Speech Perception in Early Childhood: Contending with Speaker and Accent Variation

by

Marieke Helena Gerarda van Heugten

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Department of Psychology
University of Toronto

© Copyright by Marieke Helena Gerarda van Heugten (2012)
Abstract

Understanding spoken language is much more complex than common intuition may suggest. Speaker-related variation in the realization of words due to changes in speaker gender and accent results in myriad possible realizations of the same word. Although adult listeners can easily accommodate such speaker-specific pronunciations, research suggests that children have more difficulty with this variability. In this thesis, I examine when and how children learn to contend with speaker and accent variation.

In three lines of research, I test the hypothesis that infants and toddlers may be less challenged by acoustic variability than past research suggests. In Chapter 2, I show that 15-month-old infants, like adults, can readily accommodate an unfamiliar Australian accent after hearing a 2-minute story produced in Australian English. Interestingly, speaker exposure only elicits accommodation when infants are highly familiar with the words in the story. These findings are consistent with the hypothesis that lexical feedback serves as the mechanism inducing accent adaptation. More support for this idea comes from Chapter 3, where even 20-month-olds were found to experience difficulty
accommodating a speaker’s unfamiliar accent when the speaker uses hard words that are not consistently known by children this age. Chapter 3 further shows that 28-month-olds, but not 20-month-olds, recognize words produced in sentence frames by a speaker with an unfamiliar accent without any prior exposure to the accent, thereby providing converging evidence that young children are proficient at dealing with accent variability. Chapter 4 subsequently demonstrates that even 7.5-month-olds can contend with some forms of speaker variation when tested under relatively natural listening conditions.

Taken together, these three lines of research provide compelling evidence for the view that early speech processing abilities are remarkably sophisticated and that the transition from infant to adult listener may be less abrupt than currently thought.
Acknowledgments

Finishing this thesis marks the final days of my Toronto chapter. And although I am sad to leave, I am happy that after all this time I can finally say that I am done ‘school’. I made it through!!! But I didn’t do it alone. Even though this dissertation carries my name, many people have contributed to the endeavor. Without their help and support, finishing would have been impossible. I would like to thank you all. THANK YOU! No really, thank you.

First and foremost, I would like to thank my advisor, Elizabeth Johnson, for her invaluable guidance and endless patience throughout the years. I am grateful to have such a caring and available mentor. Thank you also for your enthusiasm and encouragement, and for teaching me to write. You have always stimulated me to perform to the best of my ability, offering unfailing encouragement and support where necessary. Thank you for letting me make my own mistakes when I stubbornly insisted on doing things my way and for not saying ‘I told you so’ (too much) afterwards. Having worked together for so long, I am sure it will be strange to move on and no longer be in your lab.

I would also like to thank the other members of my committee, Craig Chambers and Eyal Reingold, for dedicating their time and energy to my research and of course for their feedback and suggestions on my work. Craig, also thank you for your continuous support whenever I dropped by your office to ask you for advice. Thanks also to Bruce Schneider, Mark Schmuckler, and Kris Onishi who have generously agreed to read this thesis and be part of my defense committee.
If you study speech perception (and English is not your first language), you need people willing to produce your stimuli. The speakers of my studies deserve a big thank you, too. Melissa Paquette-Smith, Julie Kow, Mary Lamb, and Vicky Ng: thanks! Two of the experiments in this thesis present work in collaboration with Dena Krieger. Thank you for your investments in this project. In addition, Huiwen Goy deserves thanks for helping me with various Praat scripts. And of course, thanks to all research assistants who, in one way or another, have helped me complete my research in time. In particular, I would like to thank Caterina Minaudo for running participants when I was abroad. Thank you also to the families who participated in this research.

Life would surely not have been the same without my fellow grad students. Melissa Paquette-Smith and Dena Krieger, thanks for being awesome lab mates. It is nice being surrounded by such cheerful people! Dena, thank you for Scoozzing it up with me and for your feedback on this thesis. The HCL group at UTM also provided a supportive work environment. Thank you to all the people on the 4th floor for hanging out with me, keeping me company during endless shuttle rides, and discussing research!

Who wouldn’t want to spend two summer months in Paris? Thank you to Anne Christophe for hosting me and making sure I could complete testing my experiments and to Delphine Dahan who collaborated on the project. And a big thank you to all of those who helped me while I was there. Even if my French was insufficient to thank you all, I really appreciate it! I look forward to coming back.

Finally, I would like to thank my friends for being there for me (and – only very incidentally – listening to me rant). The road to thesis completion has not been without hurdles and I am glad I met such great friends during my time here in Toronto. Turns out
that grad school is the perfect place to meet people like you. I especially want to thank Renée Biss, Trevor Fernandes, Daniel Paquin, Jessica Remedios, Kristoffer Romero, and Stephanie Spielmann. You have made grad school fun. Our weekly pub nights surely helped me put school stress in perspective. And although I am happy to be finishing, I am sad that means that we no longer get to live in the same city. Thank you for being such good friends. I will miss you. In addition, Stephanie Stalinski, I promised you I would write it, and writing it is what I do here: I wish I were awesome like you. Cara Tsang, thank you for your appreciation of linguistic wordplay; Dominique Vuvan, thank you for sharing your roof with me, and Annick Moisan, thank you for helping me keep my French up to date (sort of)! Thank you also to my friends and family back home. Sorry for not coming back more often. And of course, Mom and Dad, thanks for your support and for your belief in me.
“John, when people thought the earth was flat, they were wrong. When people thought the earth was spherical, they were wrong. But if you think that thinking the earth is spherical is just as wrong as thinking the earth is flat, then your view is wronger than both of them put together.”

~Isaac Asimov in ‘The relativity of wrong’
# Table of Contents

Acknowledgments ................................................................................................................ iv

Table of Contents ............................................................................................................... viii

List of Figures .................................................................................................................. xii

List of Appendices ........................................................................................................... xiv

Chapter 1 General Introduction ....................................................................................... 1

1 Speaker-related variability in spoken language ......................................................... 1

2 The effects of speaker-related variability on infant speech perception ............... 3
   2.1 Dealing with speaker-related information in vowels ........................................ 3
   2.2 The effect of speaker information on infants’ word mapping abilities ....... 5
   2.3 The effect of speaker information on recognizing familiar words .......... 7
   2.4 Implications for infant speech perception ...................................................... 8

3 Adult listeners’ strategies for handling speaker-related variability ................. 10
   3.1 Consequences of speaker-related variability in adulthood ....................... 10
   3.2 Speaker accommodation as a solution to between-speaker variability ...... 11
   3.3 Mechanisms underlying speaker accommodation .................................... 12
   3.4 The scope of speaker accommodation ....................................................... 14
      3.4.1 Generalizing beyond previously heard items .................................... 14
      3.4.2 Within- versus between-speaker generalization ............................... 15
   3.5 Implications for theories of adult speech perception .................................. 18
   3.6 Speaker adaptation in adults: concluding remarks ................................... 20

4 Reconciling speech perception by infants and adults: Is there speaker 
   accommodation in infancy? ....................................................................................... 20

5 Outline of this dissertation .......................................................................................... 23

Chapter 2 The development of accent adaptation in infancy ................................ 27

1 Introduction ................................................................................................................ 27
2 Experiment 1 ........................................................................................................................................ 28
  2.1 Methods ........................................................................................................................................ 29
    2.1.1 Participants ................................................................................................................................. 29
    2.1.2 Stimuli ......................................................................................................................................... 30
    2.1.3 Procedure .................................................................................................................................. 31
  2.2 Results and Discussion ..................................................................................................................... 32

3 Experiment 2 ....................................................................................................................................... 33
  3.1 Methods .......................................................................................................................................... 34
    3.1.1 Participants ................................................................................................................................. 34
    3.1.2 Stimuli and Procedure ................................................................................................................ 34
  3.2 Results and Discussion .................................................................................................................... 35

4 Experiment 3 ....................................................................................................................................... 36
  4.1 Methods .......................................................................................................................................... 38
    4.1.1 Participants ................................................................................................................................. 38
    4.1.2 Stimuli ......................................................................................................................................... 38
    4.1.3 Procedure .................................................................................................................................. 38
  4.2 Results and Discussion .................................................................................................................... 39

5 Experiment 4 ....................................................................................................................................... 41
  5.1 Methods .......................................................................................................................................... 42
    5.1.1 Participants ................................................................................................................................. 42
    5.1.2 Stimuli and Procedure ................................................................................................................ 42
  5.2 Results and Discussion .................................................................................................................... 42

6 Discussion of Experiments 1-4 .......................................................................................................... 44

Chapter 3 Toddlers’ recognition of words spoken in unfamiliar accents ............................................. 47

1 Introduction ......................................................................................................................................... 47

2 Experiment 1 ....................................................................................................................................... 49
  2.1 Methods .......................................................................................................................................... 50
    2.1.1 Participants ................................................................................................................................. 50
    2.1.2 Stimuli ......................................................................................................................................... 51
2.1.2 Stimuli ........................................................................................................... 84
2.1.3 Procedure ..................................................................................................... 85
2.1.4 Design .......................................................................................................... 85
2.2 Results ............................................................................................................ 86
3 Discussion of the current experiment ................................................................. 87
Chapter 5 General Discussion .............................................................................. 92
1 Summary of findings .......................................................................................... 92
2 Implications for models of early speech perception .......................................... 94
3 Mechanisms inducing speaker accommodation in early childhood ............... 96
4 Cross-speaker generalization ............................................................................. 99
5 Concluding remarks .......................................................................................... 102
References ........................................................................................................... 104
Appendices ......................................................................................................... 117
List of Figures

Figure 1: Orientation time in seconds (error bars indicate standard errors of the mean difference scores) for 15-month-olds to known and nonsense words in Experiments 1a (native accent) and 1b (unfamiliar accent)................................................................. 33

Figure 2: Orientation time difference between known and nonsense words (produced in Australian-accented English) in seconds (error bars indicate standard errors of the mean difference scores) in Experiments 1b (15-month-olds) and 2 (17.5-month-olds and 22-month-olds, respectively).......................................................................................................................... 34

Figure 3: Orientation time in seconds (error bars indicate standard errors of the mean difference scores) to known and nonsense words in Australian-accented English in Experiments 1b (15-month-olds), and 2 (17.5- and 22-month-olds, respectively)............. 36

Figure 4: Orientation time in seconds (error bars indicate standard errors of the mean difference scores) for 15-month-olds to known and nonsense words in Experiment 3 (unfamiliar accent with exposure). ................................................................................................. 39

Figure 5: Orientation time in seconds (error bars indicate standard errors of the mean difference scores) for 15-month-olds to known and nonsense words in Experiments 1a (native accent), 1b (unfamiliar accent), 3 (unfamiliar accent with exposure), and 4 (unfamiliar accent with book familiarization and exposure). ................................................. 43

Figure 6: Visual representation of the test trials in Experiments 1, 2, and 3. ............... 53

Figure 7: Proportion of fixations to target in the time window before and after target word onset (on the left) and the proportion of fixations to target over time from target word onset (on the right) for 28-month-olds in Experiment 1a. ................................................. 56
Figure 8: Proportion of fixations to target broken down by heard word trials and generalization trials in the time window after target word onset (on the left) and the proportion of fixations to target over time from target word onset in these two trial types (on the right) for 28-month-olds in Experiment 1a......................................................... 57

Figure 9: Proportion of fixations to target in the time window before and after target word onset (on the left) and the proportion of fixations to target over time from target word onset (on the right) for 28-month-olds in Experiment 1b........................................ 58

Figure 10: Proportion of fixations to target in the time window after target word onset broken down by experiment (on the left) and the proportion of fixations to target over time from target word onset for 28-month-olds in the two experiments (on the right) ... 59

Figure 11: Proportion of fixations to target in the time window before and after target word onset (on the left) and the proportion of fixations to target over time from target word onset (on the right) for 20-month-olds in Experiment 2........................................ 66

Figure 12: Proportion of fixations to target broken down by heard word trials and generalization trials in the time window after target word onset (on the left) and the proportion of fixations to target over time from target word onset in these two trial types (on the right) for 20-month-olds in Experiment 2. ......................................................... 67

Figure 13: Proportion of fixations to target in the time window before and after target word onset (on the left) and the proportion of fixations to target over time from target word onset (on the right) for 20-month-olds in Experiment 3........................................ 72

Figure 14: Proportion of fixations to target broken down by heard word trials and generalization trials in the time window after target word onset (on the left) and the proportion of fixations to target over time from target word onset in these two trial types (on the right) for 20-month-olds in Experiment 3. ......................................................... 73

Figure 15: 7.5-month-old infants’ mean orientation times in seconds (and standard errors of the mean difference scores) to trained and novel test words broken down by Voice Familiarity........................................................................................................ 87
List of Appendices

**Appendix A:** Language background questionnaire administered to the participating families in Chapters 2 and 3 ........................................... 117

**Appendix B:** Lists of words used in Chapter 2 ................................................................. 119

**Appendix C:** Pictures used in Chapter 3 ........................................................................ 120

**Appendix D:** Video orders used in Chapter 3 .................................................................. 123

**Appendix E:** Passages used in Chapter 4 ........................................................................ 127
Chapter 1
General Introduction

1 Speaker-related variability in spoken language

At the heart of human communication lies the ability to rapidly and effortlessly comprehend speech produced by a variety of speakers, all of whom have unique vocal tract configurations and linguistic backgrounds. This is a truly impressive feat, as factors such as speaking rate, mood, speech register, and the position of words in sentences ensure that two tokens of the same word will never be physically identical, even when produced by a single speaker. In addition, between-speaker differences, such as those in voice characteristics and accents, further increase the variability in the speech signal. But if the same word can be realized in so many different ways, then how do listeners figure out which differences are linguistically meaningful and which are merely speaker-related? Distinguishing between these two alternatives cannot simply be a matter of determining the overall acoustic similarity of two word tokens. That is, the relatively salient acoustic difference between tokens of the word fool in a male versus a female voice, for example, may be linguistically irrelevant, whereas a potentially much more subtle difference between the words fool and full in the same voice distinguishes between two completely different words. Thus, achieving perceptual constancy while at the same time preventing confusion between minimal pairs is a non-trivial task that involves highly sophisticated speech perception abilities.

Adult listeners are proficient language users and possess flexible signal-to-word mapping processes, enabling them to link phonetically variable tokens of a word to
single underlying lexical representation (see Cutler, 2008 for an overview). Infants at the early stages of language development, in contrast, neither have a lexicon full of possible word candidates stored in their memory nor have they fully tuned in to the sound inventory of their native language. A fundamental question in the field of infant speech perception thus concerns how infants, in the absence of adult-like strategies and knowledge, start extracting the relevant information from a highly variable speech signal and how children learn to perceive stable word percepts in the absence of stable acoustic word realizations whilst ignoring non-contrastive variation.

In this dissertation, I examine these issues by looking at infants’ and young toddlers’ abilities to cope with speaker-related variability during speech processing. Specifically, I will focus on the way young listeners make use of information in the speech signal to better understand a speaker’s intended message. In the remainder of this chapter, I will outline the contemporary perspectives on infants’ and adults’ ability to deal with variation in the speech signal. First, in section 2, I discuss the difficulties infants face when dealing with between-speaker variation during word recognition. Section 3 then addresses how adults, whose speech perception abilities in laboratory experiments are also hindered by speaker variability, overcome these perceptual challenges imposed by speaker variability during speech perception under natural listening conditions by tuning into the characteristics of the speaker and subsequently use this information to better understand the speaker. In section 4, I raise the question of whether infants can also adapt to speaker idiosyncrasies. Lastly, section 5 concludes with my research objectives and an outline of the experiments presented in this dissertation.
The effects of speaker-related variability on infant speech perception

As adults, we usually recognize words regardless of whether they are produced by a male or female voice, in a North American, Scottish, or Australian accent, by a child, a peer, or a boss. Is this perceptual constancy for speech present from early on, or do infants need to learn that highly different acoustic patterns can refer to the same underlying representation? In order for infants to easily comprehend the speech of unfamiliar voices on TV or of their foreign-accented daycare provider, one would hope that their ability to deal with speaker variability is present early during the course of language development. Without such abilities, any acquired linguistic competence cannot be readily applied when listening to an unfamiliar accented speaker and would thus be rendered useless under many everyday listening conditions. Thus, understanding infants’ ability to process speech in the face of acoustic variability is crucial for understanding language acquisition and speech perception.

2.1 Dealing with speaker-related information in vowels

The question of how infants deal with speaker idiosyncrasies in the pronunciation of words was first addressed in the classic speech perception literature. Data from early speech sound categorization studies show that as early as at 6 months of age, infants have some ability to deal with speaker information in the realization of vowels (Kuhl, 1979; 1984). In these studies, infants first learned to turn their heads towards a visual reinforcer when hearing a (synthesized) vowel different in vowel quality from a baseline vowel. For example, when a list of isolated vowel [a]-vowels was used as background stimuli, a headturn would need to be made towards a visual reinforcer when the vowel [i] was
subsequently presented. Such headturns would indicate that the infants distinguished between the two vowels. In the following test phase, variability in talker and pitch contour was added. Despite this additional variability, infants were found to distinguish between the [a] and [i] vowels, showing appropriate head turns when a switch from [a] to [i] occurred and refraining from head turns on trials without a vowel switch, most notably, when the [a] vowel was presented from the voice of a different gender. This ability to generalize across speaker-induced variation when distinguishing between phonemes develops early in life. Infants as young as two months of age, for example, presented with the word form bug detected the change to the word form dug. When tested in the high-amplitude sucking procedure (a procedure often used in young children measuring their attention to the presented materials), infants increased their sucking rate after the phonological form of the word changed. This increase in sucking rate was observed regardless of whether these word forms were spoken by a single speaker or by multiple speakers, thereby demonstrating an early ability to cope with some acoustic variation in the realization of phonemes. Speaker variability did, however, lead to poorer retention of the phonological features of a word. Specifically, although immediate recognition scores were unaffected by the number of speakers infants heard, recognition scores after a two minute delay were impeded by added speaker variation (Jusczyk, Pisoni, & Mullennix, 1992).

While these studies provided researchers with a basic foundation of the effects of speaker information on linguistic processing in infancy, their limiting factor is that in order to perform well in this task, infants only need to detect relevant changes between stimulus tokens presented in immediate succession. In natural speech perception, much
more advanced processing is required. First, phonemes or syllables are not simply encountered in isolation – they need to be extracted from the continuous signal and subsequently mapped onto previously stored lexical information. Second, there is typically no opportunity to carry out direct phoneme comparisons across similar-sounding tokens. Thus, rather than discriminating between two tokens, speech perception outside the laboratory requires infants to recognize speech sounds as belonging to a certain phonemic category stored in (long-term) memory. If even two minutes of silence is sufficient to break down the ability to generalize across speakers (Jusczyk et al., 1992), then one may wonder to what extent infants may be able to apply the observed abilities during speech processing in everyday life.

2.2 The effect of speaker information on infants’ word mapping abilities

To examine infants’ initial competence in contending with speaker-related variability in a somewhat more naturalistic task, researchers have started to make use of a procedure that tests infants’ ability to map two distinct realizations of a word, one of which is embedded in fluent speech, onto the same underlying representation. In such studies, infants are first familiarized with multiple tokens of two isolated word forms (e.g., feet and bike). These words are presented over loudspeakers, without semantic or otherwise contextual information being available. It has been found that in the subsequent test phase, infants prefer to listen to passages containing these trained items than to passages containing novel items (e.g., cup and dog; Jusczyk & Aslin, 1995). Studies employing this methodology have shown that although infants have very specific criteria for what qualifies as a good match (e.g., no looking preference is observed when the words contain
a mispronunciation such as *tup* instead of *cup*; Jusczyk & Aslin, 1995), they can cope with *some* speaker variation in the realization of words from early on. That is, English-learning 7.5-month-olds show evidence of generalizing across speaker idiosyncrasies by listening longer to passages containing tokens of the trained words than to passages containing novel words, even when the passages in the test phase are produced by a different same-gendered speaker than the one who produced the words in the familiarization phase (e.g., Houston & Jusczyk, 2000). These initial studies thus confirmed the findings obtained from the using isolated vowels and syllables.

Follow-up studies, however, revealed that these early word recognition abilities also appear to be quite limited. If two realizations of a word are acoustically dissimilar, infants experience difficulty mapping them onto the same word representation. 7.5-month-olds presented with word tokens in a male voice, for example, subsequently do not recognize these word forms when they are produced in an acoustically distinct female voice (Houston & Jusczyk, 2000). Likewise, word forms spoken in a happy affect are not recognized when they are later produced in a neutral affect (Singh, Morgan, & White, 2004) and infants presented with word forms in Spanish-accented (Schmale & Seidl, 2009) or even Canadian-accented English (Schmale, Cristià, Seidl, & Johnson, 2010) do not appear to recognize these word forms in a Midwestern American English accent. These word form mapping studies suggest that unless infants are familiarized with a particular word form in a diversity of voices or affects (and will subsequently show broad generalization abilities across that domain; Houston, 1999; Singh, 2008; also see Rost & McMurray, 2009 for similar findings) infants initially only recognize words when they are produced in a fashion similar to how they had heard them before (see Houston &
Jusczyk, 2003; Singh, 2008 however, for findings that variability in the immediate context can also be informative). And even when semantic referents are available, 24-month-old English-learners fail to recognize words taught by an American speaker when this word is subsequently presented in a Spanish-accented voice (Schmale, Hollich, & Seidl, 2011).

2.3 The effect of speaker information on recognizing familiar words

The observation that infants can initially cope with only a limited amount of speaker-related variation in the realization of words does not hold only for the word form mapping task described above. Older children show a similar pattern of only being able to contend with a limited amount of variation when recognizing isolated words. North-American 15-month-olds, for example, listen longer to highly familiar words such as \textit{mommy} and \textit{ball} than to low-frequency words that were likely not part of their vocabulary. This preference for familiar words, however, is only observed when the stimuli are produced by a speaker of their own variant of English. When the unknown talker speaks an (unfamiliar) Jamaican variant of English, an accent very distinct from their own North-American variant of English, children do not display this listening preference for familiar words (Best, Tyler, Gooding, Orlando, & Quann, 2009). Similarly, 19.5-month-olds will look longer to the referent of a known word (as opposed to an unrelated referent) when it is produced in their own North American accent, but not when it is produced in an unfamiliar Jamaican accent (Mulak, Best, Irwin, & Tyler, 2008) and even children up to seven years of age continue having difficulty recognizing words in unfamiliar accents (Nathan, Wells, & Donlan, 1998). While surface variability in the
realization of words impedes speech processing in adult listeners as well, adults can still recognize words in unfamiliar accents, albeit with slight delays compared to native-accented speech (e.g., Floccia, Goslin, Girard, & Konopczynski, 2006). In contrast, the additional processing costs due to surface variability appear to fully disrupt word recognition abilities for younger listeners, allegedly rendering infants unable to access words in unfamiliar accents. Infants initially thus seem much less capable than adults in dealing with naturally occurring variability in the speech signal.

2.4 Implications for infant speech perception

How does this affect the nature of early word representations? In line with exemplar-based models of early word recognition such as WRAPSA (Jusczyk, 1993) and PRIMIR (Werker & Curtin, 2005), reports that infants initially experience difficulty coping with variability during spoken word recognition have been taken to suggest that infants’ early word representations are instance-based. According to this view, infants would store each individual word they segment ‘as is’, in the exact way they encountered it, including all indexical information induced by the characteristics of the speaker. If infants’ early word form representations are truly episodic in nature, acoustic deviations from previously heard word tokens would not easily be linked to these stored instances. The surface forms of the word *fool* produced by a male versus female speaker, for example, are very distinct. In the absence of abstract word representations in the infants’ mental lexicons, such word tokens are not immediately recognized as referring to the same underlying representation. Only once a sufficient number of phonetically diverse tokens of a new word have been encountered and (individually) stored, will a distinct word cluster...
emerge. Such word cluster would enable generalizations to take place, thereby allowing infants to more easily cope with indexical variation (e.g., Jusczyk, 1993). While this would explain infants’ apparent inability to deal with acoustic divergence, accepting the view that infants’ early word representations are exemplar-based raises important questions concerning the way infants’ episodic word representations would develop into more adult-like abstract representations allowing words to be recognized despite surface variation.

The literature on infants’ abilities to deal with variability in the realization of words stresses their apparent difficulty in generalizing across acoustically distinct tokens. As it does not seem implausible that infants simply cannot accommodate between-speaker differences, few studies have been conducted examining if and how they can tune in to speakers’ accents. As adults, however, we know from our own experience that we can recognize words in foreign accents. Any observed difficulties in recognizing words that diverge from our own native accent must thus be overcome and hence the focus of adult speech perception studies lies on examining how listeners accomplish this. This striking difference between the focus of the two fields (stressing the difficulty infants experience, but the solution adults have worked out) may lead us to overlook the possibility that infants may possess at least some adaptation abilities similar to those of adults. Infants’ observed difficulty coping with speaker-related variability may, in other words, not be a direct sign of early word representations being episodic in nature but may rather be a consequence of methodological limitations. Had infants been tested in paradigms more similar to those used with adults, infants may have been able to use similar strategies as adults for working out the between-speaker mappings. In order to
draw such parallels, it would thus be informative to explore the way adults cope with speaker-induced variability during word recognition.

3 Adult listeners’ strategies for handling speaker-related variability

Adult word recognition under everyday listening conditions is remarkably efficient. Although Americans watching the recent Sherlock TV series may initially be challenged when trying to understand the British-accented Sherlock Holmes, such difficulties understanding the accents of actors generally abate by the end of the first episode (cf. Mitterer & McQueen, 2009). Thus, while adults are initially affected by speaker variation, any difficulties they may experience can be rapidly overcome. What are the exact challenges adults face in the absence of speaker exposure? And, perhaps more importantly, how do they accomplish the task of effectively recognizing words despite these initial difficulties?

3.1 Consequences of speaker-related variability in adulthood

While not generally discussed in the infant literature, speaker information continues to affect speech processing into adulthood (Goldinger, 1996; Magnuson & Nusbaum, 2007; Martin, Mullennix, Pisoni, & Summers, 1989; Mullennix & Pisoni, 1990; Mullennix, Pisoni, & Martin, 1989; Palmeri, Goldinger, & Pisoni, 1993). The identification and recall of isolated spoken words, for example, have been found to be more reliable when the speaker remains constant as compared to when the speaker changes from trial to trial (Martin et al., 1989; Mullennix et al., 1989) and word recognition judgments are more accurate when words are repeated in the same as opposed to a different voice (Palmeri et
al., 1993). In addition, listeners seem unable to ignore speaker information, even when instructed to focus exclusively on phonetic details (Mullennix & Pisoni, 1990). The picture that emerges from these studies is that linguistic information and speaker information are interdependent and that variability in the presentation of speakers can impede word recognition.

3.2 Speaker accommodation as a solution to between-speaker variability

Given listeners’ reliance on speaker information when encoding the linguistic message of an utterance, it seems possible that listeners may exploit these voice cues to tune in to the talker they are listening to. Recent studies on this topic support this intuitive impression that experience with a particular speaker allows for subsequent improvement in the comprehension of that speaker (Bradlow & Bent, 2008; Clarke & Garrett, 2004; Dahan, Drucker, & Scarborough, 2008; Maye, Aslin, & Tanenhaus, 2008). That is, listeners become familiar with the specific characteristics of a speaker (or a group of speakers) and this knowledge subsequently enhances speaker comprehension. For this reason, understanding familiar speakers tends to be easier than understanding unfamiliar speakers (Bradlow & Pisoni, 1999; Nygaard, Sommers, & Pisoni, 1994; Nygaard & Pisoni, 1998; Smith & Hawkins, 2012). Speaker accommodation, or the perceptual learning of a speaker’s speech attributes, may thus be one mechanism that enables adult listeners to readily map acoustically distinct tokens onto the same underlying linguistic representation.

In studies examining the effect of speaker exposure, adult listeners are often first presented with a passage uttered by a speaker producing speech with particular
idiosyncrasies and later tested on their comprehension of words produced by that same speaker (as compared to pre-exposure comprehension or as compared to a control group; Greenspan, Nusbaum, & Pisoni, 1988; Maye et al., 2008). Alternatively, differences in listeners’ recognition of words produced by a particular speaker are examined at various points over the exposure period (e.g., Bradlow & Bent, 2008; Clarke & Garrett, 2004; Dupoux & Green, 1997). In these latter studies, speaker adaptation is evidenced by increasingly more efficient and enhanced performance to similar types of materials over the course of an experiment. Regardless of the method used, listeners generally tend to better understand speakers once they have had some experience with the way the speaker talks. This exposure does not need to be extensive; sometimes as little as one minute of speech can be sufficient for listeners to accommodate that speaker’s characteristics (Clarke & Garrett, 2004).

3.3  **Mechanisms underlying speaker accommodation**

How does adaptation to the phonetic differences between speakers take place? To examine the processes underlying accommodation, researchers have thus far focused primarily on the retuning of single phonemes (see, however, Reinisch and Weber (2011) for evidence of adaptation to suprasegmental stress information in foreign accents). In a landmark study on the perceptual learning of speech, for example, Norris and colleagues presented Dutch subjects with a speaker who either consistently produced word-final [f] halfway between [f] and [s] (i.e., [f/s] sound; e.g., in *naaldbos* ‘pine tree’) or who produced word-final [s] as an ambiguous [f/s] (e.g., in *witlof* ‘chicory’). Their results showed that listening to a speaker who consistently realized their [s] sounds ambiguously
between [f] and [s] caused a shift in the category boundary between [s] and [f] towards [s] whereas exposure to a speaker who consistently realized their [f] sounds as [f/s] resulted in a shift of the perception of the ambiguous sound towards [f]. Importantly, this effect was only observed when subjects were exposed to the ambiguous [f/s] sounds in real words, indicating that lexical feedback is a prerequisite for speaker adaptation (Norris, McQueen, & Cutler, 2003). The use of lexical information is thus crucial for the perceptual learning of speech.

The finding that listeners need lexical information for speaker adaptation has been replicated using a variety of different methodologies (e.g., Davis, Johnsrude, Hervais-Adelman, Taylor, & McGettigan, 2005; Kraljic & Samuel, 2005), but follow-up work has shown that listeners’ ability to adapt is somewhat position-dependent. That is, retuning occurs for ambiguous word-medial and word-final segments (Kraljic & Samuel, 2005; 2006; Norris et al., 2003), yet fails to surface for segments in word-initial positions (Jesse & McQueen, 2011). Speakers of different accents, however, typically diverge in the pronunciation of more than one phoneme (and natural accents differ not just in segments, but also in other linguistic levels such as intonation and rhythm). It would thus be interesting to determine whether adaptation remains robust when multiple phonemes are systematically altered, including or not including word-initial phonemes, either with native-accent intonation or with unfamiliar intonation. Presumably, however, under natural listening conditions, many differences between typical and atypical speakers hold for segments that can occur both word-initially and word-medially or word-finally. Even if the presence of a diverging word-initial segment does not induce adaptation, the
presence of the same segment in later positions of other words would allow the listener to adapt to the atypical realization of that segment.

3.4 **The scope of speaker accommodation**

3.4.1 **Generalizing beyond previously heard items**

The research discussed in the previous section used methodologies that presented listeners with ambiguous segments midway between two native phoneme categories. In order to fully adapt to speakers, however, it is not sufficient to only accommodate reasonably small, subphonemic deviations from the standard phoneme pronunciation. Sometimes deviations can cause a full phonemic shift, such as the realization of vowels between accents (Australian *ball* for example, can sound like *bowl* to Canadian listeners). Even in these cases, where the shifts are larger, listeners can, given the appropriate listening conditions, still accommodate the speaker’s or accent’s idiosyncratic way of producing a particular phoneme (Dahan et al., 2008; Maye et al., 2008). In addition, this accommodation involves a directional shift into the experienced direction. Specifically, after having heard a speaker who lowers their front vowels (e.g., from [ɛ] to [æ], such that *get* would be produced as *gat*), listeners interpret [æ]-sounds from that speaker to represent underlying [ɛ], in line with the direction of the shift. They do not, however, consider a raised front vowel (e.g., from [ɛ] to [I], *git* instead of *get*, which involves a shift into the opposite direction) produced by that speaker a good example of the original [ɛ]-vowel. Thus, at least as long as speakers are consistent in the direction of their vowel change, listeners err on the side of caution, making speaker adaptation very specific in its scope. Listeners have unconscious a priori expectations of what words from that
particular speaker should sound like and can use that information when they process speech (e.g., Dahan et al., 2008). Of course, it is possible that speakers who are inconsistent in how they realize their vowels (e.g., some raised vowels, some lowered, some tensed, some laxed) trigger less directional shifts.

In everyday life, speakers can produce more than just a limited number of words. Even after having heard the speaker talk for a little while, it is thus likely for listeners to encounter a word they had not heard the speaker say before. For speaker adaptation to be fully efficient, any consistent segmental shift in a speaker’s realization of words should generalize to the same segments present in unheard words. Speaker adaptation, in other words, would prove most beneficial when listeners do not simply store previously heard tokens, but abstract away from the words they heard and generalize to novel words that contain the shifted segment. Studies have found that this is indeed the case (Bradlow & Bent, 2008; Clarke & Garrett, 2004; Davis et al., 2005; Maye et al., 2008; McQueen, Cutler, & Norris, 2006). In studies inducing vowel shifts, for example, listeners would not only recognize the previously heard word *gat* as an instance of *get* but also allowed the word *tent*, which did not previously occur in the training phase, to be produced as *tant*. This reveals that word representations must at least be sufficiently abstract to allow for the incorporation of changes on a segmental (or subsegmental, Dahan et al., 2008) level.

3.4.2 Within- versus between-speaker generalization

A different form of generalization may also take place at the speaker level. Accents have been proven useful to study speaker adaptation and are often, but not always, used as a
tool to examine how listeners deal with speaker-related variability in the speech signal (see Dahan & Maed, 2010; Davis et al., 2005 for laboratory-induced noise vocoded distortions and Dupoux & Green, 1997 for time-compressed utterances). Although the type of segment changes discussed in the previous section allows for great experimental control, natural accent variation – somewhat more deviant from the home accent than single phoneme shifts – provides a better representation of the demands incurred by processing speech in everyday life.

Interestingly, natural accents are characterized by substantial overlap within groups of different speakers. That is, two Australian-accented English speakers generally have much more in common than an Australian- and a Scottish-accent speaker. This may lead one to wonder whether the accommodation processes we have seen above are speaker-specific or whether they are more broadly applied to a subsequent speaker. Bradlow and Bent (2008) addressed this question by presenting listeners with English sentences produced by Chinese-accented speakers. The subjects’ task was to transcribe the accented sentences. They found that experience with a single speaker, as was the case in the majority of the studies reviewed earlier, helps understand that specific speaker’s production patterns, but not another speaker’s production patterns, even if they share their language background and hence have comparable accents (Bradlow & Bent, 2008). This is not simply due to the fact that the accent used in that study was a nonnative one. Similar results are obtained in carefully controlled studies where exposure to one speaker’s ambiguous phonemes results in an adaptation for that specific speaker, but, at least when the shift takes place along a spectral dimension, does not generalize to another
One factor that does initiate generalization, however, is the presence of multiple speakers. In their study, Bradlow and Bent (2008) also manipulated the number of speakers listeners heard. Specifically, in addition to the low variability group, where listeners were exposed to a single accented speaker, subjects in the high variability group were presented with multiple different speakers. Such access to different Chinese-accented speakers is beneficial for understanding novel Chinese-accented speakers in the test phase (in fact, it is as beneficial as is prior access to the particular speaker used in the test phase) and hence allows for more speaker-general accent adaptation, indicating that the type of variability required for adaptation is dependent on the level of generalization. For non-general speaker-specific adaptation, within-speaker variability is sufficient. To generalize across speakers, in contrast, access to more between-speaker variability is necessary (see also Stacey & Summerfield, 2007 for similar results with spectrally shifted noise-vocoded speech and Rost & McMurray, 2009 for related findings with infants). Regardless of the variability in the number of different Chinese-accented speakers participants listened to, however, none of the exposure conditions induced a better understanding of a completely different (Slovakian) accent. These results nicely align with the findings that speaker adaptation is directional, accommodating only those discrepancies between novel and typical speakers that are likely useful. A specific, directional shift readily allows for better understanding of the speaker at hand all the while limiting the number of additional possibilities that would potentially arise if they simply relaxed the criteria and thus speaks to the efficient mappings listeners compute.
3.5 Implications for theories of adult speech perception

The recent findings that linguistic and speaker-related information are interdependent have lead to a pivotal shift in the way the nature of speech processing is viewed. Traditional accounts of speech perception had argued that listeners, before extracting linguistic meaning, first normalize the speech signal. That is, through a process called perceptual normalization, listeners must first discard linguistically irrelevant indexical information before the linguistic information can be interpreted (Van Lancker & Kreiman, 1987). These early models thus considered variability in the speech signal as “noise” that interferes with the processing of linguistic cues (e.g., Joos, 1948; Ladefoged & Broadbent, 1957; Shankweiler, Strange, & Verbrugge, 1977). Based on the research described above, however, speaker information is no longer considered to be irrelevant. Instead, more recent models of speech perception propose that speaker idiosyncrasies are retained in long-term memory. Exactly how this happens is, however, still subject of debate.

On the one hand, proponents of exemplar models (e.g., Goldinger, 1996; 1998) argue that this indexical information is stored in the mental lexicon. By considering early word forms to be episodic in nature, these models can explain listeners’ apparent reliance on speaker-specific information during linguistic encoding. Previously encountered words presented in the same voice and accent share great acoustic overlap and hence are readily recognized as being the same word. Previously encountered words that differ in speaker identity or accent, in contrast, are perceptually less similar and therefore pose a greater challenge for word recognition. For this reason, in word repetition judgment tasks, words are recognized faster as ‘previously heard’ when they are consistently produced by
the same speaker than when the speaker changes between the familiarization and test phase of the study (Palmeri et al., 1993).

Proponents of more abstractionist models, in contrast, argue that word representations are abstract rather than episodic. That is, the lexicon is thought to only consist of abstract phonological representations stripped off from lexically irrelevant information. Processes such as speaker adaptation do play an important role during speech encoding, but take place at a prelexical level. Once sounds have been categorized, segmental patterns are sent to the lexical level, and lexical items can be accessed. Successful word recognition is thus not dependent on the amount of acoustic overlap between word tokens, but on listeners’ opportunity to adapt to the speaker. Voice information, in other words, would not be token-specific, but rather abstract. This would explain why North-American listeners’ difficulty understanding Sherlock Holmes’ British accent, for example, decreases over time, even when the words he uses are ones we had not heard him say before. Findings showing that familiarity with the speaker not only facilitates subsequent recognition of presented words, but also the recognition of novel items by that speaker (Bradlow & Bent, 2008; Clarke & Garrett, 2004; Davis et al., 2005; Maye et al., 2008; McQueen et al., 2006) suggest that speaker adaptation must take place prelexically and hence support this latter view. That is, generalization effects would be unexpected to occur on an exemplar-based view, as the word cluster of unheard words is not likely affected by the pronunciation of unrelated words overlapping in just a single phoneme. Instead, episodic models storing word exemplars would be more consistent with an instance-specific manifestation of perceptual learning. More generalizable mappings thus appear to be better aligned with abstract prelexical accommodation.
3.6 Speaker adaptation in adults: concluding remarks

Taken together, the picture that emerges from the wide variety of studies looking at speaker and accent accommodation is that adult listeners’ signal-to-representation mapping skills are extremely flexible. Adults readily exploit speaker information, available from everyday naturalistic speech, to accommodate a wide variety of speaker-or accent-specific idiosyncrasies. Even as little as one minute of exposure to an accented speaker can be sufficient for adults to adapt to that speaker. Any extra processing costs induced by between-speaker variability are rapidly overcome by just a little experience with listening to the speaker.

4 Reconciling speech perception by infants and adults: Is there speaker accommodation in infancy?

As we have seen above, adults require only brief speaker exposure to adapt to a speaker’s voice or accent. Previous studies showing that phonetic dissimilarity impedes infants’ word recognition (Best et al., 2009; Houston & Jusczyk, 2000; Schmale & Seidl, 2009; Schmale et al., 2010; Singh, Morgan, & White, 2004), however, have overlooked the possibility that infants, too, could potentially benefit from brief speaker exposure. In these past studies, infants were presented with lists of isolated words by an unfamiliar speaker, either for recognition or for mapping to subsequently presented words in passages. This may have failed to provide young language users with the wide range of speaker information normally contained within the speech signal encountered in everyday listening conditions. The lack of speaker exposure in these studies may have prevented
infants from adapting to the unfamiliar speaker, much as exposure to only isolated words
hinders adult word recognition.

To date, little research has been devoted to further examine this issue. One exception, however, concerns a recent study testing North-American 19-month-olds’
word recognition when the speaker shifted their pronunciation of the vowel [a] towards
[æ] (White & Aslin, 2011). Words such as bottle and block, for example, would be
produced as battle and black. Crucially, prior to the test phase, one group of toddlers was
presented with words containing this vowel shift whereas another group of toddlers was
presented with the same words produced with the original vowel. At test, only those
infants presented with the novel pronunciations were found to recognize words with the
shifted vowels. Infants who heard the unshifted vowels in the familiarization phase, in
contrast, did not increase their looks toward the referent after the target had been named.
Exposure to the speaker’s realization of the vowel shift thus enhances toddlers’ later word
recognition. This holds both for words that were previously heard and for words that were
not previously encountered (by that speaker). Just like adults, then, 19-month-olds
generalize segmental information across heard exemplars and abstract away from the
precise items they have been exposed to. Moreover, experience with this vowel shift does
not cause listeners to consider vowels as completely unreliable. That is, while toddlers
were better at recognizing words that followed the shift they were trained on, words with
different vowel shifts (i.e., [a] to [ɛ] or [ɪ], such as bettle and bittle for bottle and bleck
and blick for block) in the test phase were largely unaffected (and hence not recognized).
Note, however, that words without vowel shifts were still recognized (both block and
black, for example, were judged to be appropriate labels for a block; see Maye et al.,
2008 for similar findings with adults), suggesting that the signal-to-segment maps of the speaker’s vowels had expanded to include both vowels, not just shifted from [a] to [æ], as would have been the case in full speaker accommodation (see Dahan et al., 2008 for findings in an adult eye tracking study showing that vowel shifts can cause words to be less efficient competitors, arguing for the possibility of more advanced adaptation in adulthood). This study presents the first evidence that children as young as at 19 months of age are able to better understand speakers when provided with some speaker experience. Follow-up work shows that such adaptation abilities remain consistently present throughout the elementary school time (McQueen, Tyler, & Cutler, in press),

How does this finding translate to speech perception in everyday life? In naturally occurring accent variations, differences between accents concern more than just a single phoneme shift. In addition, the majority of the speech infants and toddlers hear consists of multiword utterances (Aslin, 1993). Exposure to a single speaker producing only isolated words that contain a single vowel shift is thus not very likely to occur. The question hence arises whether naturalistic exposure to a speaker with an unfamiliar accent similarly causes adaptation. In other words, can children deal with more extreme phonetic mismatches when given appropriate everyday-like exposure to the speaker? If adaptation abilities such as those described above also surface for more distinctive natural accents, this would indicate that although children experience difficulty recognizing words in unfamiliar accents, they do possess the ability to contend with this type of speaker variability when the listening conditions are more ecologically valid. Such findings would explain the discrepancy between the previously described word recognition studies on the one hand and the classic speech perception studies on the other hand in which
infants were found to cope with between-speaker variability on syllables or phonemes (Dehaene-Lambertz & Pena, 2001; Jusczyk et al., 1992) even when those speakers differ in gender (Kuhl, 1979). In line with these classic studies, evidence for speaker adaptation in infancy might call into question the current view that infants’ early word representations are exemplar-based and would be consistent with the idea that that speaker information may be encoded prelexically from very early on and might never stored as part of the word representation itself. This could indicate that both infants and adults encode this information at least partially in an abstract manner (Cutler, 2008), suggesting that there is continuity in the development from infant to adult listeners.

5 Outline of this dissertation

In this dissertation, I will test the hypothesis that when provided with more ecologically valid listening conditions, even very young infants can contend with speaker-related variability. Given that adults, after only brief exposure to a speaker, readily adapt to speaker idiosyncrasies and that this information is normally available under everyday-like listening conditions, it is possible that providing infants with access to speaker characteristics would show mature word recognition abilities. As this type of speaker accommodation would take place prelexically, infants’ early word representations might accordingly contain an abstract component, just like adults’ word representations do (Cutler, Eisner, McQueen, & Norris, 2010). If the nature of infants’ early word representations is similar to that of adults, this would make the transition from the infant to the adult state much less abrupt. Rather than a qualitative switch in the representational format of words, infants’ word representations may, at least to some extent, be abstract
from early on. Words may then continue to be represented in an abstract fashion over time, with the fundamental difference between infants and adults being an increase in online efficacy to interpret speaker information and the application of this information when recognizing words. According to this view, speaker information may be viewed as a source of information used to adapt to different speakers. Rather than being noise that only adds to the processing demands of speech perception, speaker information may in fact assist in unraveling the many-to-one mapping in the speech signal.

In Chapter 2, I examine the issue of whether infants, like adults, accommodate the speech of accented speakers. Infants presented with isolated words produced in an unfamiliar accent, distinct from their own, have been found to experience difficulty recognizing those words, even though they have no trouble recognizing those same words in their own accent (Best et al., 2009). If this difficulty was caused by infants’ inability to determine what information in the speech produced by a speaker with an unfamiliar accent was indexical in nature and what was linguistic, then when provided with access to the speaker’s characteristics prior to test, infants should be better at identifying the underlying representation of a word. As a result, words may be recognized despite the fact that the unfamiliar accent is acoustically very distinct from their own accent. In Experiments 1 and 2, I first replicate previous findings showing that without any speaker exposure, 15-month-old infants only recognize known words produced by speakers that sound similar to those they encounter in their everyday life (i.e., Canadian-accented speakers). In contrast, no such recognition was observed in an acoustically distinct unfamiliar Australian accent until 22 months of age. Experiments 3 and 4 then address the question of whether prior exposure to the Australian-accented speaker would help 15-month-olds accommodate the
speaker and if so, what type of information infants might rely on to accommodate the unfamiliar accent.

The methodology used in Chapter 2 gives rise to all-or-nothing results: either infants recognize familiar words or they do not. The aim of Chapter 3 is to further test children’s use of speaker characteristics when understanding accented speech employing a more fine-grained measure of word recognition. Using an eye tracking methodology, this chapter reports two experiments testing 20- and 28-month-olds’ word recognition in an unfamiliar accent, after an initial two-minute exposure to the same or a different accent.

Chapters 2 and 3 examine whether exposure to a speaker can help young children contend with accent variation. Chapter 4 tests even younger infants’ ability to cope with a different source of variability: differences in the gender of the speaker. Infants at the earliest stages of word learning have been found unable to accommodate acoustic discrepancies in the realization of words (Houston & Jusczyk, 2000; Schmale & Seidl, 2009; Schmale et al., 2010). In these studies, however, words—produced by voice actors—were first presented to children in isolation. Much like studies testing infants’ abilities to recognize words in unfamiliar accents, these studies thus also fail to provide infants with access to the speakers’ idiosyncratic ways of realizing speech. Chapter 4 examines whether infants would be better able to deal with speaker-related variability in the realization of words when the speakers are highly familiar to the infants. Long-term exposure to the speakers’ voices may enable infants to use previously established signal-to-word mappings during speech processing and may consequently facilitate the recognition of word forms. In addition, this chapter addresses the question of whether in the absence of long-term voice familiarity, speaker variability can be handled when target words are presented using
more ecologically valid listening conditions (in fluent connected speech, rather than as isolated words, and produced by parents rather than by possibly less authentic voice actors). By reintroducing the more naturalistic properties of infant-directed speech in everyday life in the current task, I aim to see whether infants can capitalize on this information in everyday speech processing.

Taken together, as is described in more depth in Chapter 5, this dissertation will provide important novel insight into the way speaker-related variability is used in speech processing during infancy and toddlerhood. Specifically, I will demonstrate how infants, like adults, can deal with acoustic variability during word recognition in an elegant fashion when presented with appropriate listening conditions. By using listening conditions that do not resemble the listening conditions for speech perception in everyday life, previous infant studies might not have provided children with a fair opportunity to exhibit their true potential. When presented with more naturalistic stimuli, infants can readily contend with speaker-induced variation in the speech signal. Under these more ecologically valid listening conditions, infants behave similarly to adults in that they can map acoustically distinct versions of a word onto the same underlying representation. This suggests that word representations are sufficiently abstract to deal with acoustic variability from very early on and that the development of word recognition over time may be characterized by changes that are quantitative, but not qualitative in nature.
Chapter 2
The development of accent adaptation in infancy

1 Introduction

This chapter examines infants’ ability to understand known words produced by speakers with unfamiliar accents. In Chapter 1, I discussed in detail the widespread view in the infant speech perception literature that infants experience much more difficulty than adults in coping with variation in the realization of words (Best et al., 2009; Houston & Jusczyk, 2000; Schmale & Seidl, 2009; Schmale et al., 2010; Singh et al., 2004). For example, the available research suggests that children under a year of age struggle to map tokens of the same word produced in different accents or by speakers of different genders onto the same underlying word form representation (Schmale et al., 2010; Houston & Jusczyk, 2000). And while North-American 15-month-olds recognize highly familiar words in their own variant of English, they fail to do so when these words are produced in a Jamaican accent, indicating that unfamiliar accents pose serious difficulties for word recognition at this age (Best et al., 2009). The picture that has emerged from these studies is that infants are, at least initially, far less adept than adults at dealing with phonetic divergence in the realization of words. Accepting this view has serious implications for theoretical models of early language and speech development (Jusczyk, 1993; Werker & Curtin, 2005) and raises questions concerning how infants first learn to map acoustically distinct tokens of the same word onto the same underlying representation.

An alternative view, however, is that infants might be better able to cope with acoustic variation in the realization of words from early on. Support for this view comes
from studies showing that adult listeners require only brief speaker exposure to adapt to the speaker’s voice (Nygaard et al., 1994; Nygaard & Pisoni, 1998; Smith & Hawkins, 2012) or accent (Bradlow & Bent, 2008; Clarke & Garrett, 2004; Dahan et al., 2008; Eisner & McQueen, 2005; Kraljic & Samuel, 2005; Maye et al., 2008; Mitterer & McQueen, 2009, Norris et al., 2003). Previous studies showing that infants’ word recognition is impeded when presented with phonetic dissimilarity (e.g., Best et al., 2009), however, have not taken into account the possibility that that brief speaker exposure may be advantageous for infants as well. That is, infants in these studies only had access to list of lists of isolated word tokens. As isolated words lack the rich information normally contained within the speech signal encountered in everyday listening conditions, access to only a limited number of isolated words may have been insufficient for the infants to create reliable signal-to-word mappings for the speaker at hand. As a result, infants may not have been able to adapt to the unfamiliar accent, much as exposure to only isolated words hinders adult word recognition. Here, I examine the possibility that exposure to fluent Australian-accented speech enables Canadian English-learning infants to work out the mapping between their native accent and the unfamiliar Australian accent.

2 Experiment 1

In Experiment 1, I test Canadian English learning infants’ ability to recognize words in Australian-accented English. Australian English is a variant of English phonetically and prosodically distinct from North-American English (Wells, 1982). Compared to North-American English, for example, Australian vowels tend to be more raised and fronted. In
addition, [I] and [ə] are often merged in unstressed syllables, and intonation patterns can also be vastly different between the two variants. Would this dissimilarity prevent Canadian 15-month-olds to recognize words in the unfamiliar accent, just like Jamaican-accented words are not recognized by American 15-month-olds (Best et al., 2009)?

Using the Headturn Preference Procedure, as previously employed by Swingley (2005), infants were presented with lists of multiple known (e.g., bottle, cup) and lists of multiple nonsense words (e.g., shammy, bog). In Experiment 1a, infants listened to the word lists presented in their own variant of English (i.e., Canadian English), whereas infants in Experiment 1b heard these words produced in an unfamiliar Australian English accent. Previous studies using a similar design have shown that infants prefer listening to known over nonsense words when these words are produced in infants’ native accent (Best et al., 2009, Hallé & De Boysson-Bardies, 1994; Swingley, 2005; Vihman, Nakai, DePaolis, & Hallé, 2004). However, at 15 months of age, no such preference is observed when words are produced in an unfamiliar accent (Best et al., 2009). I thus predict that in this study only infants in Experiment 1a (and not 1b) will listen longer to known than to nonsense words.

2.1 Methods

2.1.1 Participants

A total of 32 normally developing English-learning 14.5- to 15.5-month old infants from the Greater Toronto Area were tested, 16 in each of Experiments 1a and 1b (age range: 447-477 days; 17 boys). None of the infants had any reported hearing issues or recent ear infections and infants in Experiment 1b had not had any substantial exposure to
Australian-accented English, as established by a language questionnaire (see Appendix A for the questions parents in both Chapter 2 and 3 were asked to answer) at the end of the lab visit. An additional 4 infants (2 in each of Experiment 1a and 1b) were tested, but excluded from the analyses due to extreme fussiness. All participating infants in this and subsequent experiments received a certificate and a small gift.

2.1.2 Stimuli

Infants were presented with eight test trials. Half of these trials (i.e., the known word trials) consisted of lists of words generally known by infants at 15 months of age (daddy, bottle, diaper, mommy, grandma, kitty, ball, dog, bath, kiss, cup, shoe), as evidenced by an average word understanding of 90.1% (range 68.7%-100%) by 15-month-olds in the Lexical Development Norms for English (Dale & Fenson, 1996). The other half of the trials, referred to as the nonsense word trials (reflecting their status in the infants’ lexicon), contained lists of nonsense or low-frequency words expected to be unknown to the 15-month-olds (koddy, dimma, dapper, mitty, guttle, shammy, bog, bap, deuce, kie, koth, brall). In order to exclude potential biases due to preferences for specific speech sounds, the two types of trials were matched in phonemes. Each trial contained all twelve known or nonsense words, repeated twice for a total of 24 words. Monosyllabic and bisyllabic words were alternated and the position of each word varied across trials (see Appendix B for all word lists). For Experiment 1a, words were produced by a native female English speaker from the Greater Toronto Area. The word lists in Experiment 1b were produced by a native female English speaker from Sydney, Australia. Speakers were instructed to speak in an infant-directed fashion. For each speaker, known and
nonsense words were matched in terms of word length and average pitch level of the stressed vowel. The Canadian speaker’s known words were on average 559 ms long and the nonsense words 579 ms. The average pitch of the vowel was 358 Hz for the known words and 378 Hz for the nonsense words. For the Australian speaker, the average known word length was 516 ms and the average nonsense word length 518 ms. The average vowel pitch was 301 Hz for known words and 292 Hz for nonsense words. Words were equated for loudness and were presented with approximately 800 ms inter-word intervals. All trials lasted 34.5 s.

2.1.3 Procedure

Participants were individually tested using the Headturn Preference Procedure (Kemler Nelson et al., 1995). At test, infants were seated on their caregivers’ lap in the centre of a dimly-lit double walled sound-attenuated Industrial Acoustics Company (IAC) test booth. A red light was mounted at eye level on the panel in front of the infant. Each of the side panels held a blue light, with loudspeakers positioned directly underneath. First, the centre light flashed. Once the infants oriented towards this light, the experimenter, who monitored the infants’ behavior via a TV screen outside the booth, pressed a button to terminate the centre light. This automatically initiated the flashing of one of the sidelights. The presentation of the word lists was contingent upon the infants’ looking behavior; lists started playing once the infant oriented towards the sidelight and were presented until the infant looked away for two consecutive seconds or until the maximum trial length of 34.5 s was reached. Infants’ orientation time to known and unknown words was measured. The eight test trials were presented in random order, with the restriction
that there could be no more than two known or two unknown word lists in a row. To avoid biases, parents were naïve to the experimental predictions and listened to masking music over closed headphones throughout the whole experiment. The experiment lasted approximately 2-3 minutes.

2.2 Results and Discussion

In line with past studies, Figure 1 shows that infants only listened longer to words than to nonsense words in Experiment 1a, when presented with a speaker of their own accent, but not in Experiment 1b, where the speaker spoke in an unfamiliar accent. Specifically, all infants in Experiment 1a listened to longer to lists of known words (on average 11.55 s) than to nonsense word lists (on average 6.29 s). Infants tested in Experiment 1b listened to known words for 9.39 s on average and to nonsense words for 10.06 s, with 6 out of 16 infants listening longer to the known words. A 2 x 2 ANOVA with word status (known vs. nonsense word) as a within-participant factor and accent familiarity (native vs. unfamiliar accent) as a between-participant factor revealed a main effect of word status ($F(1,30) = 7.039; p = 0.013$), indicating that infants preferred to listen to known over nonsense words. Crucially, this main effect was modulated by the interaction between word status and accent familiarity ($F(1,30) = 11.698; p = 0.002$). That is, only infants presented with a Canadian speaker preferred to listen to known over nonsense words ($t(15) = 4.224; p = .001$ in Experiment 1a; $t(15) = -.552; p = .589$ in Experiment 1b).

Thus, while infants at 15 months of age recognize words in their own accent, the phonetic dissimilarity of an unfamiliar accent prevents them from recognizing isolated words in unfamiliar accents. These findings conceptually replicate previous work examining
Figure 1: Orientation time in seconds (error bars indicate standard errors of the mean difference scores) for 15-month-olds to known and nonsense words in Experiments 1a (native accent) and 1b (unfamiliar accent).

American infants’ understanding of Jamaican-accented English (Best et al., 2009) with a different population and different accent.

3 Experiment 2

Without any prior exposure to the speaker, 15-month-olds in Experiment 1 fail to recognize words in a variant of English distinct from their own native accent. While adults do experience additional demands when exposed to accented speakers, word recognition is still above chance (Bradlow & Bent, 2008). This raises the question of how and when infants develop more proficient word recognition abilities. Previous studies using a different population and a different accent (Jamaican English) have suggested that such competences develop between 15 and 19 months of age (Best et al., 2009). Of course, different accents vary in how distinct they are from the infants’ target accent and
it is possible that Jamaican-accented English is more or less distinct from Connecticut English than Australian-accented English is from Canadian English. In order to verify that the accent and stimuli used in the current study gives rise to a similar developmental pattern and to more precisely assess when infants are able to overcome these phonetic differences, Experiment 2 examines the developmental trajectory of infants’ word recognition in Australian-accented English using the same materials as in Experiment 1b. If the ability to recognize words in Australian English in the absence of speaker information develops as a function of linguistic maturity, the older the infants are, the greater their preference may be to listen to known over nonsense words.

3.1 Methods

3.1.1 Participants

An additional 32 normally developing English-learning infants from the Greater Toronto Area were tested. Sixteen of these infants were between 17 and 18 months of age (age range: 523-548 days; 12 boys). The remaining 16 infants were between 21.5 and 22.5 months of age (age range: 662-686 days; 6 boys). As was the case in Experiment 1b, none of the infants had had any substantial exposure to Australian-accented English and all were free of known hearing issues or recent ear infections. One additional 17.5-month-old and three additional 22-month-olds were tested, but were excluded from the analyses due to extreme fussiness.

3.1.2 Stimuli and Procedure

The stimuli and procedure are identical to those in Experiment 1b.
3.2 Results and Discussion

To examine how Canadian English-learning infants’ ability to recognize words in Australian-accented develops over time, infants’ performance on this task was compared across the three age groups (including the 15-month-olds in Experiment 1b). The differences in orientation time between known and nonsense words were submitted to a one-way ANOVA with age testing for potential trends. The difference scores were found to follow a linear trend ($F(1,45) = 4.639; p = .037$; see Figure 2). There was no effect for the quadratic term ($F(1,45) = 0.40; p = .843$). This demonstrates that with time, infants gradually start preferring to listen to known over nonsense words, with paired samples t-tests indicating that this difference reaches significance at 22 months of age ($t(15) = 2.663; p = .018$ for 22-month-olds; $t(15) = .281; p = .783$ for 17.5-month-olds; see Figure 3). Similar to previous studies (Best et al., 2009), children gain the ability to cope with unfamiliar accents when recognizing words sometime between 15 months of age and

![Figure 2: Orientation time difference between known and nonsense words (produced in Australian-accented English) in seconds (error bars indicate standard errors of the mean difference scores) in Experiments 1b (15-month-olds) and 2 (17.5-month-olds and 22-month-olds, respectively).](image-url)
their second birthday. The finding that the words in the Australian accent can be recognized by just slightly older Canadian English-learning infants than those in Experiment 1 shows that Australian English can be understood by young children. Would exposure to the characteristics of the unfamiliar Australian accent enable the younger group of infants to similarly recognize words in the Australian accent?

4   Experiment 3

Experiments 1 and 2 reveal that, without prior exposure to the speaker, infants are not able to recognize words in an unfamiliar variant of English until the later half of their second year of life. From the adult literature we know, however, that only brief experience with a speaker can help listeners accommodate that speaker’s accent (Bradlow
& Bent, 2008; Clarke & Garrett, 2004; Dahan et al., 2008; Eisner & McQueen, 2005; Kraljic & Samuel, 2005; Maye et al., 2008; Mitterer & McQueen, 2009, Norris et al., 2003). As the twelve isolated words in the previous experiments do not provide infants with a large amount of exposure to the characteristics of the accent, it is possible that this limited availability of accent idiosyncrasies is insufficient for younger infants to adjust to the accent. This raises the possibility that infants, too, might be better able to cope with unfamiliar accents after having benefited from speaker exposure. That is, prior access to the characteristics of an accent may help infants work out the inter-accent signal-to-word maps, which they can later use to recognize the test items. Thus, in Experiment 3, I seek to find out whether Canadian 15-month-olds, who experience difficulty recognizing words in Australian English, can benefit from speaker exposure and adapt to the speaker’s accent, thereby overcoming their inability to recognize words in an unfamiliar accent. In order to examine this, Canadian-English-learning 15-month-olds were tested on the same Australian-accented test stimuli used in Experiment 1b. This time, however, the test phase was preceded by a two-minute recorded version of ‘The Very Hungry Caterpillar’ storybook (Carle, 1994), read by the same Australian speaker who had recorded the test items. Crucially, the story did not contain any of the words used in the test phase. In order to succeed at recognizing the known words in the test phase, infants would thus not only need to learn the abstract mapping between the pronunciation of words in their native Canadian accent and the unfamiliar Australian accent during the exposure phase, but they would also need to apply this abstract knowledge when recognizing known words that they had never heard spoken by the Australian speaker. If
the exposure video indeed induces such form of accent adaptation, infants should prefer to listen to the known word list over the nonsense word list in the test phase.

4.1 Methods

4.1.1 Participants
Sixteen normally developing English-learning 14.5- to 15.5-month old infants from the Greater Toronto Area were tested (age range: 449-473 days; 10 boys). As in Experiments 1b and 2, infants were free of any known hearing issues or recent ear infections and none of the infants had had any substantial exposure to Australian-accented English. An additional 4 infants were tested, but were excluded from the analyses due to extreme fussiness.

4.1.2 Stimuli
Test phase stimuli were identical to those in Experiment 1b and 2. Prior to test, however, infants were presented with a video recording of The Very Hungry Caterpillar story read in an infant-directed fashion by the same Australian speaker who also recorded the isolated words. The two-minute exposure video displayed a close-up of the speaker from the chest up against a green background.

4.1.3 Procedure
The procedure was identical to that in Experiment 1 and 2, besides the addition of an exposure phase. During the exposure phase, infants sat on their caregivers’ lap and
watched the two-minute video on a large TV screen. The movie played continuously until the end, after which the test phase started.

4.2 Results and Discussion

Contrary to my predictions, the results of this experiment indicate that 15-month-olds failed to adapt to the Australian speaker’s accent after hearing the story (see Figure 4). That is, infants listened to lists with known words for 8.93 s and to lists with novel words for 8.09 s, with 8 out of 16 infants listening longer to the known words. These orientation times are not significantly different from one another ($t(15) = .627; p = .540$).

Experiment 3 shows that the two minutes of exposure to the Very Hungry Caterpillar story did not enable infants to tune into the abstract accent-specific characteristics, at least not to an extent that it would help them recognize previously unheard words. There

![Figure 4](image)

*Figure 4:* Orientation time in seconds (error bars indicate standard errors of the mean difference scores) for 15-month-olds to known and nonsense words in Experiment 3 (unfamiliar accent with exposure).
are at least two possible explanations for why this may be the case. One possibility is that speaker exposure may not be beneficial at this age. Perhaps, in line with previous reports (Best et al., 2009; Jusczyk, 1993; Werker & Curtin, 2005), infants’ early word representations may lack the abstract components needed to accommodate unfamiliar accents and accent exposure may only be used to work out the signal-to-word maps once infants have developed more abstract representations. If this were the case, this would imply a discontinuity between the word representations early in life and those later in life, leaving one wondering how these early episodic representations would develop into more mature representations that carry abstract components (Cutler, 2008).

Alternatively, exposure to the Very Hungry Caterpillar story may not have induced adaptation because the story contains many key words 15-month-olds are unlikely to know (e.g., *caterpillar, sausage, cocoon*). Both adult listeners and school-aged children use lexical feedback for the perceptual learning of a specific speaker’s pronunciation of sounds (Eisner & McQueen, 2005; McQueen et al., in press; Mitterer & McQueen, 2009; Norris et al., 2003), so if access to words is a prerequisite for accent adaptation even in infancy, infants’ inability to accommodate the Australian accent may be due to an inability to access the words in the exposure phase. Not recognizing words in the exposure phase, in other words, would be comparable to adults’ inability to recognize nonsense words and would prevent tuning into an unfamiliar accent (Norris et al., 2003).\(^1\) Accent adaptation may thus only be only educed once infants recognize a sufficient

---

\(^1\) Note that although 15-month-olds were likely familiar with the function words in the story (e.g., Shady, 1996; Shafer, Shucard, Shucard, & Gerken, 1998; Shi, Werker, & Cutler, 2006), the reduced nature of the vowels in this category of words may not have provided them with ample information regarding the between-accent vowel mappings.
number of content words spoken in the unfamiliar accent. Experiment 4 examines this possibility.

5 Experiment 4

In Experiment 4, I test the possibility that it was infants’ inability to recognize words in the exposure phase of Experiment 3 that prevented them from adaptation. If this were the case, then increased familiarity with the words in the story, and hence more robust representations of the word forms associated with them, may allow infants to access the story words in the exposure phase, even when produced in Australian-accented English. Accessing the words in the exposure phase may in turn allow infants to use the (abstract) phonological code needed to guide accent accommodation. In order to accomplish this and yet test the same population under the same listening conditions, infants would essentially have to gain greater knowledge of the words in the story. Given that storybook reading can prompt word learning in children (Horst et al., 2011), Canadian-English-learning infants were familiarized with the Very Hungry Caterpillar story prior to their visit to the lab. Parents were asked to read the book at home to their child once a day for the two weeks prior to their lab visit. During their visit, at which point infants had heard the story read to them a minimum of 14 times in their own Canadian English accent, infants were exposed to the exact same Australian-accented exposure video and test trials from Experiment 3. If speaker exposure enables infants to work out the mapping between their native accent and the unfamiliar accent, but only when they have access to lexical information, then being familiar with the story should cause infants to listen longer to the known word list than the nonsense word list in the test phase.
5.1 Methods

5.1.1 Participants

Sixteen normally developing English-learning 14.5- to 15.5-month old infants from the Greater Toronto Area were tested in Experiment 4 (age range: 448-469 days; 7 boys). As in Experiments 1b, 2, and 3, all infants were free of any known hearing issues or recent ear infections and none of the infants had had any substantial exposure to Australian-accented English. An additional 3 infants were tested, but excluded from the analyses due to extreme fussiness. To ensure that parents read the storybook to their infants, they were asked to complete a diary, which inquired about the time of day the story was read and the person who read the story. The data from an additional two infants were replaced due to the parents’ failure to read the book at home once a day for two weeks. According to the diaries, all infants in the final sample listened to the story at least 14 times before participating in the story. Reading was mostly carried out by direct family members.

5.1.2 Stimuli and Procedure

Stimuli and procedure were identical to Experiment 3.

5.2 Results and Discussion

Figure 5 presents an overview of the three experiments with 15-month-olds presented in this study. As can be seen from the right panel of this figure, infants in Experiment 4 preferred to listen to the known words over the nonsense words. Infants listened to known words for 15.23 s and to nonsense words for 8.90 s, with 14 out of 16 infants listening longer to the known words. This difference is statistically significant ($t(15) = 3.616; p =$
.002), indicating that infants do adapt to the Australian speaker’s accent after hearing the familiar story.

In order to verify that infants in Experiment 4 did behave differently from those in Experiment 3, a mixed model ANOVA with word status as a within-participant factor and exposure story familiarity as a between-participant factor was conducted. This yielded a main effect of word status ($F(1,30) = 10.624; p = 0.003$). Importantly, however, this effect was qualified by an interaction with familiarity to the storybook ($F(1,30) = 6.212; p = 0.018$), demonstrating that only infants familiarized with the story at home preferred to listen to known over nonsense word lists.

---

*Figure 5:* Orientation time in seconds (error bars indicate standard errors of the mean difference scores) for 15-month-olds to known and nonsense words in Experiments 1a (native accent), 1b (unfamiliar accent), 3 (unfamiliar accent with exposure), and 4 (unfamiliar accent with book familiarization and exposure).
6 Discussion of Experiments 1-4

In order to efficiently communicate with the people around us, it is crucial to recognize words despite variation in speakers or accents. Although young infants experience difficulty recognizing familiar words produced in an unfamiliar Australian accent until they near their second birthday, 15-month-olds readily overcome this inability after brief experience with the accent. With only two minutes of exposure to the speaker, infants can accommodate the speaker’s Australian accent, as observed by infants’ subsequent preference for known over nonsense words. Crucially, infants recognize words spoken in an unfamiliar accent in the test phase even though none of those words occurred in the exposure story, indicating that the adaptation to the speaker’s word realizations in the exposure phase generalized to previously unheard words (in that accent). A speaker’s typical production pattern may thus establish speaker-dependent prelexical signal-to-sound mapping strategies, suggesting that infants’ early word representations are suitably abstract to deal with acoustic variability due to between-accent differences. These results nicely align with classic findings from speech sound discrimination studies showing that infants can, to some extent, cope with substantial indexical variation in the realization of phonemes or syllables (Dehaene-Lambertz & Pena, 2001; Kuhl, 1979; 1984). Taken together, these findings support the notion that infants store abstract linguistic representations in their mental lexicon.

Speaker exposure only assisted word recognition when infants had regularly heard the story in the two weeks preceding the lab visit. There are at least two separate, albeit closely related explanations for this finding. First, it is possible that without recent routine exposure to the story, the number of words infants knew (and accessed) in the exposure
phase was insufficient to allow for lexically-guided accent adaptation. The unfamiliar words in the story would essentially be treated the way adult listeners treat nonsense words and would hence fail to prompt adaptation. Infants who were read the storybook for a minimum of 14 times, in contrast, presumably became familiar with many of the previously unfamiliar words in the story. The knowledge of these words may have allowed more words to be accessed during the exposure phase, which in turn may have induced adaptation to the unfamiliar accent. Alternatively, encoding a familiar story may be less demanding than encoding a less familiar story, such that more processing resources may have been available to recognize and adjust to deviant pronunciations of words. For example, it is possible that recognition of a story activates words associated with that story. These increased activation levels would arguably facilitate word recognition, such that words can be recognized despite the acoustic divergence in surface forms. While this study does not distinguish between these two possibilities, both explanations have as their basic premise that adaptation is facilitated by accessing words during the exposure phase. The idea of speaker accommodation being lexically mediated is consistent with adult work showing similar effects (Eisner & McQueen, 2005; Mitterer & McQueen, 2009; Norris et al., 2003). The finding that infants, who are just beginning to build their lexicon, may also use this lexical knowledge demonstrates the importance of lexically-guided adaptation strategies across the life span.

In Experiments 3 and 4, the exposure story and the test items were produced by the same Australian-accented speaker. This means that infants’ ability to recognize the Australian-accented words in Experiment 4 could have been the result of either speaker adaptation or accent adaptation. Specifically, appropriate exposure to one Australian
speaker induced a better understanding of that accented speaker, but it is unsure whether it would also generalize to a different Australian speaker. The adult literature suggests that adaptation to one speaker does not always transfer to another speaker of the same accent and that only exposure to multiple different speakers of the same accent induces full speaker-independent accent adaptation (Bradlow & Bent, 2008). Whether infants would follow a similar generalization pattern is an empirical question and will be discussed in more detail in the General Discussion in Chapter 5. Either way, infants’ ability to overcome difficulty understanding speakers in unfamiliar accents after only two minutes of exposure suggests that in everyday life, they might be able to understand accented speakers of their own language, regardless of whether they have had prior exposure to their accent. This makes speech perception in infancy remarkably efficient.

Mapping physically variable signals onto abstract linguistic representations is arguably one of the most impressive cognitive feats accomplished by humans. This study has shown that speaker exposure allows even infants, who are only at the initial stages of learning to speak, to adapt to unfamiliar accents and hence work out sophisticated speaker-dependent signal-to-word maps. This suggests that word representations are sufficiently abstract to deal with acoustic variability from very early on and that the transformation of word recognition from infancy into adulthood may be a gradual development that is solely quantitative in nature.
Chapter 3
Toddlers’ recognition of words spoken in unfamiliar accents

1 Introduction

Chapter 2 shows that, when provided with appropriate listening conditions, Canadian 15-month-olds can accommodate to Australian-accented English. This is, to my knowledge, the first study to show that infants this young can adapt to speakers of a naturally occurring unfamiliar accent, enabling them to recognize words in that accent. These findings give rise to a large body of further questions regarding the exact nature of these adaptation processes. How much exposure is necessary? How many words need to be recognized in order to adapt? Is exposure to more words always beneficial when dealing with an unfamiliar accent or does the advantage of prior exposure level off towards an asymptote at some point? Does the adaptation extend to other speakers of the same accent? Does it generalize to speakers of a closely related accent? How quickly do children recognize the accented words? Does speaker tuning change over the course of development? And to what extent is recognition of accented words as robust as the recognition of native accented words, especially at the early stages of speech perception?

This chapter examines these last two questions. In Chapter 2, word recognition was measured by means of orientation times to trials containing known or nonsense words. While a preference for known over nonsense words suggests that infants do recognize the words in the unfamiliar accent, it does not reveal whether these words were also accessed as efficiently as native-accented words. Speaker accommodation in adulthood is known to lead to rapid activation of lexical items in subsequent speech
produced by the speaker. In fact, under some conditions, brief speaker exposure induces speaker adaptation such that any difficulties adults experience when recognizing words in unfamiliar accents are fully overcome (Clarke & Garrett, 2004). Does the same hold for infants, whose word recognition abilities are plausibly somewhat less efficient than those of adults? One way in which the degree of recognition accuracy could be assessed is by employing an online measure of lexical activation.

This chapter implements an eye tracking technique to examine the accuracy of word access in an unfamiliar accent. Using the Preferential Looking Paradigm (as described by Fernald, Zangl, Portillo, & Marchman, 2008; Johnson & Zamuner, 2010; Swingley, 2012), I measure toddlers’ eye movements as they listen to an instruction to look at one of two depicted objects on a TV screen. Based on the evidence available, there is reason to believe that recognizing words in unfamiliar accents remains a non-trivial task even beyond infancy. Specifically, previous work similarly making use of a preferential looking procedure has shown that without any prior exposure to the speaker, toddlers experience difficulty mapping accented words onto their underlying linguistic representations until around 30 months of age (Schmale et al., 2011). However, children in this study were taught new words whose representations might arguably be somewhat less robust than the representations of frequently heard familiar words. Although caution should be exercised when directly comparing results across paradigms (especially when different accents are used), this relatively unstable representation may be one of the reasons why toddlers failed to recognize the accented words in that study, even though considerably younger children at 19 and 22 months of age are able to recognize words in an unfamiliar Jamaican (Best et al., 2009) and Australian accent (Chapter 2),
respectively. Here I test to see whether toddlers are able to recognize accented words in a preferential looking task when these words are highly familiar. Moreover, I examine whether the accuracy of accessing lexical items depends on the speaker’s accent and listeners’ exposure to the speech characteristics of the accented speaker.

2 Experiment 1

Experiment 1 tests Canadian English learning 28-month-old toddlers’ ability to recognize words produced in an unfamiliar accent. Without any speaker exposure, toddlers do not start to recognize words taught in their own accent when those words were subsequently produced in a foreign accent until sometime in their third year of life (Schmale et al., 2011). For this reason, testing 28-month-olds provides an excellent means to examine whether exposure to a speaker enables toddlers to recognize words more efficiently, similar to adults requiring only brief speaker exposure to accommodate an accented voice.

Toddlers were presented with recorded instructions to look at one of two depicted target items. The target words were selected to be commonly known to toddlers at this age. In Experiment 1a, the test phase containing these instructions was preceded by an exposure phase in which the toddlers listened to a recording of the same speaker who also produced the instructions for the test phase. As in Experiment 3 and 4 in Chapter 2, the Very Hungry Caterpillar story was used to provide children with access to the characteristics of the accent. In Experiment 1b, in contrast, the exposure phase preceding the test phase presented the toddlers with a speaker who spoke a different unfamiliar accent. If prior exposure to the speaker of the same accent used in the test phase helps
toddler adapt (and if, unlike the 15-month-olds in Chapter 2, toddlers have a good grasp of the words used in the story), toddlers in Experiment 1a should recognize test items more reliably than those in Experiment 1b. I thus predict that toddlers, upon hearing the target words in the test phase, are more likely to shift their gaze towards the target picture in Experiment 1a, where toddlers may have adapted to the test phase speaker during the exposure phase, than in Experiment 1b, where any adaptation to the speaker during story presentation renders irrelevant for recognizing words in the test phase by a different speaker of a different accent.

By including words that occurred in the story as well as words that were never heard before in that speaker’s accent, I also examine whether previously heard words are recognized more easily or whether the adaptation process is fully abstracted away from the individual words. If adaptation is restricted to the particular words heard in the story, the words that occurred in the storybook should be recognized more efficiently than the new words. If, however, the adaptation process is sufficiently abstract, no differences should be observed between the two types of words.

2.1 Methods

2.1.1 Participants

A total of 32 normally developing English-learning 27- to 29-month-old toddlers from the Greater Toronto Area were tested (age range: 824-883 days; 14 boys). Half of the toddlers participated in Experiment 1a and the other half participated in Experiment 1b. All infants were free of any known hearing issues or recent ear infections. In addition, as established by a language questionnaire at the end of the lab visit (see Appendix A), none
of the toddlers had had any substantial exposure to Australian- or Scottish-accented English, the accents used in this study. An additional 5 toddlers were tested, but excluded from the analyses due to extreme fussiness or failure to complete the study.

2.1.2 Stimuli

A total of 16 nouns (ball, boat, book, butterfly, cake, car, cat, cheese, cup, cow, dog, duck, fork, soup, strawberry, and toast) were selected to be used as target words in the test phase this study. These nouns are generally known by children at 28 months of age, as evidenced by an average word production rate of 89.9% (range 69%-98.3%) by 28-month-olds in the Lexical Development Norms for English (Dale & Fenson, 1996). For each noun, an image was selected to represent the word. These images were matched for approximate size and interest (see Appendix C for all images). A questionnaire administered to the parents at the end of their visit asking about their judgment of their toddlers’ recognition of each of the pictures in the study indicates that the pictures clearly depicted the appropriate nouns (the average picture was reportedly recognized as depicting the target item by 95.1% of the toddlers).

Nouns were embedded in each of two carrier sentences. One sentence was an imperative (Look at the [noun]!) and the other one a question (Where’s the [noun]?). Nouns always occurred in sentence-final position. These sentences were recorded by a native female English speaker from Scotland, UK in a child-directed fashion. Target words lasted on average 759 ms. In addition, tokens of auditory attention attractors (aww, hey, look, wow) were recorded to be used direct toddlers’ attention to the screen prior to
sentence onset. To increase toddlers’ interest in the task, the speaker also recorded positive statements about the stimuli (e.g., Fantastic, eh? or How cute!).

The Very Hungry Caterpillar story presented to the toddlers prior to test was read by the same Scottish speaker for toddlers in Experiment 1a and by an Australian speaker (the same speaker who had also recorded the stimuli in Chapter 2) for toddlers in Experiment 1b. Unlike the video recording in Chapter 2, this recording did not display the speaker. Instead, it displayed the illustrations from the storybook (with the text erased) corresponding to the accompanying part of the story. The exposure video lasted approximately 2 minutes and 21 seconds.

2.1.3 Design

During the test trials, the movie presented two pictures (a target and a distracter) side-by-side against a white background for a total of seven seconds. To maintain the toddlers’ interest in the video, both objects gradually increased and decreased in size. Approximately 2 to 2.5 s after the appearance of the two objects, toddlers were instructed to look at one of those two pictures. Target words occurred exactly 3 s after picture onset. The instruction was preceded by an attention-getting statement and followed by a positive lead-out sentence, commenting about the target item (see Figure 6 for a trial outline).

Within each of Experiments 1a and 1b, toddlers were randomly assigned to one of four video orders (see Appendix D). Each video consisted of 16 test trials. In the four heard word trials, the two pictures on the screen referred to words present in the Very Hungry Caterpillar story. In the twelve generalization trials, neither of the two depicted objects represented words previously mentioned by the speaker.
Each of the 16 nouns was presented as target once, such that each combination of pictures appeared twice during the study. The order of a picture being target or distracter as well as the position of the picture on the left or right of the screen was counterbalanced across videos. Within each video, targets occurred equally often on left and right.

2.1.4 Procedure

Toddlers were individually tested using the Preferential Looking Paradigm. They were seated on their caregivers’ lap approximately 1 m away from a Sony LDC TV screen in a double walled sound-attenuated IAC test booth. Once the toddler oriented towards this screen, the experimenter started the video. Sessions were videotaped by a camera below the screen for subsequent offline coding of gaze position. To avoid biases, parents were naïve to the experimental predictions and listened to masking music over closed headphones throughout the whole experiment. The experiment lasted approximately 5 minutes. After their toddler had watched the video, parents completed a questionnaire regarding the toddlers’ comprehension of the target words and their recognition of the pictures used in the study.
2.1.5  **Off-line coding**

Sessions were imported for off-line frame-by-frame coding using SuperCoder (http://hincapie.psych.purdue.edu/Splitscreen/home.html). For each 33-ms frame, the gaze position of the toddler was judged to be a look towards the left, right or neither picture. The coder was blind to both the audio and the video components of the trials. Four sessions were randomly selected to be recoded by a second coder. The agreement on individual fixation durations was consistently high between the two coders (mean correlation = .996).

2.1.6  **Data analysis**

Following previous studies employing a similar procedure (e.g., Johnson, McQueen, & Huettig, 2011; Swingley, 2007; Van Heugten & Johnson, 2011; Zangl & Fernald, 2007), the proportion of fixations to the target picture was used as the dependent variable. This proportion was calculated by dividing the fixations to the target by the sum of the fixations to the target as well as the distracter. Thus, as only fixations toward one of the two pictures were taken into account (and looks away from the screen or shifts between pictures did not affect this ratio), a value of .5 suggests that the two pictures are fixated equally. Values greater than .5 are indicative of a looking preference toward the target picture. All data were checked for normality.

Recall that picture pairs were counterbalanced within videos, such that each picture appeared twice: once as the target and once as the distracter. This should prevent any inherent biases for either the target or the distracter picture. Nonetheless, in order to ensure that no such bias was present, the target fixation proportion was first compared to
chance level (.5) during the one second interval before target onset. To examine whether toddlers recognized the words in the study, the target fixation proportion was analyzed during a two-second time window, starting 300 ms after target word onset. Fixations occurring prior to that time window are likely initiated before target word onset and are hence not triggered by the target word (c.f. Canfield, Smith, Brezsnyak, & Snow, 1997; Haith, Wentworth, & Canfield, 1993; Matin, Shao, & Boff, 1993; see also Johnson & Huettig, 2011; Van Heugten & Johnson, 2011; Zangl & Fernald, 2007 for similar use of time windows).

2.2 Results

2.2.1 Results of Experiment 1a

To verify that toddlers were equally likely to look at each of the two pictures prior to target word onset, a one-sample t-test was conducted on the proportion of fixations towards the target in the second prior to target onset. This analysis revealed that the average target fixation proportion of .46 (SEM: .028) did not differ from the .5 chance level ($t(15) = -1.252; p = .230$), indicating that toddlers did not display a preference to either the target or the distracter picture before they had heard the target word.

The one-sample t-test after target word onset, in contrast, showed that the target fixation proportion of .65 (SEM: .035) did differ significantly from the .5 mark ($t(15) = 4.288; p = .001$; see Figure 7). That is, during the two seconds after target word onset (corrected for eye movement latencies), toddlers looked more towards the target than towards the distracter picture. One may wonder, however, if this effect may be driven by the heard word trials, which contained target words the toddlers previously heard the
Figure 7: Proportion of fixations to target in the time window before and after target word onset (on the left) and the proportion of fixations to target over time from target word onset (on the right) for 28-month-olds in Experiment 1a. Dashed lines indicate chance level. Dotted vertical line indicates average word offset.

speaker utter in the story. Figure 8, which displays the fixation proportions to targets separately for heard word trials and generalization trials, indicates that this is not likely the case. Separate analyses of the two trial types confirm this observation. That is, the fixation preference for target words is present in both the heard word trials (mean target fixation proportion: .64 (SEM: .041); \( t(15) = 3.499; p = .003 \)) and the generalization trials (mean target fixation proportion: .65 (SEM: .041); \( t(15) = 3.768; p = .002 \)). There are no differences between the two trial types (\( t(15) = .244; p = .811 \)).

These results demonstrate that after only brief exposure to a speaker with a previously unfamiliar Scottish accent, toddlers are able to recognize words produced by the accented speaker and this ability appears to be independent on whether the toddlers had previously heard the speaker say the intended word. However, whether word recognition is induced by hearing the Very Hungry Caterpillar story read by the same
Figure 8: Proportion of fixations to target broken down by heard word trials and generalization trials in the time window after target word onset (on the left) and the proportion of fixations to target over time from target word onset in these two trial types (on the right) for 28-month-olds in Experiment 1a. Dashed lines indicate chance level. Dotted vertical line indicates average word offset.

Scottish-accented speaker in the exposure phase is currently unclear. The results of Experiment 1b, where toddlers heard an Australian-accented speaker rather than the Scottish-accented speaker read the storybook, will address this question.

2.2.2 Results of Experiment 1b

As in Experiment 1a, the fixation proportion to target before target onset is compared to chance level. A one-sample t-test showed that the target and distracter were fixated equally (mean proportion of fixations to target: .47 (SEM: .023); t(15) = -1.224; p = .240), indicating that there were no biases prior to target word onset. After target word onset, however, toddlers preferred to look at targets over distracters (t(15) = 6.914; p<.000), as revealed by an average target fixation proportion of .67 (SEM: .02468; see Figure 9). Thus, both after hearing a Scottish-accented and after hearing an Australian-
Figure 9: Proportion of fixations to target in the time window before and after target word onset (on the left) and the proportion of fixations to target over time from target word onset (on the right) for 28-month-olds in Experiment 1b. Dashed lines indicate chance level. Dotted vertical line indicates average word offset.

Accented speaker, toddlers showed evidence of recognizing Scottish-accented target words in the test phase. Given that the speaker and accent changed between the exposure and test phase, this effect could not have been solely driven by a recollection of the test phase speaker’s previous production of the items in heard word trials.

2.2.3 Comparison between Experiment 1a and Experiment 1b

Although toddlers recognize the target words in both Experiment 1a and in Experiment 1b, it is possible that recognition accuracy may be higher in Experiment 1a, where toddlers have had greater exposure to the speaker in the test phase, compared to Experiment 1b, where toddlers heard a speaker of a different accent. To assess this possibility, the two groups of toddlers are compared in terms of accuracy (i.e., fixating the correct target image). Inspection of Figure 10 shows that the fixation proportion to
Figure 10: Proportion of fixations to target in the time window after target word onset broken down by experiment (on the left) and the proportion of fixations to target over time from target word onset for 28-month-olds in the two experiments (on the right). Dashed lines indicate chance level. Dotted vertical line indicates average word offset.

target over time is similar across the groups and an independent-samples t-test indicates that the average proportion of target fixations after target word onset does not differ based on the speaker who read the story ($t(30) = -.489; p = .628$).

2.3 Discussion

In Experiment 1, 28-month-olds were tested on their recognition of words in an unfamiliar Scottish accent. To test the effect of speaker accommodation, prior to test toddlers in Experiment 1a were presented with a version of the Very Hungry Caterpillar story produced by the same Scottish-accented speaker. Toddlers in Experiment 1b, in contrast, were presented with the same story, but read by a different speaker with an Australian accent. At test, toddlers in both groups reliably recognized the target words. This speaks to 28-month-olds’ ability to readily deal with accent variation in the
pronunciation of words, with or without prior exposure to the speaker or accent, and is impressive given suggestions in the literature that children at this age still experience difficulty comprehending accented speech in some contexts (Schmale et al., 2011).

Word recognition was robust independent of whether toddlers had previously heard the same speaker or a different speaker with a different accent, in that the recognition levels did not differ across the two groups. Why does exposure to the speaker’s accent not enhance subsequent word recognition? One explanation for the absence of a difference between the two experiments would be that listening to a speaker with an unfamiliar accent may have relaxed (rather than shifted) toddlers’ phoneme signal-to-word mappings. That is, because the speaker speaks in an unfamiliar accent, toddlers may have put less weight on the exact phonetic realizations of words, allowing words to be recognized even if they diverge from the pronunciations in the toddlers’ accent. This explanation is unlikely, however, as numerous studies outlined in detail in the Introduction have indicated that listening to a speaker who shifts certain phonemes leads to accommodating the specific shift only, both for toddlers and for adults. That is, no general broadening of the mapping between that speaker’s realization of those phonemes and the underlying linguistic representations was observed (Maye et al., 2008; White & Aslin, 2011). Further experiments are needed, however, to test this possibility more directly. For example, follow-up work could test 28-month-olds’ ability to recognize Scottish-accented words after listening to a Canadian speaker. Listening to a native-accented speaker, whose phonetic realizations are similar to those in the child’s typical environment, should not induce loosening of the signal-to-word maps. As a result, recognition patterns of the Scottish-accented target words in that study comparable to the
patterns in the current study would challenge the idea that words were recognized because of general relaxation of the mapping between the surface form and its underlying representation.

An alternative explanation for why speaker exposure did not enhance word recognition in this study would be that the task of recognizing words in the test phase did not impose any significant problems for the toddlers in this study. There are at least three reasons for why this may be the case. First, it is possible that 28-month-olds tested in this task do not experience any noticeable difficulties dealing with unfamiliar accents. Hearing words in sentence frames, rather than in isolation may have contributed to such prompt recognition processes. That is, the sentence frame may have provided toddlers with sufficient information to allow for immediate accommodation. If this were the case, the exposure phase would be largely irrelevant given that the speaker information extracted from the immediately preceding sentence would be sufficient for the toddlers to fully compensate the previously unfamiliar accent. Second, hearing words in sentence frames could have possibly assisted word recognition in an additional fashion. Specifically, the sentence frame may have set up the syntactic structure of the sentence such that it may have been easier to recognize words as the predicted final noun. Even when listening to native-accented speech, toddlers have been shown to better recognize words when they are imbedded in sentences rather than when they are presented in isolation (Fernald & Hurtado, 2006), possibly because of this reason. Together, the rapid accommodation of an unfamiliar speaker and the benefit of the presence of syntactic structure may mask any difficulties toddlers would have experienced when recognizing isolated words. A third explanation for why toddlers could cope so well with the accented
speaker in this study pertains to the paradigm I used. In this study, there was a three-second delay between the appearance of the two pictures on the screen and the onset of the target words in the auditory domain. It is thus possible that the presence of the pictures on the screen activated the depicted words (see Mani & Plunkett, 2010 for evidence that children implicitly name objects they are seeing on the screen). The toddlers’ task in this study would then have been to determine which of the two activated words would resemble the accented word most closely. Since both pictures in this study corresponded to familiar words, mapping an ‘oddly-produced’ word onto one of two activated lexical candidates is arguably not a demanding task and may have eased word recognition compared to a situation where no visual information is available.

Regardless of which of these explanations (or combination thereof) is correct, the finding that 28-month-old toddlers can recognize words in unfamiliar accents is intriguing and aligns nicely with the idea that children at this age are quite competent at dealing with speaker-related differences in the realization of words. However, it also raises the question of why toddlers in the word learning study conducted by Schmale and colleagues (2011) performed so poorly. In that study, 24-month-olds did not succeed at recognizing that words in an unfamiliar accent were the same as those just learned in their native accent (though 30-month-olds tested using the same materials and procedure did succeed). One obvious reason for the discrepancy between that and the current study may be the difference in age. The 28-month-olds tested here fall in between the 24- and 30-month-olds in the word learning study and the 30-month-olds in that study did show evidence of generalizing from native to accented speech. Had they tested 28-month-olds instead, they may have found the same effect. The development of the ability to deal with
accent variability during the recognition of words may in other words develop between 24 and 28 months of age. A second possibility is that Spanish-accented English may be a particular hard accent to grasp. Children are generally better at detecting non-native accents than native, but unfamiliar accents (Floccia, Butler, Girard, & Goslin, 2009; Girard, Floccia, & Goslin, 2008), likely because of the more salient divergence from the home accent. It seems plausible that detecting an accent is caused by an inability (or increased difficulty) to understand the speaker and if this were the case, that would explain why non-native Spanish might have been harder to understand than the native accent used in Experiment 1. Finally, a third possibility may have to do with the fact that the study by Schmale and colleagues not only tested children’s ability to adapt to different accents, but also incorporated a word learning component. While it is indeed possible that 24-month-olds were unable to recognize words they had learned in their own accent when they were later presented in Spanish-accented English, it is equally well possible that these toddlers failed to learn the words in the first place. A study examining the effects of word learning and accent adaptation separately would resolve this confound.

Taken together, the results of Experiment 1 suggest that toddlers can, at least under some conditions, recognize words embedded in sentence frames in unfamiliar accents without any prior exposure to the speaker they are listening to. This opens the ground for further examination of this competence in younger children. When does it develop? Would younger children, for example, be equally able to recognize words in unfamiliar accents? To address this question, Experiments 2 and 3 test 20-month-old toddlers’ word recognition abilities in an unfamiliar accent using a similar paradigm.
3 Experiment 2

Experiment 1 shows that at 28 months of age, toddlers are, at least to some extent, able to cope with speaker-related variability, recognizing familiar words in an unfamiliar Scottish accent even without any prior access to the speaker’s characteristics. This experiment examines the developmental trajectory of this ability by looking at 20-month-old children’s ability to overcome difficulties imposed by phonetic mismatches between words in Canadian-accented English and words in an unfamiliar Australian accent. The Australian-accented speaker from Chapter 2 was used to produce the stimuli in both the exposure phase and the test phase. Given that infants as young as 15 months of age can, under some conditions, adapt to the Australian-accented speaker, this reveals that any difficulty the accent may impose should not be impossible to overcome.

Twenty-month-olds were chosen to participate in this study for two reasons. First, Chapter 2 suggests that by 22 months of age, children possess some ability to recognize words by the same Australian-accented speaker even in the absence of any speaker context. By selecting just slightly younger children, I aim for a listener group whose speech perception abilities in the face of accent divergence are still somewhat fragile, at least without prior exposure to the speaker. Second, although 20-month-olds’ word recognition skills may not have fully matured yet, there is some evidence using the same paradigm suggesting that by this age, toddlers can accommodate single vowel changes (White & Aslin, 2011). While unfamiliar natural accents typically involve much larger discrepancies from the native accent than just a sole segment shift, these findings indicate that at least some adaptation skills are in place and, perhaps even more importantly, that these adaptation skills can be picked up by the current procedure. Whether 20-month-olds
also possess the competence to adjust to natural accents, however, is an empirical question; one that will be addressed in the current study. If the toddlers in this experiment would accommodate the speaker and consequently recognize the target words, they should look reliably more towards the target picture than towards the distracter once the target word unfolds.

3.1 Methods

3.1.1 Participants

A total of 16 normally developing English-learning 19.5- to 21-month old toddlers from the Greater Toronto Area were tested (age range: 602-641 days; 8 boys). As was the case in all previous experiments, no hearing issues or recent ear infections were reported. In addition, none of the infants had had any substantial exposure to Australian-accented English, as established by a language questionnaire at the end of the lab visit. An additional 6 infants were tested, but excluded from the analyses due to extreme fussiness or failure to complete the study.

3.1.2 Stimuli

The visual stimuli for Experiment 2 were the same as those for Experiment 1. The only difference between the two experiments was that in Experiment 2 all auditory stimuli, both the story and the test phase, were recorded by the Australian speaker who also recorded the story in Experiment 1b. Target words lasted on average 646 ms.
3.1.3 Design, Procedure, and Off-line Coding

The design and procedure were the identical to those in Experiment 1. Toddlers’ eye movements were also coded in an identical fashion as in Experiment 1, with 4 videos being recorded by a second coder. As before, inter-coder reliability on the individual fixation durations was consistently high (mean correlation = .987).

3.2 Results

Data analysis for the current experiment was comparable to that of Experiment 1. First, the fixations to the target picture were measured as a proportion of fixations towards either the target or the distracter in the second prior to target word onset. The average fixation proportion of .49 (SEM: .025) did not differ significantly from chance (t(15) = -.297; p = .771). Before hearing the target word, toddlers were thus equally likely to look at the target and distracter picture.

![Figure 11: Proportion of fixations to target in the time window before and after target word onset (on the left) and the proportion of fixations to target over time from target word onset (on the right) for 20-month-olds in Experiment 2. Dashed lines indicate chance level. Dotted vertical line indicates average word offset.](image-url)
This pattern of results was the same for the two seconds after target word onset (starting 300 ms after target word onset). Even when hearing the target word, toddlers’ fixation proportion to the target (mean: .53, SEM: .016) did not differ significantly from the .5 chance level ($t(15) = 1.799; p = .092$; see Figure 11). In order to examine the possibility that 20-month-olds, unlike the 28-month-olds tested in Experiment 1a, might only recognize words they had heard the speaker produce before (such effect may be masked in the analysis including all items, as there are many more trials containing words children had not heard before than trials containing words that had been heard before), separate one-sample t-tests were conducted for heard word trials and generalization trials. However, neither the average proportion of fixations to target for previously heard items of .55 (SEM: .038) nor for the words that were not previously heard by the speaker of .52 (SEM: .015) exceeded chance level ($t(15) = 1.255; p = .229$ and $t(15) = 1.375; p = $.

![Figure 12](image.png)

*Figure 12:* Proportion of fixations to target broken down by heard word trials and generalization trials in the time window after target word onset (on the left) and the proportion of fixations to target over time from target word onset in these two trial types (on the right) for 20-month-olds in Experiment 2. Dashed lines indicate chance level. Dotted vertical line indicates average word offset.
.189, respectively) and the proportions of fixations to target did not differ across the trial types \((t(15) = .709; p = .489)\). This is reflected in Figure 12, displaying the target fixation proportion over time for each of the two trial types.

3.3 Discussion

Unlike the 28-month-olds tested in Experiment 1, no evidence was obtained indicating that the 20-month-olds tested in the current experiment recognized the target words in the test phase of the study. Even after having been exposed to the speaker’s vocal and accent characteristics, toddlers did not succeed in accommodating the speaker and consequently failed to look longer at the target than at the unrelated distracter picture.

What could be the reason for the absence of word recognition in this study? From Experiment 4 in Chapter 2, we know that adaptation to this particular speaker with this particular accent is, at least under certain conditions, possible. If 15-month-olds who have been given the opportunity to become familiar with the relevant characteristics of the speaker are able to recognize words produced by this speaker, it seems unlikely that the accent is generally too hard to deal with for 20-month-olds. Two alternative possibilities remain. First, it is possible that the Very Hungry Caterpillar story contained too many unfamiliar words for 20-month-olds, just like the story was too difficult for the 15-month-olds who were not highly familiar with the story in Experiment 3 of Chapter 2 and may hence not have induced adaptation. If lexical feedback is required for adaptation, as Chapter 2 suggests, this would explain why both infants in Experiment 3 of that chapter and the toddlers in this experiment experienced difficulty dealing with phonetic
divergences in the realization of words between accents even after they had been exposed to the accented speaker.

A second possibility is that it is not the words in the story, but the words used in the test phase that were too difficult for 20-month-olds to understand. Although the words that were selected tend to be learned early in life, the stimuli were originally designed to be used with 28-month-olds. In fact, whereas the average production rate of the words in the test phase was on average 89.9% for the older age group, the production rate of the younger age group for the same words was 62.6% according to the Lexical Development Norms for English (Dale & Fenson, 1996). In addition, from the questionnaire administered to the parents at the end of their visit, it is clear that 28-month-olds generally know these words better than 20-month-olds (on average 95.1% of the words and pictures were judged to be recognized at 28 months of age; this percentage dropped to 81.6% for the 20-month-olds). While recognition scores indicate that even 20-month-olds should still recognize the used target words, it is possible that (contrary to the previously observed tendencies for parents to underestimate their infants’ vocabularies; Houston-Price, Mather, & Sakkalou, 2007) parental judgments in this study overestimated the words infants know. As a result, infants may have adapted to the speaker, but these adapted signal-to-word mappings were rendered useless here because infants simply did not know the words in the study. To rule out this latter possibility, Experiment 3 presents toddlers with Canadian-accented English words.

4 Experiment 3

Experiment 2 shows that 20-month-olds experience difficulty recognizing words in
Australian-accented English, even if they have had some exposure to the speaker. To test whether toddlers’ understanding of the target words is sufficient to recognize them in the absence of phonetic variability in their realization, toddlers in Experiment 3 were presented with the same materials as the toddlers in Experiment 1 and 2, but pronounced in their own Canadian English accent. If the words in this study would be sufficiently accessible for 20-month-olds, then presenting the words in Canadian English should induce longer looks towards the target picture rather than towards the distracter.

4.1 Methods

4.1.1 Participants

Another 16 normally developing English-learning 19.5- to 21-month old infants from the Greater Toronto Area were tested (age range: 595-636 days; 9 boys). No hearing issues or recent ear infections were reported. An additional 5 toddlers were tested, but excluded from the analyses due to extreme fussiness or failure to complete the study.

4.1.2 Stimuli

The stimuli of Experiment 3 closely followed those of Experiment 1 and 2. Unlike the stimuli in the first two experiments in this chapter, however, all auditory stimuli for the current experiment were recorded by a native Canadian-English speaker from the Greater Toronto Area. Toddlers thus heard both the story and the test phase in their own native accent. The Very Hungry Caterpillar story that was presented prior to test was included to ensure that toddlers had spent the same time in the test booth watching a movie before the test trials started and hence control as well as possible for attention span differences.
between the experiments. Visual stimuli were identical to those in Experiment 1 and 2. Target words lasted on average 698 ms.

### 4.1.3 Design, Procedure, and Off-line Coding

The design, procedure, and coding practice were identical to those in Experiment 1 and 2. Four videos were recorded by a second coder. As before, inter-coder reliability on the individual fixation durations was consistently high (mean correlation = .996).

### 4.2 Results

The analyses of the data as well as the time windows are identical to those in Experiment 2. The proportion of fixations to the target (as a function of the sum of target and distracter fixations) was first assessed in the second before target word onset to check for any inherent biases for the target words or distracters. A one-sample t-test revealed that the average target fixation proportion of .50 (SEM: .022) did not differ from the chance level test statistic of .5 ($t(15) = .096; p = .925$), indicating that toddlers fixated on the target and distracter equally before target word onset.

If the target words were recognized in Canadian English, the proportion of target fixation should exceed .5 in the time window after target word onset. This is indeed the case. The average target picture fixation proportion was .63 (SEM: .027), significantly different from chance, as revealed by a one-sample t-test ($t(15) = 4.982; p < .001$; see Figure 13). This effect, however, appears to be driven by the words that did not occur in the story only. While these words were recognized quite efficiently with an average proportion of fixation to the target of .66 (SEM: .030), significantly exceeding chance
level ($t(15) = 5.297; p < .001$), no such effect was observed for the words that did occur in the story (mean target fixation proportion: $.56$ (SEM: $.037$); $t(15) = 1.587; p = .133$). A direct comparison between the two trial types indicates that they indeed give rise to different recognition patterns ($t(15) = 2.501; p = .024$; see Figure 14). This may indicate that the story words may have been too advanced for 20-month-olds to be recognized in the current paradigm, even when presented in their own Canadian accent.

4.3 Discussion

When presented with a speaker of their own variant of English, Canadian 20-month-olds readily shift their gaze toward the appropriate target picture, indicating that the toddlers have no trouble recognizing the words in the test phase. As the target words

![Figure 13: Proportion of fixations to target in the time window before and after target word onset (on the left) and the proportion of fixations to target over time from target word onset (on the right) for 20-month-olds in Experiment 3. Dashed lines indicate chance level. Dotted vertical line indicates average word offset.](image)
Figure 14: Proportion of fixations to target broken down by heard word trials and generalization trials in the time window after target word onset (on the left) and the proportion of fixations to target over time from target word onset in these two trial types (on the right) for 20-month-olds in Experiment 3. Dashed lines indicate chance level. Dotted vertical line indicates average word offset.

In Experiment 2 and 3 are identical, this suggests that toddlers’ failure to recognize words in Experiment 2 is not likely to be due to 20-month-olds’ inability to recognize the particular words chosen in this study.

Interestingly, unlike the findings for the 28-month-olds (who were tested on an unfamiliar accent), the recognition scores for 20-month-olds were lower on heard word trials than on generalization trials. In fact, as Figure 12 shows, average recognition scores for words that had previously occurred in the Very Hungry Caterpillar story peaked at around 60%, whereas recognition scores for familiar words that did not occur in the story peaked at around 80%. This demonstrates that even when the story words were presented in toddlers’ native accent, recognition of these words was not as robust as the recognition of other, previously unheard, words in the
test phase. This discrepancy could potentially be due to the fact that the words that occurred in the story were somewhat harder than the previously unheard words. In line with this explanation, parents judged their toddlers to understand and recognize 85.9% of the words in the generalization trials, but only an average of 68.8% of the words in the heard word trials. While certainly not conclusive, these findings do support the notion that toddlers’ failure to recognize Australian-accented words in Experiment 2 was due to their inability to access sufficient words during the presentation of the story and suggest that had toddlers been familiar to the story or had easier, more accessible, words been used, 20-month-olds in Experiment 2 might have adapted to the speaker’s accent (see Van der Feest & Johnson, 2011 for evidence that 24-month-old toddlers can adapt to speakers’ natural language category boundaries after exposure with familiar words).

5 Discussion of Experiments 1-3

In this chapter, I examined 20- and 28-month-old toddlers’ abilities to recognize words in unfamiliar accents. Most impressively, 28-month-olds evidently have no trouble mapping target words in unfamiliar accents onto their underlying word representations in this task, even without prior exposure to the characteristics of a speaker. While previous studies have claimed that toddlers at 24 months of age are unable to map words in foreign accents onto the linguistic representation of recently-learned words (Schmale et al., 2011), the 28-month-olds tested here performed much better and readily recognized words in an accent they had never heard before and did so without any apparent difficulty, similar to the 22-month-olds in Chapter 2. The
finding that 28-month-olds in this study can cope with a substantial amount of accent variation regardless of whether they had had prior exposure to the speaker further adds to the idea that word representations are abstract from early on in development.

In contrast to the 28-month-olds, the 20-month-olds in the current study did not appear to be able to deal with variation in the pronunciation of words across different accents. Although caution needs to be taken when comparing the results of toddlers who listened to different unfamiliar accents (Scottish-accented English in Experiment 1 and Australian-accented English in Experiment 2), dramatic improvements appear to take place between 20 and 28 months of age. It is possible, of course, that 20-month-olds do possess the underlying competence to deal with accent-induced differences in the phonetic realization of words, but that the information that drives speaker adaptation is unavailable from the Very Hungry Caterpillar story. This might be the case if accommodation is indeed lexically mediated and the words in the exposure phase may have been too hard to access.

The idea that 20-month-olds can accommodate at least some phonetic anomalies when tested in a similar type of paradigm comes from a study showing that 19-month-olds, after being exposed to words containing a phoneme shift by a particular speaker later recognize novel words that are realized with the shifted version of that phoneme (Aslin & White, 2011). Without the relevant exposure, these words were not recognized. Adaptation to the speaker in this case, however, only involved a shift in the mapping of a single phoneme. It seems plausible that with more extreme variability, as is the case for natural accents, toddlers – at least at this age – may need exposure to a wider variety of words before the complete signal-to-word
mappings are in place. The finding that words that had occurred in the story were only retrieved with difficulty, even when they were presented in toddlers’ native accent, suggests that accessing them when the speaker has an Australian accent may have been extremely challenging. Given that the test items that also occurred in the storybook (i.e., butterfly, cake, cheese, and strawberry) constituted some of the easiest words from the story (many of the key many key words such as caterpillar, sausage, and cocoon are much more advanced words that are learned later than the relatively easy words used in the test phase), it is possible that toddlers, at least at 20 months of age, may have been unable to access majority of the words in the exposure phase when produced in Australian English. Thus, similar to the 15-month-olds in Chapter 2, the words in the story may have been somewhat too difficult to for toddlers to use for speaker accommodation.

In Experiment 3 in Chapter 2, exposure to the Very Hungry Caterpillar story in the unfamiliar accent alone did not induce accommodation in 15-month-olds. Similarly, in the current study 20-month-olds fail to adapt to the speaker’s unfamiliar accent when presented with a read version of the same book in the unfamiliar accent. Albeit admittedly speculative, the convergence between these results is consistent with the proposal offered in Chapter 2 that a lack of lexical access may prevent accent adaptation. Speaker accommodation, in other words, may be lexically guided such that only when children access a sufficient number of words can they work out the signal-to-word maps. Both adult listeners and six-year-old children have previously been shown to use lexical feedback for adapting to a shift in the speaker’s pronunciation of sounds (Eisner & McQueen, 2005; McQueen et al., in press; Mitterer & McQueen, 2009;
Norris et al., 2003). Lexical access may similarly be a prerequisite for accent adaptation in younger children. If this is indeed the case, then toddlers may be better able to accommodate a speaker’s accent when the speaker first produces words that are better known by the toddlers. Follow-up studies should further investigate this possibility.

Taken together, this chapter shows that by 28 months of age, toddlers are remarkably proficient at dealing with accent-related variation in the realization of words, at least in a preferential looking task such as the one conducted here. Not only does this study show that 28-month-olds recognize words in accented speech (similar to the 22-month-olds in Chapter 2), it also demonstrates that they do so rapidly and accurately, regardless of whether they had heard the accented speaker beforehand. Hearing accented words in sentences might thus be sufficient to accommodate the speaker’s accent, even for toddlers, whose vocabularies are much smaller than those of adults. Such an ability to accommodate speaker accents ‘on the fly’ likely develops sometime between 20 and 28 months of age, with the possibility that the younger children would accommodate the speaker when first presented with more accessible words produced by that speaker. In combination with Chapter 2, these data suggest that children’s early word representations are sufficiently abstract to deal with accent variation from very early on and that the process of deciphering the signal-to-word maps becomes increasingly more advanced over time.
Chapter 4

Cross-gender word form recognition in early infancy

1 Introduction

Chapters 2 and 3 have demonstrated that children in their second and third year of life are much more proficient at dealing with accent variability during word recognition than previous work has suggested. Accents, however, are not the sole source of acoustic variability in the speech signal. In fact, variation due to accents may be qualitatively different from other types of speaker-related variation, such as variation due to the gender or the affect of a speaker. Specifically, accents may be largely characterized by shifts in the speakers’ realization of multiple phonemes (as well as differences in rhythm and intonation), whereas variability in speaker gender (e.g., from high to low pitch when dealing with female and male speakers) or affect (e.g., the differences correlated with speech produced in a happy vs neutral tone) involves an overall shift of the words in acoustic space. Thus, in contrast to accents, factors such as speaker gender causes acoustic variability across word tokens while leaving the phonological information largely similar to infants’ native accent. To examine if and how infants can also contend with these other speaker-related factors inducing variability in the first year of life, this chapter looks at 7.5-month-old infants’ abilities to map words spoken by speakers of different genders onto the same underlying representation. At an age when infants have neither fully tuned into the specific phoneme inventory (and boundaries) of their native language or accent (e.g., Werker & Tees, 1984) nor built up a rich vocabulary that can be used for lexical feedback
in accent adaptation, manipulating speaker gender is an efficient way to examine whether infants can cope with basic forms of acoustic variability during speech perception.

As discussed in Chapter 1, previous studies examining the effect of speaker gender or affect on infants’ word mapping skills have suggested that infants are initially unable to accommodate gender- or affect-induced dissimilarities in the acoustic realization of words (Houston & Jusczyk, 2000; Singh et al., 2004). At 7.5 months of age, word forms presented in a female voice in the training phase, for example, are easily recognized when produced by a similar-sounding female voice in the test phase, but tend to remain unrecognized in a perceptually distinct male voice (Houston & Jusczyk, 2000). Similarly, words spoken in a happy affect are recognized when they are later presented in the same happy affect, but not when they are produced in a neutral affect (Singh, Morgan, & White, 2004). What does that mean for early word form representations? Logically speaking, if infants fail to recognize acoustically dissimilar instances of the same word as referring to the same underlying representations, they should be developing separate representations for word forms spoken in a male voice and those spoken in a female voice, and similarly for word forms spoken in a happy affect and word forms spoken in a neutral affect. Such inability to recognize the phonological constancy of words would thus be devastating for lexical development. Of course, the acoustics of a word form can differ along many different dimensions and it seems highly implausible that infants initially build hundreds of individual representations for each word they acquire.

However, taking a closer look at these word mapping studies reveals that these studies have the same limitation as word recognition studies that present infants with isolated words in unfamiliar accents (e.g., Best et al., 2009; Mulak et al., 2008; Nathan et
al., 2008). That is, infants in this paradigm are trained on lists containing isolated words produced by voice actors. In the subsequent test phase, infants’ acceptance of perceptually dissimilar variants of word forms as instances of these trained words is measured. Because these words are initially presented in isolation in a potentially somewhat unnatural fashion, infants have only limited access to the speaker’s natural voice characteristics. This reduction in the availability of the speaker’s characteristics during the training phase compared to natural listening conditions may impede the mapping of the word form onto the intended underlying representation and could consequently hinder later recognition. Without any speaker context, infants may thus not be presented with sufficient information to allow them to interpret the isolated words as a function of the speaker they are listening to. For this reason, previous research may have underestimated infants’ early speech perception abilities in the face of acoustic variability due to speaker gender or affect.

The goal of the current study is twofold. First, I test the possibility that long-term familiarity with the speaker may help 7.5-month-old infants overcome the lack of speaker context when mapping acoustically distinct tokens onto the same underlying representation. In everyday life, much of the speech children hear is produced by their parents or caregivers – speakers to whom they have had repeated exposure. In naturalistic listening conditions, infants may thus use previously established signal-to-word maps to better understand a speaker’s intended message. This is plausible as even adults’ speech perception benefits when they are listening to speech produced by familiar rather than unfamiliar speakers (e.g., Nygaard & Pisoni, 1998; Nygaard et al., 1998; Smith & Hawkins, 2012). Infants tested in this study have about 7.5 months of experience with their parents’ voices and have likely adapted to the way their parents produce speech, if
not at the segmental level, conceivably at least at the prosodic level. Specifically, the early preference for maternal over strangers’ voices merely a few days after birth (DeCasper & Fifer, 1980; Mehler, Bertoncini, Barrière, & Jassik-Gerschenfeld, 1978) develops into an enhanced ability to segregate two separate speech streams when one of the talkers is their own mother by 7.5 months of age (Barker & Newman, 2004). It is thus possible that infants are also able to use this stored knowledge of their parents’ voices when trying to figure out the mapping between the surface form of a word to its underlying representation. Parental voices thus provide an excellent means to test whether coping with variability in the realization of words may be facilitated by long-term exposure to the speakers.

A second goal of this study is to test whether infants may be better able to contend with acoustic variability due to the gender of the speaker when presented with more everyday-like listening conditions. Even though much of the speech infants hear is produced by familiar speakers, they are occasionally exposed to people they have not met before. Would infants be able to accommodate such novel speakers? Much like Chapter 2, this study thus aims to present infants with the information they need to establish the mapping between acoustically distinct realizations of the same word, such that the previously observed difficulties in cross-gender signal-to-word mapping (Houston & Jusczyk, 2000) can be overcome. Unlike Chapter 2 (and 3), however, the current study does not include an exposure phase where infants are first familiarized with an unfamiliar speaker. Instead, speaker information is presented to children by using passages that contain repeated tokens of a target word. Infants are then tested on their recognition of these target words when produced in isolation. Familiarizing infants with passages
containing target words rather than lists of isolated words likely increases exposure to the speaker-specific idiosyncrasies in the passages and introduces infants to a wider range of the speaker’s idiosyncratic realizations of speech segments and utterance-level prosody. Unlike isolated words, fluent speech may thus provide infants with the opportunity to adapt to the systematic variation of an unfamiliar speaker, thereby allowing them to start recognizing acoustically distinct variants of a word as referring to the same underlying representation (cf. Bradlow & Bent, 2008 and Clarke & Garrett, 2004 for perceptual learning with adults after brief exposure and Chapter 2 for an infant study showing similar results). In addition, the increased variability in the realization of the target words available from the training passages (relative to the word lists in previous studies) may also facilitate word form extraction (cf. Rost & McMurray, 2009; Singh, 2008).

To potentially further facilitate word form recognition in this study, the stimuli were produced by speakers who, at the time of recording, were parents of 7.5-month-olds rather than voice actors. Parents, unlike the voice actors in previous studies, are speaking to infants on a daily basis, and this may induce speech that is more representative of the input typically received by children. Because infant-directed speech differs from adult-directed speech in terms of pitch, accent, and vowel space (Burnham, Kitamura, & Vollmer-Conna, 2002; Kuhl et al., 1997), parental recordings may, from a child’s perspective, contain more naturalistic acoustic cues that enhance word recognition than the recordings made by actors used in past studies.
2 Experiment

In order to examine whether infants are able to map acoustically distinct tokens onto the same word representations and whether any such competence may be enhanced when listening to familiar speakers, two groups of 7.5-month-olds are tested. The first group is exposed to passages recorded by their mothers. They are then tested on lists of isolated word tokens spoken by their fathers. Both speakers are thus highly familiar. The second group of infants hears the same stimuli as the first group. As the voices belong to another infant’s parents, infants in this group have not had any prior access to the speakers’ voices, apart from the training phase. If the use of more naturalistic stimuli in the training phase indeed yields more robust word encoding, infants should succeed in the cross-gender word recognition task. I thus hypothesize that infants will listen longer to trained than to novel words (cf. Jusczyk & Aslin, 1995; Seidl & Johnson, 2006) despite the change in gender of the speaker between the training and the test phase. Moreover, if voice familiarity plays a role, this effect should be more pronounced for those infants tested on their own parents’ voices than for those tested on unknown voices. The predicted preference for trained over novel words may, in other words, be larger for the former as compared to the latter group.

2.1 Methods

2.1.1 Participants

Forty-eight normally developing monolingual English-learning 7.5-month-olds with no

---

2 Data from this experiment have been reported in Van Heugten & Johnson (2012). This paper can be accessed online at http://jslhr.asha.org/cgi/content/full/55/2/554
reported hearing problems from the Greater Toronto Area were tested in this study (age range: 220 - 248 days; 25 girls). An additional eleven infants were tested, but excluded from the analysis due to failure to complete the study or extreme fussiness.

2.1.2 Stimuli

Prior to test, twenty-four mothers and fathers whose infants were recruited to participate in the study were audio taped in a sound-attenuated booth. All parents were either native English speakers or had learned English before five years of age in an English-speaking country. Some parents, however, were born in regions of the world outside of North America where English is spoken as a first language, such as India, England, and Trinidad and had non-Canadian accents. Mothers recorded two of four six-sentence passages in infant-directed speech (see Appendix E). Each passage contained a target word (boat, cup, pear, or toque – in Canadian English, ‘toque’ is the commonly used word for a knitted hat) that was repeated once in every sentence. Within a passage, this target word appeared twice in sentence-initial position following a function word (e.g., Her boat had white sails), twice sentence-medially (e.g., That horn on the boat was really loud), and twice in sentence-final position (e.g., This girl will steer my big boat). For each mother, a training video was created consisting of three alternating repetitions of each of the two recorded passages accompanied by a static photo of the mother. All training videos were between 110 and 120 seconds long and contained 18 tokens of each of the target words.

Because infants were tested on each of the four target words (two being trained and two being novel), fathers recorded multiple isolated instances of all four target words,
which were edited to create different test word lists. In each list, five tokens of a single target word were repeated three times. Individual tokens were separated by a pause of approximately 600 ms and test lists were 17.31 seconds long.

2.1.3 Procedure

Infants were seated on their parent’s lap in front of a TV screen in a double-walled sound-attenuated booth. The experimenter, located outside the booth, started the training phase as soon as the infant oriented towards the TV screen. The movie continuously played until the end, after which the test phase started. In the test phase, an adapted version of the Headturn Preference Procedure (as used by Jusczyk & Aslin, 1995) tested infants’ recognition of the trained words across different genders. First, a red light at the panel in front of the infant started flashing. Once the infant oriented towards this light, one of the two lights at the side panels started to flash. As soon as the child turned towards this flashing light, the word list started playing from the loudspeaker mounted underneath the blinking light. Trials either played until the end of the list or until the infant looked away for two seconds. Parents listened to masking music over closed headphones so that they could not bias the child’s behavior.

2.1.4 Design

Half of the infants were presented with their own parents’ voices, while the other half were yoked pairs listening to the same stimuli in unfamiliar voices. The target words used during training (boat and toque, cup and boat, toque and pear, or pear and cup) were counterbalanced across conditions, as was the order of presentation of the two passages in
the training phase. Infants were tested on all four test lists. For each infant, two of these test lists contained trained and two contained novel words. Test lists were randomly presented once in each of three blocks (twelve test trials in total).

2.2 Results

First, orientation times were calculated for each infant in each trial. Orientation times for trials more than 2.5 standard deviations away from the infant’s mean (3 out of 576 data points) were discarded. Mean orientation times to trained and novel words were then calculated for each infant separately. On average, infants tested on their parents’ voices listened to lists with trained words for 10.18 s and to lists with novel words for 9.45 s, with 18 out of 24 infants listening longer to the trained words. Infants tested on unfamiliar voices listened to lists with trained words for 10.31 s on average and to lists with novel words for 9.51 s, with 16 out of 24 infants listening longer to the trained words (see Figure 15). A mixed $2 \times 2$ ANOVA with Word Familiarity (trained vs. novel words) as a within-subject factor and Voice Familiarity (own parents’ voices vs. unknown voices) as a between-subject factor revealed a main effect of Word Familiarity ($F(1,46) = 4.224; p = .046$), showing that infants listened longer to word lists containing trained words as opposed to word lists containing novel words. No other significant main effects or interactions were found (all $Fs < 1$), indicating that the performance of infants presented with their own parents’ voices did not differ from those presented with the voices of another infant’s parents.
Discussion of the current experiment

These results demonstrate that infants as young as 7.5 months of age are capable of coping with naturally occurring surface variation in the realization of words. More specifically, this study is the first to show that 7.5-month-olds can map word tokens produced by a female and a male speaker onto the same underlying representation in a word segmentation task. This is consistent with the view that early word representations are independent of the speaker’s gender, such that infants are able to map perceptually dissimilar realizations of the same word onto the same linguistic representation.

An important difference between this study and earlier studies using the same paradigm is that infants in this study first heard the target words embedded in fluent speech and spoken by parents rather than hearing them in isolation spoken by voice actors. The use of fluent speech instead of isolated word tokens may have assisted

Figure 15: 7.5-month-old infants’ mean orientation times in seconds (and standard errors of the mean difference scores) to trained and novel test words broken down by Voice Familiarity.
performance on this task in at least two ways. First, exposure to fluent speech provides infants with a wider range of the speaker’s segmental and prosodic information. This may better enable infants to adapt to the speaker, even when the speaker was previously unfamiliar. Better speaker adaptation may, in turn, allow for the generation of more robust word representations. Second, unlike words in isolation, fluent speech embedded in context, comparable to what infants typically experience in their everyday life, naturally contains some degree of variability that could be useful for extracting robust representations of word forms.

The use of authentic (parental) voices rather than voice actors may also have assisted performance in this study in at least two ways. First, conversational partners often display convergence in speaking rate (Webb, 1972), pitch (Gregory, 1990) and phonetic realizations (Pardo, 2006). For frequently interacting partners, such as the parents in this study, this type of accommodation could lead to long-term effects evidenced even in the absence of the partner and might have led to greater acoustic similarity between the word tokens produced in the male and female voice. Note, however, that infants’ ability to map acoustic realizations of a word produced by male and female speakers has since been replicated using speakers who did not know each other (Johnson, Seidl, & Tyler, in preparation), rendering it unlikely that vocal convergence would be the sole explanation for infants’ behavior in the current study.

The use of the more authentic parental voices in this study may have assisted performance in a second way. As discussed before, speech recorded by infants’ parents (as opposed to voice actors in previous studies) may better reflect speech encountered by infants in their everyday environment, even if the speakers themselves are unfamiliar.
This might decrease task demands such that more cognitive resources can be devoted to mapping the realization of one word onto the realization of another. Although I cannot determine the exact factor(s) responsible for infants’ early cross-gender word recognition success in the current study, this does not detract from the primary finding that when provided with more natural listening conditions, infants are competent in coping with variability in the speech signal.

In this study, I used cross-gender acoustic differences to examine infants’ ability to deal with the lack of invariance in the speech signal. However, variability in the speech signal is, of course, not restricted to differences in speaker gender. Other factors, such as emotional affect and accent, also affect the acoustic realization of words. The results of previous studies, similar to findings on cross-gender word form generalization (Houston & Jusczyk, 2000), suggest that infants initially experience difficulty overcoming these forms of variability (Schmale et al., 2010; Schmale & Seidl, 2009; Singh et al., 2004). These studies, however, all have the same limitation: they have presented infants with lists of isolated words. This could be potentially problematic, as isolated words do not contain the acoustic, segmental, and prosodic richness of fluent speech and may therefore reduce infants’ opportunity to adapt to an unfamiliar speaker. Using isolated word tokens may, in other words, have failed to provide infants with the information they need to link two acoustically distinct word tokens to the same underlying representation of a word form. Although speculative, it is thus possible that infants would have been found successful at coping with acoustic variability in the speech signal due to affect or perhaps even accent, had these previous studies used naturalistic fluent speech material similar to the current study instead.
The finding that infants generalize across acoustically distinct word tokens raises theoretically important questions regarding the nature of infants’ early word representations. While the results of this study argue against extreme episodic models lacking any form of abstraction, on their own, they do not rule out exemplar-based theories of early speech perception in general. That is, detailed memory traces including talker-specific information may be stored in infants’ mental lexicons, attached to an emergent and more abstract prototypical representation of a word. Alternatively, and in line with the results obtained with 15-month-olds Chapter 2 and 28-month-olds in Chapter 3, however, word representations may be speaker-independent even at the earliest stages. For example, the developing lexicon may consist of abstract phonological words stripped off from lexically irrelevant information. In such an abstractionist view, indexical information would still play an important role in early word recognition, but only at the prelexical level (see Eisner & McQueen, 2005; Kraljic & Samuel, 2006, 2007; McQueen et al., 2006 for evidence that talker adaptation takes place at the prelexical level in adults).

Interestingly, infants’ ability to cope with cross-gender variation in the realization of words held regardless of whether infants were presented with their own parents’ or unfamiliar voices, suggesting that long-term exposure to speaker-specific idiosyncrasies might not have an effect over and above the effect of short-term exposure to a previously unknown speaker’s voice. While previous work demonstrates that speaker familiarity can enhance infants’ speech processing when the familiar speaker is presented concurrently with an unfamiliar speaker in the background (Barker & Newman, 2004), this finding may have been due to the more challenging listening conditions in that study. That is, if
speaker adaptation in fluent speech is fast, and hearing a few sentences at most is sufficient for accommodating unfamiliar speakers in clear speech, it is plausible that long term voice familiarity only helps under more adverse listening conditions (e.g., speech in noise or multiple background talkers), where accommodation would be more challenging. This interpretation nicely aligns with the observation that adult studies showing an advantage for processing speech produced by familiar over unfamiliar speakers typically involve the presentation of speech in noise (e.g., Nygaard & Pisoni, 1998; Nygaard et al., 1998; Smith & Hawkins, 2012).

In short, this study suggests that infants’ early signal-to-word mapping skills under reasonably naturalistic listening conditions are sufficiently flexible to allow for non-trivial generalizations, such as those across adult speakers’ gender. Even at 7.5 months of age, infants may thus possess a readily available abstraction mechanism that allows them to generalize across phonologically irrelevant acoustic dimensions of the speech signal. Future models of early word recognition need to take into account these enhanced speech perception abilities when describing the transition from infant into adult listeners.
Chapter 5
General Discussion

1 Summary of findings

In this dissertation, I tested the hypothesis that, when provided with ecologically valid listening conditions, infants and toddlers can efficiently cope with speaker-related variability in the realization of words. This hypothesis was motivated by the lack of continuity observed between the infant and adult literature on speech perception. While adult studies have shown that listeners make great use of prior experience with specific speakers to better understand their subsequent speech (Bradlow & Bent, 2008; Clarke & Garrett, 2004; Dahan et al., 2008; Maye et al., 2008), infants’ word mapping or word recognition in the face of speaker variability has mainly been tested using words in isolation produced by a speaker the infant had never heard before (Best et al., 2009; Houston & Jusczyk, 2000; Schmale & Seidl, 2009; Schmale et al., 2010; Singh et al., 2004). Thus, infant studies have largely failed to take into account that exposure to the speaker may be beneficial for infants as well. That is, the reduced amount of speaker information available from word tokens produced in isolation may have prevented infants from accommodating to the speaker, even if in everyday life, when access to speaker idiosyncrasies is the standard, infants are able to adjust to the speaker. Providing infants with more ecologically valid listening conditions is thus crucial to uncover infants’ natural word recognition abilities used in everyday speech perception.

In three lines of research, I have shown that infants and toddlers are indeed generally better able to deal with speaker-related variability than previous work has
suggested. First, in Chapter 2, Canadian 15-month-olds were found to be unable to recognize known words in an unfamiliar, Australian accent without any speaker context. They did, however, recognize these same Australian-accented words when, prior to test, they had been able to extract relevant information about the speaker’s characteristics from a two-minute story read by the Australian speaker. Even only very brief access to the speaker’s typical way of pronouncing words thus helps infants adapt to the speaker. This enables them to overcome difficulties that had long been viewed as impossible to surpass during word recognition in unfamiliar accents. Without such exposure to the speaker, it is not until the latter half of the second year of life that infants start recognizing words in unfamiliar accents in this paradigm. By that age, the combination of more mature signal-to-word mapping abilities and more precise representations of words in the mental lexicon may have enabled children to recognize words produced in an unfamiliar Australian accent. It is, of course, possible that prior exposure to this speaker would have led to even more robust recognition levels.

In Chapter 3, I then demonstrated that 28-month-olds readily recognize words produced by a Scottish-accented speaker in a Preferential Looking Paradigm, even without any prior exposure to the speaker. Target words in this study were embedded in sentences, which may have provided the toddlers with sufficient context for interpreting the words. Given 20-month-olds’ inability to similarly succeed at recognizing words produced by an Australian-accented speaker in this task, the ability to use this immediate sentence context to deal with speaker-related variation likely develops between 20 and 28 months of age. It is plausible that to recognize and fully access the words in the study, the younger toddlers may have needed more access to speaker information prior to the study. Had 20-month-
olds been presented with a more accessible form of exposure to the accented speaker (e.g.,
easier words), they may have been able to better adapt to the speaker and recognize the
accented words. Follow-up work should further examine this possibility.

Finally, in Chapter 4, I showed that under relatively naturalistic listening
conditions, even 7.5-month-olds perform better than past studies suggest. Unlike previous
reports indicating that 7.5-month-olds experience great difficulty mapping acoustically
dissimilar tokens of the same word onto the same underlying representation, infants
presented with target words embedded in fluent speech spoken by parents were able to
build robust and arguably at least somewhat abstract word form representations.
Specifically, when trained on words forms produced by a female speaker, they later
recognized those words when produced by a male speaker, regardless of whether they were
familiar with the speakers. Taken together, these studies support the notion that infants’
early signal-to-word mapping skills flourish under relatively naturalistic listening
conditions and that infants’ early word representations are sufficiently flexible to allow for
acoustic-phonetic divergence in the realization of words.

2 Implications for models of early speech perception

The finding that infants are able to cope with speaker-related variability has important
implications for models of early speech perception. Specifically, the competence of
mapping acoustically distinct word tokens onto the same underlying word representation
requires infants’ early word representations to be abstract in nature. If word representations
were fully exemplar-like in nature, infants should have shown great difficulty recognizing
words in an unfamiliar accent, even if they had had access to the speaker characteristics.
Only prior exposure to the to be recognized words themselves in the unfamiliar accent would have helped them to subsequently recognize these words. As infants and toddlers in Chapter 2 and 3 recognized test items that did not occur in the storybook (with the 28-month-olds in Chapter 3 recognizing previously unheard words equally quickly and reliably as words that they had heard the speaker say before), it would be a challenge for exemplar-based models to provide an adequate explanation of the results. Furthermore, episodic models of speech perception would have difficulty explaining the results of Chapter 4, where infants were able to generalize word forms across speakers of different genders. According to these models, infants will only generalize across acoustic variability once they have heard a sufficient number of similarly variable word tokens. If infants’ word representations were truly episodic, the 7.5-month-olds in Chapter 4 who were trained on words in a female voice only should hence not recognize these words in an acoustically distinct male voice unless these words were previously uttered by a more variable group of speakers.

Models of speech perception allowing for prelexical speaker accommodation, in contrast, would be able to elegantly explain the current findings. According to such models, access to speaker information may help infants build the sophisticated maps between the speech signal and the linguistic units. Speaker information is not necessarily part of the representations of words in the mental lexicon and processes such as speaker adaptation take place at prelexical levels. Once sounds have been categorized, abstract linguistic patterns are sent to the lexical level, upon which lexical items can be created or accessed. This would explain why word recognition might not dependent on acoustic overlap per se, but may rather rely on infants’ ability to adapt to the speaker. In light of the
results of the current experiments, this would clarify why infants can generalize across acoustic divergence in the realization of words when these words are embedded in a type of context that infants can exploit. Abstraction processes like these are inconsistent with extreme episodic models (e.g., Goldinger, 1996). Even at the very early stages of speech processing, word representations are thus, at least to some extent, abstract.

Research on the nature of adult speech perception has indicated a similar need for abstraction (Cutler et al., 2010). If both infant and adult representations contain at least partially abstract linguistic units, this would make the transition from the infant to the adult state much more gradual than previous work has suggested. That is, even the earliest word representations infants form may already be abstract in nature. Over time, infants may become more efficient in recognizing words. In the first year of life, the linguistic representations may refine to accommodate the fact that infants are still tuning into the sound patterns of their native language. Later, once most words in the child’s mental lexicon are represented in a more mature fashion, efficiency may be increase by acquiring better adaptation mechanisms.

3 Mechanisms inducing speaker accommodation in early childhood

If speaker adaptation is evident from early on, one may ask what the mechanism is that induces these adaptation processes. The data from Chapter 2 put forward the idea that even 15-month-old infants, whose vocabularies are still limited, rely on lexical information to accommodate novel speakers. In Experiment 3 of that chapter, when infants did not have repeated exposure to the Very Hungry Caterpillar storybook prior to their participation in the study, infants were unable to accommodate the speaker. In
Experiment 4, in contrast, when infants were read the storybook once a day for two weeks and were hence presumably more familiar with the words in the story, infants did accommodate the novel speaker’s Australian accent. Without repeated exposure to the story, the story words (involving caterpillars, sausages, and cocoons) may thus have been difficult to understand for children at 15 months of age. Further evidence for this idea comes from Experiment 3 in Chapter 3, where even 20-month-olds displayed some difficulty recognizing the storybook words even when they were produced in the toddlers’ own accent. If lexical feedback is indeed crucial for speaker adaptation, this initial difficulty recognizing the words in the story may also explain why the 20-month-olds in Experiment 2 of that chapter failed to accommodate the Australian-accented speaker. It is thus possible that 20-month-olds would have adapted to the speaker had they been able to access the words in the exposure phase. In future work, I am planning to examine whether prior exposure to the story in toddlers’ own accent facilitates speaker adaptation when the story is subsequently produced by the Australian-accented speaker, much like reading the storybook at home helped 15-month-olds accommodate the same Australian speaker in Chapter 2.

If speaker adaptation is indeed lexically guided, then the question arises what strategies younger infants, such as those in Chapter 4, may use to accommodate acoustic divergence in the realization of words. 7.5-month-olds know only very few words (though see Bergelson & Swingley, 2012; Tincoff & Jusczyk, 1999; 2012 for findings that they do know some words) and most of the words used in the passages in Chapter 4 are likely unknown at this age. In addition, 7.5-months-olds are still figuring out the phoneme inventory of their language. Even if they were able to use lexical information to
figure out the mapping between the speech of different speakers, these signal-to-word maps would presumably link to rough (potentially phonologically unanalyzed) whole words. One possibility is that unlike older infants, younger infants may rely more on the global information about the speaker’s voice. For example, listening to a particular speaker may provide infants with a general idea of the speaker’s vocal range and just a rough estimation of where the speaker falls in acoustic space may be sufficient to map the speaker’s realizations of the words onto the linguistic representations. According to this view, bottom-up acoustic cues are more important than top-down lexical information at the early stages where lexical information is limited. Only when infants’ processing abilities become more mature (and involve the language-specific representations of phonemic and lexical units), will adaptation become lexically mediated.

An alternative option is that speaker information is lexically guided when dealing with accents, but that accommodating variability in speaker gender or affect would not be lexically mediated. Whereas accent differences typically involve shifts in the realization of multiple speech sounds, differences in the gender or affect of the speaker take place at levels that are linguistically less relevant (in English, changes in pitch, pitch level, and energy do not alter the linguistic identity of a word). Changing the gender of a speaker, for example, mainly results in pitch level differences in the realization of words. A single directional shift in a single (linguistically irrelevant) dimension of one word token towards another one would cause the two word tokens to sound roughly similar. Differences in accents, on the other hand, are much less unidimensional and mapping two words tokens produced by speakers of different accents might hence likely be much more complex. Lexical access may only be necessary to contend with the more challenging.
type of phonetic discrepancies, where listeners need to shift their signal-to-phoneme maps. Although this dissertation does not distinguish between these two possibilities, the idea that coping with differences in the realizations of words may be easier across speakers of the same accent than across speakers of different accents is supported by findings showing that young infants are able to overcome pitch differences during word recognition (Singh, White, & Morgan, 2008) more readily than accent differences (Schmale & Seidl, 2009; Schmale et al., 2010).

4 Cross-speaker generalization

The studies in this dissertation take a first step at looking at whether and how young children can overcome previously observed difficulties dealing with variability in the realization of words. The finding that infants and toddlers can adapt to speakers of different accents is impressive and gives rise to a whole body of follow-up questions to consider. The first question to ask would be how broadly this adaptation is applied. While the discrepancy between two speakers of different accents is typically greater than the discrepancies between two speakers of the same accent, each individual speaker of an accent has their idiosyncratic way of realizing their phonemes. Nonetheless, adult listeners can often readily group people based on their accent despite the between-speaker differences that surface within the same accent (e.g., Clopper & Pisoni, 2004). This raises the question of whether listeners’ adaptation to an unfamiliar accent generalizes to other speakers of that accent they have never heard before. A recent study aimed to address this question by presenting listeners with non-native Chinese-accented speakers of English (Bradlow & Bent, 2008). Much like the infants in Chapter 2, exposure to a particular
speaker improved adults’ subsequent word recognition for words uttered by that speaker. It did not, however, improve word recognition for other speakers of the same language background, who therefore had a similar accent. It is thus possible that any between-speaker differences may be too large to be overcome, and speaker exposure may only enable listeners to cope with the smaller within-speaker differences in the realization of words. Extrapolating from this reasoning, I may not have found an improved ability for word recognition had I given the infants in Chapter 2 exposure to a different Australian-accented speaker reading the story than the one used in the test phase.

There are, however, at least three different factors that might modulate listeners’ ability to generalize across different speakers of the same accent. First, speakers in the Bradlow and Bent (2008) study were non-native speakers of English. Non-native speakers often differ in proficiency. And even if the intelligibility between two speakers is kept constant (as is the case for one of the speaker sets in that study), the exact manifestation of the accent can differ dramatically between two non-native speakers, even if they share the same language background (see Derwig & Munro, 1997 for findings showing that, at least for adults, intelligibility of non-native speech is not always fully correlated with perceived accentedness). Accent variation between native speakers of a given language, in contrast, tends to be much more homogeneous (cf. Floccia et al., 2006). This decrease in similarity between non-native speakers’ accents may have blocked generalization, even if listeners would generalize across speakers with the same native regional variant. Two non-native English speakers with Chinese as their first language may, in other words, be perceived to be as different as an Australian- and a Scottish-accent speaker, or, in less severe cases as an Australian- and a New Zealand-
accented speaker. It would thus be interesting to see whether listeners generalize their adapted signal-to-word maps from one native-accented speaker to other speakers of that same native accent. If so, the next step would be to examine the scope of this adaptation. How far can it be stretched? Would adaptation to an Australian-accented speaker, for example, spill over to a New Zealand-accented speaker? And would this differ between children and adults?

Previous research indicates that non-native speakers are not the only group for whom listeners do not show evidence of generalization. Adaptation to native speakers who consistently produce one phoneme (carrying spectral information) in an ambiguous fashion is also speaker-specific (Eisner & McQueen, 2005; Kraljic & Samuel, 2007; Trude & Brown-Schmidt, 2012). This may raise the question of whether this selective adaptation observed in these studies undermines the possibility that listeners generalize their signal-to-word maps across native speakers of a previously unfamiliar language. This does not necessarily need to be the case. In natural accents, there are typically many cues that listeners may use to determine what accent group a speaker belongs to. Besides the greater number of deviating speech segments (different accents generally diverge in more than just a single phoneme), natural accents also differ from ‘single shift alternations’ in prosodic features such as rhythm and intonation. This likely facilitates the realization that the speaker speaks with a specific accent and may hence promote full generalization. In the case of a single ambiguous phoneme, in contrast, much less evidence is available and generalization may not immediately take place. This makes the salience of the accent a second factor to possibly modulate listeners’ ability to generalize across different speakers of the same accent.
A third factor moderating generalization concerns the presence of speaker variance. Although exposure to a single Chinese-accented speaker did not induce generalization to another Chinese-accented speaker, exposure to multiple different Chinese-accented English speakers did allow listeners to generalize to a novel Chinese-accented speaker (Bradlow & Bent, 2008). In fact, recognition levels after exposure to multiple Chinese-accented speakers was equally advantageous as exposure to the same speaker used in the test phase. Thus, hearing a variety of different speakers may generate a more abstract, speaker-independent form of adaptation compared to hearing just a single speaker. Even if exposure to a single speaker may not generate speaker-general accent adaptation, exposure to multiple speakers likely will, at least in adults.

To my knowledge, no studies to date have examined the scope of accent adaptation in early childhood. To gain a better understanding of the development of speaker adaptation, future research should address this issue. Differences between infant and adult listeners in the extent to which they allow for generalization when listening to speakers who have similar accents to (but are unambiguously different from) other speakers they have encountered would be greatly informative for mapping out the developmental pattern from child listeners to adult language users. In addition, such studies would allow a more thorough grasp on the nature of infant speech perception and early word representations.

5 Concluding remarks

The three lines of research presented in this dissertation provide evidence for the view that even early in development, young children can contend with speaker and accent variation.
While acoustic variability may indeed pose a challenge to infant listeners, this is not necessarily a challenge they cannot overcome. In fact, when tested under more everyday-like listening conditions, young children demonstrate the possession of adaptation abilities similar to those of adults. This bodes well for their abilities to tune in to new speakers with an unfamiliar accent, for example when they start daycare or watch an international TV show.

In sum, the findings in this dissertation suggest that infants’ early speech perception abilities are much more sophisticated than previously thought. In addition, the perceptual constancy for speech, even at the early stages of development, suggests that infants’ word representations are sufficiently abstract to deal with acoustic variability in the realization of words. The transformation of word recognition from infancy into adulthood may thus be less abrupt than current theories of early speech perception suggest.
References


Appendices

Appendix A: Language background questionnaire administered to the participating families in Chapters 2 and 3

1. What percentage of the time does your child hear English spoken by people who learned English in Ontario?
2. What percentage of the time does your child hear other languages or variants of English?
3. How old were you when you first learned English?
4. How old were any other primary caregivers when they first learned English?
5. How would you describe the variant of English you speak (e.g. Southern Ontario, Australian, Singapore, etc?)
6. How would you describe the variant of English other primary caregivers speak?
7. Do you speak (or have you spoken in the past) any other languages besides English?
8. Do other primary caregivers speak (or have they spoken in the past) any other languages besides English?
9. Do you consider English your dominant language? If not, please explain.
10. Do other primary caregivers consider English their dominant language? If not, please explain.
11. Please specify when you first learned the other language(s), and in what context you use(d) the other language(s).
13. Does your child attend daycare? How often? What is the dominant language at daycare?

14. What language(s) do your friends and family members who spend the most time with your child speak? What language/dialect do they generally speak in your child's presence? How much time do they spend with your child?

15. Does your child listen to Australian-accented (Chapter 3: or Scottish-accented) English? If yes: by whom (include TV shows)? Also, please explain how often/how many hours a week.

16. Do you travel often with your child to locations where Canadian English is not the dominant language/dialect? How much time do you spend in those locations?

17. Do you live in a neighborhood where a language besides Canadian English is the dominant language/dialect? Please explain.

18. Do you expect English to be your child's only language? If not, do you expect it to be your child's dominant language?

19. Can you please confirm the percentage of Canadian English your child hears?

20. (Chapter 3 only) Does your child know “The very hungry caterpillar” story? How familiar do you think s/he is with the story on a scale from 1-7?
Appendix B: Lists of words used in Chapter 2

Known words

list 1
dog – grandma – cup – daddy

grandma – dog – kitty – shoe

list 3
mommy – kiss – bottle –
cup – mommy – bath – diaper – kiss

Nonsense words

list 1
dapper – brall – dimma – bap – koddy

list 2
dapper – brall – mitty – deuce

list 3
dapper – brall – dimma –

list 4
shammy – bog – guttle – kie
Appendix C: Pictures used in Chapter 3

Pictures used in heard word trials

butterfly

cake

cheese

strawberry
Pictures used in generalization trials

- ball
- boat
- book
- cat
- car
- cow
cup

dog

duck

fork

soup

toast
**Appendix D: Video orders used in Chapter 3**

**list 1**

<table>
<thead>
<tr>
<th>Trial</th>
<th>Type</th>
<th>Picture on left</th>
<th>Picture on right</th>
<th>Target word</th>
<th>Auditory stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>G</td>
<td>dog</td>
<td>duck</td>
<td>dog</td>
<td>Look! Where's the dog? How cute!</td>
</tr>
<tr>
<td>2</td>
<td>G</td>
<td>boat</td>
<td>ball</td>
<td>ball</td>
<td>Wow! Look at the ball! Do you like it?</td>
</tr>
<tr>
<td>3</td>
<td>G</td>
<td>book</td>
<td>fork</td>
<td>fork</td>
<td>Aww! Where's the fork? Fantastic, eh?</td>
</tr>
<tr>
<td>4</td>
<td>H</td>
<td>butterfly</td>
<td>strawberry</td>
<td>butterfly</td>
<td>Hey! Look at the butterfly! Can you find it?</td>
</tr>
<tr>
<td>5</td>
<td>G</td>
<td>cow</td>
<td>cat</td>
<td>cat</td>
<td>Aww! Look at the cat! Isn't it pretty?</td>
</tr>
<tr>
<td>6</td>
<td>H</td>
<td>cheese</td>
<td>cake</td>
<td>cake</td>
<td>Look! Where's the cake? Can you see it?</td>
</tr>
<tr>
<td>7</td>
<td>G</td>
<td>car</td>
<td>cup</td>
<td>car</td>
<td>Hey! Where's the car? Amazing, eh?</td>
</tr>
<tr>
<td>8</td>
<td>G</td>
<td>soup</td>
<td>toast</td>
<td>soup</td>
<td>Wow! Look at the soup! What do you think?</td>
</tr>
<tr>
<td>9</td>
<td>G</td>
<td>dog</td>
<td>duck</td>
<td>duck</td>
<td>Look! Where's the duck? How cute!</td>
</tr>
<tr>
<td>10</td>
<td>G</td>
<td>boat</td>
<td>ball</td>
<td>boat</td>
<td>Wow! Look at the boat! Do you like it?</td>
</tr>
<tr>
<td>12</td>
<td>H</td>
<td>butterfly</td>
<td>strawberry</td>
<td>strawberry</td>
<td>Hey! Look at the strawberry! Can you find it?</td>
</tr>
<tr>
<td>13</td>
<td>G</td>
<td>cow</td>
<td>cat</td>
<td>cow</td>
<td>Aww! Look at the cow! Isn't it pretty?</td>
</tr>
<tr>
<td>14</td>
<td>H</td>
<td>cheese</td>
<td>cake</td>
<td>cheese</td>
<td>Look! Where's the cheese? Can you see it?</td>
</tr>
<tr>
<td>15</td>
<td>G</td>
<td>car</td>
<td>cup</td>
<td>cup</td>
<td>Hey! Where's the cup? Amazing, eh?</td>
</tr>
<tr>
<td>16</td>
<td>G</td>
<td>soup</td>
<td>toast</td>
<td>toast</td>
<td>Wow! Look at the toast! What do you think?</td>
</tr>
</tbody>
</table>

*H = heard word trials containing storybook words; G = generalization trials*
### list 2

<table>
<thead>
<tr>
<th>Trial</th>
<th>Type</th>
<th>Picture on left</th>
<th>Picture on right</th>
<th>Target word</th>
<th>Auditory stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>G</td>
<td>dog</td>
<td>duck</td>
<td>duck</td>
<td>Look! Where's the duck? How cute!</td>
</tr>
<tr>
<td>2</td>
<td>G</td>
<td>boat</td>
<td>ball</td>
<td>boat</td>
<td>Wow! Look at the boat! Do you like it?</td>
</tr>
<tr>
<td>4</td>
<td>H</td>
<td>butterfly</td>
<td>strawberry</td>
<td>strawberry</td>
<td>Hey! Look at the strawberry! Can you find it?</td>
</tr>
<tr>
<td>5</td>
<td>G</td>
<td>cow</td>
<td>cat</td>
<td>cow</td>
<td>Aww! Look at the cow! Isn't it pretty?</td>
</tr>
<tr>
<td>6</td>
<td>H</td>
<td>cheese</td>
<td>cake</td>
<td>cheese</td>
<td>Look! Where's the cheese? Can you see it?</td>
</tr>
<tr>
<td>7</td>
<td>G</td>
<td>car</td>
<td>cup</td>
<td>cup</td>
<td>Hey! Where's the cup? Amazing, eh?</td>
</tr>
<tr>
<td>8</td>
<td>G</td>
<td>soup</td>
<td>toast</td>
<td>toast</td>
<td>Wow! Look at the toast! What do you think?</td>
</tr>
<tr>
<td>9</td>
<td>G</td>
<td>dog</td>
<td>duck</td>
<td>dog</td>
<td>Look! Where's the dog? How cute!</td>
</tr>
<tr>
<td>10</td>
<td>G</td>
<td>boat</td>
<td>ball</td>
<td>ball</td>
<td>Wow! Look at the ball! Do you like it?</td>
</tr>
<tr>
<td>11</td>
<td>G</td>
<td>book</td>
<td>fork</td>
<td>fork</td>
<td>Aww! Where's the fork? Fantastic, eh?</td>
</tr>
<tr>
<td>12</td>
<td>H</td>
<td>butterfly</td>
<td>strawberry</td>
<td>butterfly</td>
<td>Hey! Look at the butterfly! Can you find it?</td>
</tr>
<tr>
<td>13</td>
<td>G</td>
<td>cow</td>
<td>cat</td>
<td>cat</td>
<td>Aww! Look at the cat! Isn't it pretty?</td>
</tr>
<tr>
<td>14</td>
<td>H</td>
<td>cheese</td>
<td>cake</td>
<td>cake</td>
<td>Look! Where's the cake? Can you see it?</td>
</tr>
<tr>
<td>15</td>
<td>G</td>
<td>car</td>
<td>cup</td>
<td>car</td>
<td>Hey! Where's the car? Amazing, eh?</td>
</tr>
<tr>
<td>16</td>
<td>G</td>
<td>soup</td>
<td>toast</td>
<td>soup</td>
<td>Wow! Look at the soup! What do you think?</td>
</tr>
</tbody>
</table>

_H = heard word trials containing storybook words; G = generalization trials_
### list 3

<table>
<thead>
<tr>
<th>Trial</th>
<th>Type</th>
<th>Picture on left</th>
<th>Picture on right</th>
<th>Target word</th>
<th>Auditory stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>G</td>
<td>cup</td>
<td>car</td>
<td>car</td>
<td>Aww! Look at the car! Can you find it?</td>
</tr>
<tr>
<td>2</td>
<td>G</td>
<td>cat</td>
<td>cow</td>
<td>cat</td>
<td>Look! Where's the cat? How cute!</td>
</tr>
<tr>
<td>3</td>
<td>H</td>
<td>cake</td>
<td>cheese</td>
<td>cheese</td>
<td>Wow! Look at the cheese! Amazing, eh?</td>
</tr>
<tr>
<td>4</td>
<td>G</td>
<td>toast</td>
<td>soup</td>
<td>toast</td>
<td>Hey! Where's the toast? Do you like it?</td>
</tr>
<tr>
<td>5</td>
<td>G</td>
<td>ball</td>
<td>boat</td>
<td>ball</td>
<td>Look! Where's the ball? Fantastic, eh?</td>
</tr>
<tr>
<td>6</td>
<td>G</td>
<td>duck</td>
<td>dog</td>
<td>dog</td>
<td>Aww! Look at the dog! Isn't it pretty?</td>
</tr>
<tr>
<td>7</td>
<td>G</td>
<td>fork</td>
<td>book</td>
<td>book</td>
<td>Hey! Look at the book! What do you think?</td>
</tr>
<tr>
<td>8</td>
<td>H</td>
<td>strawberry</td>
<td>butterfly</td>
<td>strawberry</td>
<td>Wow! Where's the strawberry? Can you see it?</td>
</tr>
<tr>
<td>9</td>
<td>G</td>
<td>cup</td>
<td>car</td>
<td>cup</td>
<td>Aww! Look at the cup! Can you find it?</td>
</tr>
<tr>
<td>10</td>
<td>G</td>
<td>cat</td>
<td>cow</td>
<td>cow</td>
<td>Look! Where's the cow? How cute!</td>
</tr>
<tr>
<td>11</td>
<td>H</td>
<td>cake</td>
<td>cheese</td>
<td>cake</td>
<td>Wow! Look at the cake! Amazing, eh?</td>
</tr>
<tr>
<td>12</td>
<td>G</td>
<td>toast</td>
<td>soup</td>
<td>soup</td>
<td>Hey! Where's the soup? Do you like it?</td>
</tr>
<tr>
<td>13</td>
<td>G</td>
<td>ball</td>
<td>boat</td>
<td>boat</td>
<td>Look! Where's the boat? Fantastic, eh?</td>
</tr>
<tr>
<td>14</td>
<td>G</td>
<td>duck</td>
<td>dog</td>
<td>duck</td>
<td>Aww! Look at the duck! Isn't it pretty?</td>
</tr>
<tr>
<td>15</td>
<td>G</td>
<td>fork</td>
<td>book</td>
<td>fork</td>
<td>Hey! Look at the fork! What do you think?</td>
</tr>
<tr>
<td>16</td>
<td>H</td>
<td>strawberry</td>
<td>butterfly</td>
<td>butterfly</td>
<td>Wow! Where's the butterfly? Can you see it?</td>
</tr>
</tbody>
</table>

*H = heard word trials containing storybook words; G = generalization trials*
<table>
<thead>
<tr>
<th>Trial</th>
<th>Type</th>
<th>Picture on left</th>
<th>Picture on right</th>
<th>Target word</th>
<th>Auditory stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>G</td>
<td>cup</td>
<td>car</td>
<td>cup</td>
<td>Aww! Look at the cup! Can you find it?</td>
</tr>
<tr>
<td>2</td>
<td>G</td>
<td>cat</td>
<td>cow</td>
<td>cow</td>
<td>Look! Where's the cow? How cute!</td>
</tr>
<tr>
<td>3</td>
<td>H</td>
<td>cake</td>
<td>cheese</td>
<td>cake</td>
<td>Wow! Look at the cake! Amazing, eh?</td>
</tr>
<tr>
<td>4</td>
<td>G</td>
<td>toast</td>
<td>soup</td>
<td>soup</td>
<td>Hey! Where's the soup? Do you like it?</td>
</tr>
<tr>
<td>5</td>
<td>G</td>
<td>ball</td>
<td>boat</td>
<td>boat</td>
<td>Look! Where's the boat? Fantastic, eh?</td>
</tr>
<tr>
<td>6</td>
<td>G</td>
<td>duck</td>
<td>dog</td>
<td>duck</td>
<td>Aww! Look at the duck! Isn't it pretty?</td>
</tr>
<tr>
<td>7</td>
<td>G</td>
<td>fork</td>
<td>book</td>
<td>fork</td>
<td>Hey! Look at the fork! What do you think?</td>
</tr>
<tr>
<td>8</td>
<td>H</td>
<td>strawberry</td>
<td>butterfly</td>
<td>butterfly</td>
<td>Wow! Where's the butterfly? Can you see it?</td>
</tr>
<tr>
<td>9</td>
<td>G</td>
<td>cup</td>
<td>car</td>
<td>car</td>
<td>Aww! Look at the car! Can you find it?</td>
</tr>
<tr>
<td>10</td>
<td>G</td>
<td>cat</td>
<td>cow</td>
<td>cat</td>
<td>Look! Where's the cat? How cute!</td>
</tr>
<tr>
<td>11</td>
<td>H</td>
<td>cake</td>
<td>cheese</td>
<td>cheese</td>
<td>Wow! Look at the cheese! Amazing, eh?</td>
</tr>
<tr>
<td>12</td>
<td>G</td>
<td>toast</td>
<td>soup</td>
<td>toast</td>
<td>Hey! Where's the toast? Do you like it?</td>
</tr>
<tr>
<td>13</td>
<td>G</td>
<td>ball</td>
<td>boat</td>
<td>ball</td>
<td>Look! Where's the ball? Fantastic, eh?</td>
</tr>
<tr>
<td>14</td>
<td>G</td>
<td>duck</td>
<td>dog</td>
<td>dog</td>
<td>Aww! Look at the dog! Isn't it pretty?</td>
</tr>
<tr>
<td>16</td>
<td>H</td>
<td>strawberry</td>
<td>butterfly</td>
<td>strawberry</td>
<td>Wow! Where's the strawberry? Can you see it?</td>
</tr>
</tbody>
</table>

*H = heard word trials containing storybook words; G = generalization trials*
Appendix E: Passages used in Chapter 4

Her boat had white sails. This girl will steer my big boat. That horn on the boat was really loud. He bought himself a new red boat. His boat could go quite fast. We always store your boat in our garage.

Your toque was soft and warm. She wore a red toque in the snow. Their brother had knitted this big toque. She liked how her toque covered my ears. Our friends also fancied the toque. His toque was blue and green.

Your pear came from my fridge. She washed her pear thoroughly. They wanted to eat a red pear. The pear in our basket looked good. Next to his pear was an apple. He enjoyed eating this big pear.

The cup was bright and shiny. A clown drank from that big cup. Some milk from his cup spilled on our rug. Your cup was filled with hot milk. They put her cup on their table. She then picked up a red cup.