differing selection criteria of target words, Experiment 2 by necessity was arranged so that each participant only heard a carrier sentence a maximum of two times, with the same target word. Hearing the sentences twice as often in Experiments 3 and 4 compared to Experiment 2 may have given children more opportunity to correctly segment the sentence and learn where the target word is located. Thus while the increased stimulus load relative to previous studies may have increased the cognitive demand of Experiments 2-4, it should also be kept in mind that the cognitive demand of Experiment 2 may be greater still.

11.2.3 Familiarity versus novelty preferences in listening preference task.

In the non-native condition in Experiment 2, in which French suprasegmental cues carried native AusE consonants and vowels, 12-month-olds showed a trend towards a novelty preference, tending to listen to sentences containing unfamiliarized words longer than sentences containing familiarized words. This is in contrast to the findings from similar prior studies, in which discrimination occurs via a familiarity preference. The direction of preference by the 12-month-olds in Experiment 2 is also in contrast to the direction of preference by the 15-month-olds in Experiment 3, and the 12- and 19-month-olds in Experiment 4.

This change in direction of discrimination is intriguing. It is thought that novelty preferences tend to arise when participants have been able to fully process and learn the familiarization stimulus – when they have been able to fully encode the familiarization stimuli during the familiarization period (Houston-Price & Nakai, 2004). In contrast, familiarity preferences are thought to result when the
encoding may not be fully complete, though that does not mean that participants have failed to learn or become familiar with the stimuli. In this way, factors that can slow the rate at which participants can fully encode the stimuli can affect whether a familiarity preference is seen over a novelty preference. Factors such as amount of familiarization time and age of the participants have been proposed to affect how quickly participants are able to fully encode the stimuli, and thus affect whether a familiarity or novelty preference is seen. However, in a previously published study that compared 3.5- and 6.5-month-olds' direction of preference in a visual preference task, it was found that the direction of the preference was dependent more on amount of familiarization time than age (Rose, Gottfried, Melloy-Carminar, & Bridger, 1982). Still, all experiments in this thesis shared the same familiarization time per word, and thus familiarization time alone cannot account for the discrepancy between the novelty trend found in 12-month-olds in the non-native accent condition in Experiment 2 and the familiarity effects found in 15-month-olds in Experiment 3 and 12- and 19-month-olds in Experiment 4. Further, a previous study using a similar procedure with 12-month-olds exposed children to twice as much familiarization time, yet found a familiarity preference (Schmale et al., 2010).

Eliciting a familiarity or novelty preference is also dependent on the complexity of the stimuli and task (e.g., whether or not stimuli must be learned cross-modally). However it has been proposed that the point at which a familiarity preference changes to a novelty preference is when the encoding of a stimulus has reached a point at which there is no discrepancy between the encoding and the stimulus (see Houston-Price & Nakai, 2004). In Schmale et al. (2010),
familiarization was in either the participants' native regional accent, or a non-native regional accent, and participants were subsequently tested on sentences in the alternate accent. In those cases, words in the test condition differed segmentally from those words participants had been familiarized to, and thus there was a discrepancy between the encoded words and subsequent test stimuli, contributing to a familiarity preference. Likewise, in Experiments 3 and 4, participants were familiarized to stimuli in their native regional accent, and tested on sentences in a non-native regional accent, in which segmental information was non-native but suprasegmental features were kept native. This too resulted in a discrepancy between the encoded stimuli and the test stimuli. In Experiment 2, there was no cross-accent discrepancy in the segmental information between the familiarization and test stimuli. In the non-native condition, familiarization stimuli were produced by a speaker of AusE. Segmental information in the test condition was also from a speaker of AusE, with the suprasegmental features from a speaker of FrE synthesized onto the stimuli. This could explain why a trend towards a novelty preference was found only in this condition, particularly as we proposed that children's tendency to discriminate stimuli in the non-native regional accent was due to their ability to ignore the non-informative suprasegmental information (see Section 9.3). Ignoring the non-informative non-native suprasegmental features, we posit that participants in Experiment 2 may have been able to recognize the segmentally similar pronunciations, whereas in Experiments 3 and 4, it was the segmental information that was more disparate compared to the familiarization stimuli, thus eliciting a familiarity preference.
However, while this account is promising, 12-month-olds in Experiment 4 showed a familiarity preference in the native regional accent condition in which native segments were synthesized with native suprasegmental information. If a familiarity preference only occurs when there is a mismatch between encoded familiarization stimuli and test stimuli, 12-month-olds should have instead shown a novelty preference. It should be noted that participants in Experiment 4 heard test sentences a maximum of four times, twice containing a familiarized word and twice containing an unfamiliarized word, whereas participants in Experiment 2 heard each sentence a maximum of two times, with only one word. However, it is unclear how this difference could lead to a change in the direction of looking preference. While on the whole this remains a side issue to the main questions of the thesis, it still provides interesting avenues for further research. A likely approach would involve systematic manipulation of the length of familiarization time and number of familiarization stimuli. However, the primary concerns of this thesis focused only on whether or not discrimination was shown. That is, the direction of participants' preference was not of theoretical importance here, as a preference in either direction implies that segmentation has occurred.

11.2.4 Implications.

11.2.4.1 The importance of controlling task demands. As has been discussed previously (see Fennell & Werker, 2003), task demands can affect experimental outcomes, particularly when examining the development of particular abilities. The results obtained in this series of experiments underscore the importance of carefully controlling task demands in infant research. In Experiment 2, we found that 12-month-olds were unable to discriminate sentences containing
familiarized versus unfamiliarized words in their native regional accent, despite their ability to do so previously in a similar study (Schmale et al., 2010). This difference in findings is likely to be due to our having used twice as many stimuli, which increased the difficulty to the point that children could not segment familiarized words from the test sentences. As well, each test sentence was played a maximum of only two times, and each time contained the same familiarized or unfamiliarized words in Experiment 2, but each sentence was played a maximum of four times in Experiment 4, and twice contained a familiarized word, and twice an unfamiliarized word. The reduced number of exposures may have further taxed 12-month-olds' ability to segment the test sentences in Experiment 2, relative to Experiment 4.

**11.2.4.2 Fragility of listening preference tasks.** Listening preference tasks are popular due to their minimal response demand – while infants are unable to reliably indicate directly whether or not they recognize an item, they have been found to passively convey this via looking preferences. However, while the task has had successes in testing children 7.5-13 months of age (Jusczyk & Aslin, 1995; Schmale & Seidl, 2009), to our knowledge ours was the first to test older children, at 15 and 19 months. Our results suggest that the listening preference task may not be ideal for children at these ages. While 12-month-olds discriminated sentences containing familiarized words from those containing unfamiliarized words in the native regional accent in Experiment 4, 15- and 19-month-olds did not. However, 15-month-olds showed discrimination for sentences in which the familiarized and unfamiliarized words contained CS vowel variations (Experiment 3), and 19-month-olds did so for CS consonant variations (Experiment 4). Thus, rather than showing
a total failure to perform the task, they were able to perform the task in a condition involving more challenging stimuli, which suggests that the task in the native accent conditions may not have been sufficiently engaging at that age, and is more suitable for younger participants.

However, even the younger participants demonstrated a fragile performance. As discussed in the previous section, the task in the present research was substantially harder than in previously published studies. In Experiment 2, we believe that the increased stimulus load overwhelmed the 12-month-olds' ability to perform the task in the native accent condition. While this observation highlights the importance of controlling task demands in infant research, it also underscores that behavior that is seen within a laboratory setting may itself be fragile, and perhaps not yet fully emerged as to be fully applicable in real life situations. Identifying the robustness of behaviors rather than simply their emergence could be of equal importance in language development research.

**11.2.4.3 Further directions (methodological).** To clarify the findings in Experiment 2, ideally the study should be run with only half of the stimulus load. The unexpected performance of the 12-month-olds in Experiment 2 was attributed to interference of task demands. Reducing these demands would thus help to clarify their behavior, and allow us to better relate the findings to the existing literature. Reducing the stimulus load might also elucidate the results from Experiments 3-4 as well. While 12-month-olds showed discrimination in the native regional accent in Experiment 4, running the study with a reduced stimulus load would confirm whether 12-month-olds were unable to discriminate the stimuli in the non-native segments conditions. Further, we have proposed that the listening preference task
was insufficiently engaging for the 15- and 19-month-old participants. Their failure to show discrimination in the native regional accent condition with a reduced stimulus load would confirm that their performance in the native accent conditions in Experiment 3 and 4 was based on the unengaging nature of the task, and not somehow due to an increased stimulus load.

11.3 Development of Phonologically Specified Word Forms

Despite the methodological issues discussed above, these experiments have revealed some novel findings, and have been able to inform the developmental progression of phonologically specified word forms.

11.3.1 Emergence of phonological constancy. Experiment 1 demonstrated that phonological constancy is more developed at 19 months than at 15 months. While both 15- and 19-month-olds looked more at named objects in their native regional accent, only 19-month-olds did so in the non-native regional accent, which contained both suprasegmental and segmental variation from native pronunciations. This finding is in line with previous results demonstrating that only 19-month-olds prefer to listen to high frequency early vocabulary words more than to low frequency adult vocabulary words in a non-native regional accent, whereas both ages show this preference in the native regional accent (Best et al., 2009).

11.3.2 Role of vocabulary. Expressive vocabulary size was found to positively correlate with performance in the non-native regional accent in Experiment 1, and on closer examination the correlation lay in the 15-month-olds' performance. This finding is in line with previous results demonstrating a link between vocabulary and grasp of phonological distinctiveness (e.g., Werker et al.,
The importance of vocabulary is further highlighted by the fact that the span between 15- and 19-month-olds envelops the onset of the vocabulary spurt, at around 17-18 months. As noted earlier, however, not all children demonstrate a sudden and dramatic increase in expressive vocabulary (Ganger & Brent, 2004). Still, there appears to be a qualitative change in children's performance that is dependent on vocabulary – perhaps on reaching a critical number of expressive words. This is supported by the finding that 15-month-olds' identification of non-native accented words was correlated with their expressive vocabulary, whereas it was not correlated with the 19-month-olds' identification. Regardless of whether rate of vocabulary growth underwent a spurt between those ages or only gradually increased, 19-month-olds still have larger expressive vocabularies to 15-month-olds. It is difficult to deduce whether the increasing lexicon fosters phonologically specified word forms, or if developing phonological knowledge eases the cognitive load required in learning new word forms. Most likely both processes are in play, such that vocabulary reaches a critical point when the child has acquired sufficient knowledge that allows the derivation of higher-order structures of words. Attunement to these higher-order structures then makes the acquisition of additional words more efficient, resulting in a steady rate of vocabulary development.

**11.3.3 Sensitivity to suprasegmental features.** As young children rely heavily on metrical patterns for early word segmentation (see Section 3.4.2 Children's use of the MSS), the purpose of Experiment 2 was to determine when children can segment words in the face of suprasegmental variation – that is, variation that would alter the metrical patterning, making it a less reliable or non-informative cue to word boundaries. The ability to do so would indicate participants
are progressing towards more adult-like word segmentation, where word segmentation occurs primarily via processing of the segmental information in speech, and occurs in tandem with word recognition. Twelve-month-olds demonstrated a tendency towards an ability to segment words from a synthesized accent containing non-native suprasegmental features, but native segmental information, indicating that they may be at a stage where they are not only able to segment words based on the segmental structure, but can also do so when the suprasegmental information is less- or non-informative. By our proposal, 12-month-olds' ability to discriminate sentences containing familiarized words from sentences containing unfamiliarized words in the non-native suprasegmental condition was due to their ability to ignore the non-informative suprasegmental features. This follows from our proposal that their inability to do so in the native accent condition was due to the cognitive demand imposed by the high stimulus load causing them to rely heavily on the metrical pattern for word segmentation, which directed their attention away from the segmental features. Thus, rather than resolving the non-native suprasegmental features through attention to the higher-order phonological structure of a word, it might instead be ignored. In mature word segmentation, suprasegmental features are believed to take on a supplementary, though still active, role (e.g., Norris & McQueen, 2008). We propose that in mature word segmentation, when suprasegmental metrical cues are found to be non-informative, attention may be diverted away from them, and segmentation continues via increased attention to the segmental features. This may be a component of the increased cognitive processing found in adult recognition of foreign-accented speech (e.g., Adank et al., 2009; Floccia et al., 2006; Snijders et al., 2007). The
results from Experiment 2 suggest that by 12 months children may still fall back on
the MSS for word segmentation in cognitively demanding situations, but may
nonetheless be able to segment words on the basis of the segmental structure, if
their attention is directed away from non-native metrical patterns that are non-
informative. While further investigation is needed to clarify this claim, we believe
that this ability to segment words across suprasegmental variation is not necessarily
tied to the development of phonologically specified word forms.

11.3.4 Sensitivity to segmental features. Experiments 3 and 4 revealed a
developmental trend across the ability to segment words from speech containing
segmental variation to vowels and to consonants. In Experiment 3, 15-month-olds
discriminated sentences containing familiarized words from sentences containing
unfamiliarized words when the words contained CS variations to vowels. In
Experiment 4, they did not show this sensitivity for CS variations to consonants,
whereas 19-month-olds did, indicating a developing sensitivity to vowels at 15
months, and to consonants at 19 months. Further, as neither 15- nor 19-month-olds
showed discrimination in the native accent conditions, but did so for CS variations
in the non-native segmental conditions, this suggests that they are at a point where
they are specifically attentive to variations to vowels and consonants. As we know
that by 19 months children have developed phonologically-specifed word forms,
children's performance in Experiments 3 and 4 suggest that we have pinpointed that
children are developing phonological constancy for vowels at 15 months, and
consonants at 19 months. This is consistent with the 15-month-olds' performance in
Experiment 1. Nineteen-month-olds, having developed phonological constancy to
both vowels and consonants, were able to understand non-native accented
pronunciations of words. Fifteen-month-olds, having developed phonological constancy only for vowels, were unable to understand the accented pronunciations which contained consonant variation as well as variation to vowels.

It should be noted that although we are proposing a developmental progression of emerging phonological constancy for vowels at 15 months and consonants at 19 months, this refers to development of phonological word forms, and not of individual consonants and vowels. That is, the perceptual reorganization here is analogous to, but should not be confused with, the perceptual reorganization that takes place by a child's first birthday, at which point native phonemic categories emerge for vowels and consonants. While the reorganization observed in the present research likely comes about via similar processes, at this point it is occurring at the word level, rather than at the phoneme level.

11.3.5 Category Goodness versus Category Shifting. In Experiments 3 and 4, while 15- and 19-month-olds respectively showed attention to CS vowels and CS consonants via their discrimination in those conditions, this discrimination did not extend to CG variations. Instead, they showed non-discrimination in these conditions, mirroring their performance in the native regional accent condition. Though further studies would verify this claim, it appears that the CG variations were treated as native-like. This is consistent with the predictions put forth by PAM (Best, 1994b, 1995). As CG variations map onto the corresponding native phonemic categories, they are theoretically more likely to be perceived as native-like than CS variations which map onto another category. As well, by 10.5 months children have developed a more primitive constancy that allows them for the first time to understand speech across speaker variation (e.g., Houston & Jusczyk, 2000).
be that CG variation falls within the extent of variation infants have experienced across speakers within a regional accent, and thus can be understood earlier than CS variations, and without the phonologically specified word forms required to perceive more varied speech.

11.4 Developmental Course of Word Segmentation and Phonologically Specified Word Forms: A Proposed Model

These results help clarify the developmental progression children travel as they develop robust spoken word recognition. We know from existing research that by 10.5 months, children can segment words that do not adhere to the predominant metrical pattern in their language (Jusczyk et al., 1993; Jusczyk, Houston, et al., 1999), suggesting a reduced reliance on the MSS for word segmentation, and perhaps indicating children are instead beginning to segment words on the basis of their segmental structure. This may follow from increased (receptive) vocabulary allowing for effective use of the possible word constraint as a segmentation strategy. From Experiment 2 we can further say that by 12 months children may be able to segment words that do not adhere to native metrical patterning, but instead have non-informative metrical cues. This is an important developmental step, because it demonstrates that by 12 months, children's reliance on the MSS has perhaps taken on a supplementary role, as it may be able to be ignored and word segmentation continue on the basis of processing the segmental input alone. Such a step is necessary for robust word recognition to occur, as it enables the perceiver to recognize speech across foreign-accented second language speakers who may impose their native language metrical patterning onto the spoken words of their second language (Gut, 2003). Accommodating this type of variation, then, does not
reflect the development of phonologically specified word forms, but rather is
reflective of an important qualitative shift in word segmentation that is progressing
toward more mature processing. This similarly depends on the expanding
vocabulary and cognitive resources to segment words without recourse to metrical
cues.

Thus it appears children may be able to segment words on the basis of
segmental structure by 12 months, and that with this proposed ability may come the
ability to segment words across suprasegmental variation, which emerges from the
ability to ignore non-informative suprasegmental information. However, this robust
word recognition does not extend to segmental variation yet, as 12-month-olds are
still lacking phonologically specified word forms. Instead, it appears that a
significant step in the phonological specification of word forms begins to take place
at around 15 months. Fifteen-month-olds in Experiment 3 were able to segment
words containing CS vowel variation, but were unable to do so for stimuli
containing CS consonant variation. This suggests that while they have begun to
develop phonological word forms, the specificity for vowels and consonants
develop at different paces. At 15 months children's word forms are apparently only
tolerant of vocalic variation. It is not until 19 months that phonological word forms
develop specificity for consonants as well, mirroring the developmental trend seen
in the development of native language phonetic tuning in the first year of life. These
phonological skills may be the result of continued exposure to a broad range of
distributions of segmental production, allowing for better recognition of invariant
structures, allowing the infant to recognize systematic changes to productions of
phonemes that occur between speakers and accents. Furthermore, once word
learning has begun to take off, it is increasingly likely that the words they hear are already known words, allowing for recognition that a variant production is less likely to be a novel word than an unusual pronunciation of a known word.

This progression is in tune with the results from Experiment 1. Fifteen-month-olds could not identify words in the non-native regional accent due to their inability to overcome the consonant variation, despite their ability to overcome the variation to vowels. Nineteen-month-olds, having word forms that contain phonological specification for both consonants and vowels, were able to recognize the words across both types of segmental variation. Importantly, for the non-native conditions, the performance of the 15- and 19-month-olds in Experiment 4 is mirrored in Experiment 1. Further, Experiment 4 required more from participants – segmenting the words from continuous speech, whereas Experiment 1 instead examined children's ability to identify individual words without requiring segmentation, as words were spliced into the sentence, occurred always at the end (so that placement was predictable), and involved a repetition of the word in isolation. The results of Experiment 4 provide more evidence that at 19 months children have become capable of robust, mature spoken word recognition. They are at a stage where they may be able to not only ignore irrelevant suprasegmental variation, but as well their recognition and segmentation abilities appear to coincide, and they can overcome systematic segmental variation to both consonants and vowels.

Based on these findings we propose a model of the development of segmentation. The components of this model are as follows:
1. Early word segmentation and word recognition are two processes in children, and one process in adults.

   a) Adults primarily segment words via lexical competition driven by processing the incoming stream of segments. Other cues such as attention to metrical patterning, knowledge of systematic cues to word boundaries (e.g., allophonic variation, phonotactics) serve as supplementary.

   b) To learn words children use cues that would be supplemental in adults to discern possible word boundaries, and from there begin to learn the segmental structures of words and build early word forms. This may be through a statistical or exemplar based process.

2. Early word forms are phonetically specified.

3. As the lexicon expands, supplemental cues can be shed, but word forms remain phonetically specified. This is likely to be dependent on having a large enough vocabulary to map possible phrase segmentations onto words, which must depend on both exposure to sufficient words and likely the development of cognitive resources to handle this.

4. Word forms become phonologically specified in a progression, with permissible variation to vowels (15 months) preceding that for consonants (19 months).

5. This development coincides with the vocabulary spurt, possibly in a bootstrapping manner: as more words are learnt, there is more data to be used in the identification of invariant structures, and this in turn allows for further segmentation of new words.
6. The emergence of phonologically specified words allows for more robust, mature word segmentation in two ways:

a) They allow effective segmentation in the face of speaker variation, and are able to handle even segmental variation across regional accents, while the lexicon allows supplemental cues to be ignored.

b) These phonologically specified word forms are specified to a higher-order than phonetically specified word forms, and are thus less cognitively taxing to access, allowing for robust segmentation across increased task demands.

The findings from these experiments are broadly consistent with the PRIMIR model (Werker & Curtin, 2005). The model posits that children's early word forms are phonetically specified, but that they gain access to higher-order word forms around the time of the vocabulary spurt. Challenging to this approach, however, is that the authors maintain that word representations are phonemically, rather than phonologically, specified. Were word forms phonemically specified, we would not expect children to recognize words with systematic variations that violate native phonemic categories, as was found in Experiment 1, and particularly in Experiments 3 and 4. In this respect, the findings are more in line with the Perceptual Attunement approach (Best et al., 2009), which proposes that children shift from phonetically specified word forms to phonologically specified word forms. These are more abstract than phonemically specified forms, and thus can allow systematic shifts in phonetic patterns across existing phoneme boundaries. No approach makes specific predictions regarding differing developmental progressions.
for cross-accent perception of consonants and vowels. However, a large body of research has documented both qualitative differences in perception and differences in developmental progression of language-specific perception for consonants and vowels. Experiments 3 and 4, where 15-month-olds display sensitivity to cross-accent CS variations to vowels and 19-month-olds' sensitivity to cross-accent CS variations to consonants in word recognition, are consistent with this apparent difference in development and may prefigure the perceptual differences found in adults. Finally, our finding that more phonologically specified word forms capable of resolving variation to both consonants and vowels develop between 15 and 19 months, as shown in Experiment 1, is in line with approaches which posit a developmental shift in word form specificity at around the time of the vocabulary spurt (e.g., Best et al., 2009; Werker & Curtin, 2005).

These results are also largely consistent with Shortlist B (Norris & McQueen, 2008), in that both a large vocabulary and knowledge of native language word confusions are required. As such, word segmentation should improve as vocabulary increases and exposure increases, which the results from Experiment 1 support. While Shortlist B does not explicitly define the processes involved in cross-accent listening, there is a rough correspondence between tolerance for phonetic variation found here and the native-accent based confusion probabilities in which Shortlist B is grounded. It may be a fruitful avenue of research to explore this model in a non-native context.

The results here do not fit well with statistical learning accounts (e.g., Saffran, Newport, et al., 1996), as these accounts propose that word forms have increasing phonetic specificity that builds with experience. While this does
successfully predict phonological distinctiveness improving as a skill as children get older, it does not allow for the emergence of phonological constancy. By this account, children in these experiments should never have been able to recognize cross-accent pronunciations, as they were phonetically varied from previously experienced tokens. Similarly, lexical density accounts (Brown & Matthews, 1997; Metsala & Walley, 1998) and distributional accounts (Thiessen, 2007) both predict that as vocabulary increases, so too does the specificity of word forms. This implies that younger children should be able to recognize more varied pronunciations that would fit with their relatively underspecified word forms. Our data show that as vocabulary and/or age increase, children are more, rather than less, able to accommodate phonetic shifts, which is a finding more in line with approaches such as PRIMIR (Werker & Curtin, 2005) and the Perceptual Attunement approach (Best et al., 2009), which propose a qualitative shift in attention from phonetically to phonologically (or phonemically, in the case of PRIMIR) defined word forms.

11.5 Conclusion

This research has uncovered evidence that robust, mature spoken word recognition may emerge in two separate phases. In the first phase, at around 12 months, children may be able to segment words via attention to the segmental structure of words, which coincides with their proposed ability to ignore non-informative suprasegmental cues to word boundaries, which has previously served a primary role in early word segmentation. In the second phase, phonologically specified word forms develop, with specification for vowels occurring at 15 months, and specification for consonants occurring at 19 months. These higher-order word
forms not only allow recognition of words containing segmental variation, but are less cognitively demanding to access, and thus accompany a robustness to cognitive demands. It remains to be seen at what age children are able to extract new words from running speech in the face of phonetic variation, or at what age novel words are learned as phonologically specified in running speech. The principles uncovered here, of separating phonological constancy and distinctiveness, and separating segmental variation from suprasegmental, should prove to be crucial in exploring these, and other, research questions.
12 References


Acoustical Society of America, 118, 1661–1676. doi:10.1121/1.2000774


219


spectrum, F0, and aperiodicity estimation (pp. 3933–3936). Presented at the
IEEE International Conference on Acoustics, Speech and Signal Processing,
2008, Las Vegas, Nevada. doi:10.1109/ICASSP.2008.4518514

Kitamura, C., Panneton, R., & Best, C. T. (resubmitted). The development of
attention to native versus non-native accents.

Kluender, K., Diehl, R., & Killeen, P. (1987). Japanese quail can learn phonetic

doi:10.1068/p3155

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L., Kozhevnikova, E. V.,
phonetic units in language addressed to infants. *Science*, 277(5326), 684–
686. doi:10.1126/science.277.5326.684

Kuhl, P. K., & Miller, J. D. (1975). Speech perception by the chinchilla: voiced-
voiceless distinction in alveolar plosive consonants. *Science*, New Series,
190(4209), 69–72. doi:10.1126/science.1166301

Kuhl, P. K., & Padden, D. M. (1982). Enhanced discriminability at the phonetic
boundaries for the voicing feature in macaques. *Perception &

context in the unattended ear. *Cognition, 1*, 359–372. doi:10.1016/0010-
0277(72)90002-9

prosodically-guided speech understanding* (Univac Report No. PX10791).
St. Paul: Sperry Univac.

Levitt, C., & van de Vijver, R. (1998). Syllable types in cross-linguistic and
developmental grammars. Presented at the Third Biannual Utrecht
Phonology Workshop, Utrecht.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M.


231

Mulak, K. E., Best, C. T., Tyler, M. D., & Kitamura, C. (2012). Twelve-month-olds may segment words from continuous speech in the face of non-native suprasegmental features. Presented at the International Conference on Infant Studies, Minneapolis, USA.


Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition


235


prosodic bootstrapping of phrases: Evidence from prelinguistic infants. 


Vitevitch, M. S., & Luce, P. A. (1998). When words compete: Levels of processing
doi:10.1111/1467-9280.00064


doi:10.1121/1.2168414


Werker, J. F., & Stager, C. L. (2000). Developmental changes in infant speech perception and early word learning: Is there a link? In M. B. Broe & J. B. Pierrehumbert (Eds.), *Papers in Laboratory Phonology V: Acquisition and
the lexicon (pp. 181–193). New York: Cambridge University Press.


13 Appendix A: Additional Analyses for Experiment 1
Figure A-1. Fifteen- and 19-month-olds' raw horizontal location of gaze during two target word repetitions in AusE, organized into 100 ms bins. Average word repetitions occurred at 0-677 ms and 2000-2677 ms. Data has been corrected so that looking above 400 reflects looking to the side of the screen corresponding to the target image, and points below 400 correspond to looking to the side of the screen containing the distractor image. Gray shaded areas indicate periods at which 19-month-olds' gaze was significantly different from chance (400). At no point did 15-month-olds' gaze differ from chance.
Figure A-2. Fifteen- and 19-month-olds' raw horizontal location of gaze during two target word repetitions in JaME, organized into 100 ms bins. Average word repetitions occurred at 0-677 ms and 2000-2677 ms. Data has been corrected so that looking above 400 reflects looking to the side of the screen corresponding to the target image, and points below 400 correspond to looking to the side of the screen containing the distractor image. The gray shaded area indicates the period at which 19-month-olds' gaze was significantly different from chance (400). At no point did 15-month-olds' gaze differ from chance.
Figure A-3. Fifteen- and 19-month-olds' raw horizontal location of gaze during two target word repetitions in AusE, organized into 200 ms bins. Average word repetitions occurred at 0-677 ms and 2000-2677 ms. Data has been corrected so that looking above 400 reflects looking to the side of the screen corresponding to the target image, and points below 400 correspond to looking to the side of the screen containing the distractor image. The gray shaded areas indicate periods at which 19-month-olds' gaze was significantly different from chance (400). At no point did 15-month-olds' gaze differ from chance.
Figure A-4. Fifteen- and 19-month-olds' raw horizontal location of gaze during two target word repetitions in JaME, organized into 200 ms bins. Average word repetitions occurred at 0-677 ms and 2000-2677 ms. Data has been corrected so that looking above 400 reflects looking to the side of the screen corresponding to the target image, and points below 400 correspond to looking to the side of the screen containing the distractor image. The gray shaded area indicates the period at which 19-month-olds' gaze was significantly different from chance (400). At no point did 15-month-olds' gaze differ from chance.
Figure A-5. Fifteen- and 19-month-olds' raw horizontal location of gaze during two target word repetitions in AusE, organized into 400 ms bins. Average word repetitions occurred at 0-677 ms and 2000-2677 ms. Data has been corrected so that looking above 400 reflects looking to the side of the screen corresponding to the target image, and points below 400 correspond to looking to the side of the screen containing the distractor image. The gray shaded areas indicate periods at which 19-month-olds' gaze was significantly different from chance (400). At no point did 15-month-olds' gaze differ from chance.
Figure A-6. Fifteen- and 19-month-olds’ raw horizontal location of gaze during two target word repetitions in JaME, organized into 400 ms bins. Average word repetitions occurred at 0-677 mms and 2000-2677 ms. Data has been corrected so that looking above 400 reflects looking to the side of the screen corresponding to the target image, and points below 400 correspond to looking to the side of the screen containing the distractor image. The gray shaded area indicates the period at which 19-month-olds’ gaze was significantly different from chance (400). At no point did 15-month-olds’ gaze differ from chance.
Figure A-7. Fifteen- and 19-month-olds' raw horizontal location of gaze during two target word repetitions in AusE, organized into 500 ms bins. Average word repetitions occurred at 0-677 ms and 2000-2677 ms. Data has been corrected so that looking above 400 reflects looking to the side of the screen corresponding to the target image, and points below 400 correspond to looking to the side of the screen containing the distractor image. The gray shaded areas indicate periods at which 19-month-olds' gaze was significantly different from chance (400). At no point did 15-month-olds' gaze differ from chance.
Figure A-8. Fifteen- and 19-month-olds' raw horizontal location of gaze during two target word repetitions in JaME, organized into 500 ms bins. Average word repetitions occurred at 0-677 ms and 2000-2677 ms. Data has been corrected so that looking above 400 reflects looking to the side of the screen corresponding to the target image, and points below 400 correspond to looking to the side of the screen containing the distractor image. The gray shaded area indicates the period at which 19-month-olds' gaze was significantly different from chance (400). At no point did 15-month-olds' gaze differ from chance.
Table A-1

Mean time of first peak to the target image across 15- and 19-month-olds in native and non-native regional accent conditions

<table>
<thead>
<tr>
<th></th>
<th>Native regional accent (AusE)</th>
<th>Non-native regional accent (JaME)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>15 mos</td>
<td>19 mos</td>
</tr>
<tr>
<td>Mean (ms)</td>
<td>387.50</td>
<td>493.75</td>
</tr>
<tr>
<td>SD (ms)</td>
<td>202.90</td>
<td>354.91</td>
</tr>
</tbody>
</table>

Note. A 2 x 2 ANOVA comparing accent and age group failed to reveal main effects of accent, $F(1, 30) = 0.32, p = .58$, or age group, $F(1, 30) = 0.70, p = .41$, and did not reveal an interaction between accent x age group, $F(1, 30) = 0.32, p = .58$.

Table A-2

Mean time of first peak to the target image with a value of at least 420 pixels, across 15- and 19-month-olds in native and non-native regional accent conditions

<table>
<thead>
<tr>
<th></th>
<th>Native regional accent (AusE)</th>
<th>Non-native regional accent (JaME)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>15 mos</td>
<td>19 mos</td>
</tr>
<tr>
<td>Mean (ms)</td>
<td>1275.00</td>
<td>1231.25</td>
</tr>
<tr>
<td>SD (ms)</td>
<td>768.12</td>
<td>778.65</td>
</tr>
</tbody>
</table>

Note. A 2 x 2 ANOVA comparing accent and age group failed to reveal main effects of accent, $F(1, 30) = 0.02, p = .88$, or age group, $F(1, 30) = 0.01, p = .95$, and did not reveal an interaction between accent x age group, $F(1, 30) = 0.02, p = .88$.

Table A-3

Mean time of first peak to the target image with a value of at least 440 pixels, across 15- and 19-month-olds in native and non-native regional accent conditions

<table>
<thead>
<tr>
<th></th>
<th>Native regional accent (AusE)</th>
<th>Non-native regional accent (JaME)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>15 mos</td>
<td>19 mos</td>
</tr>
<tr>
<td>Mean (ms)</td>
<td>1875.00</td>
<td>1487.50</td>
</tr>
<tr>
<td>SD (ms)</td>
<td>1186.31</td>
<td>598.75</td>
</tr>
</tbody>
</table>

Note. A 2 x 2 ANOVA comparing accent and age group failed to reveal main effects of accent, $F(1, 30) = 0.27, p = .60$, or age group, $F(1, 30) = 0.17, p = .69$, and did not reveal an interaction between accent x age group, $F(1, 30) = 1.12, p = .30$. 

252
14 Appendix B: Participant Information and Consent Forms
Figure B-1. Participant information form for Experiment 1.
CONSENT FORM FOR PARTICIPANTS

Please read the information sheet and instructions before signing this.

1. I, ____________________________,
   of ____________________________,
   agree to participate, along with my baby, as a participant in the experiment described in
   the participant information statement attached to this form.

2. I acknowledge that I have read the participant information statement, which explains why I have
   been selected, the aims of the experiment and the nature and the possible risks of the investigation,
   and the statement has been explained to me to my satisfaction.

3. I understand that I can withdraw from the recording at any time, and I understand that my decision
   whether or not to participate in or subsequently withdraw from this study will not affect any current
   or future relationship to the University of Western Sydney, or the receipt of any advertised benefits.

4. I agree that research data gathered from the results of the study may be published or provided to
   other researchers, provided that I cannot be identified.

5. I understand that if I have any questions relating to my participation in this research, I may contact
   Professor Catherine Best (tel: 9772 6760) or Ms Suzana Bicanic (tel: 9772 6696) who will be
   happy to answer them.

6. I acknowledge receipt of a copy of this Consent Form and the Participant Information Statement.

7. I understand the purpose of the study and what is being asked of me, and that I can stop
   participating at any time without loss of any of the advertised rewards. With this understanding, I
   agree to take part in this research.

NOTE: This study has been approved by the University of Western Sydney Human Research Ethics Committee. If
you have any complaints or reservations about the ethical conduct of this research, you may contact the Ethics
Committee through the Research Ethics Officers (tel: 02 4370 1136). Any issues you raise will be treated in
confidence and investigated fully, and you will be informed of the outcome.

_________________________  __________________________  Date______________
Signature of subject        Please PRINT name
EXAMINING THE EMERGENCE OF PHONOLOGICAL CONSTANCY
AS A MECHANISM FOR WORD SEGMENTATION

PARTICIPANT INFORMATION SHEET
(PARENT/CARETAKER)

Who is carrying out the study?
You and your child are invited to participate in a study conducted by Karen Mulak, a PhD Candidate at MARCS Auditory Laboratories at the University Of Western Sydney. The study will form the basis for the PhD degree at the University Of Western Sydney under the supervision of Prof Catherine T. Best.

What is the study about?
The purpose is to investigate whether children’s ability to recognise words across foreign accents and non-native dialects (i.e., that differ phonologically from the native pronunciation) is related to their ability to segment the words from continuous speech.

What does the study involve?
This study will involve your child listening to some sentences and words while they look at a screen with either a checkerboard or pictures on it. The entire time your child will be seated on your lap.

Data will be collected today during testing and stored in password protected computer files for 5 years post-publication, after which they will be destroyed. The data can/will be accessed by the Primary Investigator, her supervisory panel, and research assistants directly involved in the project. Data will be used to obtain group results which will be published in scientific journals and a PhD thesis. No identifiable information concerning you or your child will be made public in any way. If you have concerns about what has been recorded, you may request access of recordings of your child within the period of storage. These recordings can be accessed by contacting Karen at (02) 9772 6573 or by email at k.mulak@uws.edu.au.

Children not participating in the study will be watched by the principal researcher or a research assistant during the time the study is being carried out.

How much time will the study take?
The study will take approximately 10 minutes. Today’s visit in total will last approximately 30 minutes.

Figure B-3. Participant information sheet for Experiments 2-4 (page 1 of 2).
Will the study benefit me or my child?
The study will not benefit you or your child, but will inform us about the development of
word segmentation and language development in children.

Will the study involve any discomfort for me?
There are no discomforts in this study. However, should you or your child be discomforted
at any time, please let us know, and we can stop immediately. Your participation is entirely
voluntary.

How is this study being paid for?
The study is being sponsored by the University Of Western Sydney.

Will anyone else know the results? How will the results be disseminated?
All aspects of the study, including results, will be confidential, and only the researchers will
have access to information on participants. The group results will be published in journal
articles, and a PhD thesis. No individual results will be published.

Can I withdraw my child from the study?
Your and your child’s participation is entirely voluntary. You are not obliged to consent to be
in the study. If you do choose to participate, you may withdraw your child from the study at
any time at which point all written and audio records of your child’s participation will be
destroyed.

Can I tell other people about the study?
Yes, you can tell other people about the study by providing them with the chief
investigator’s contact details. They can contact the chief investigator to discuss their
participation in the research project and obtain an information sheet.

What if I require further information?
When you have read this information, Karen will discuss it with you further and answer any
questions you may have. If you would like to know more at any stage, please feel free to
contact Karen at (02) 9772 6573.

What if I have a complaint?
This study has been approved by the University Of Western Sydney Human Research Ethics
Committee. The Approval number is [enter approval number]. If you have any complaints or
reservations about the ethical conduct of this research, you may contact the Ethics
Committee through the Office of Research Services on (02) 4736 0083, or by email at
humeethics@uws.edu.au. Any issues you raise will be treated in confidence and
investigated fully, and you will be informed of the outcome.

If you agree to participate in this study, you will be asked to sign a consent form.

Figure B-4. Participant information sheet for Experiments 2-4 (page 2 of 2).
EXAMINING THE EMERGENCE OF PHONOLOGICAL CONSTANCY AS A MECHANISM FOR WORD SEGMENTATION

PARENT/CARETAKER CONSENT FORM

I have been asked to participate with my child in the research EXAMINING THE EMERGENCE OF PHONOLOGICAL CONSTANCY AS A MECHANISM FOR WORD SEGMENTATION conducted by Karen Mulak and give my free consent by signing this form. I acknowledge that:

1. The research project will be carried out as described in the information sheet, a copy of which I have retained. I have read and understood the Information Sheet and have had the opportunity to discuss the information and my child’s involvement in the project with the researcher(s) to my satisfaction.

2. My child’s involvement is confidential and the information gained during the study may be published, but no personal information about my child will be used in any way that reveals my child’s identity.

3. My consent and my child’s consent to participate are voluntary. I may withdraw from the study at any time, without affecting my or my child’s relationship with UWS, or the MARCS Baby Lab. I do not have to give a reason for the withdrawal of my consent.

Please tick 'yes' or 'no':

Yes      No

☐ ☐ I consent to audio/video taping of the testing session. I understand the recording will be used only to aid data collection, and will not contain additional identifying information about my child.

Signed (Parent/Caretaker): .................................................................

Name: ............................................................................................

Date: ..............................................................................................

Figure B-5. Consent form for Experiments 2-4.