The Development of Intonation and Information Structure:

Perceptual and Productive Investigations

by

Jill Caroline Thorson

B.A., University of Rochester, 2005

M.A., University of Rochester, 2007

Submitted in partial fulfillment of the requirements

for the Degree of Doctor of Philosophy in the

Department of Cognitive, Linguistic, and Psychological Sciences at Brown University

Providence, Rhode Island

May 2015

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

Date _____ _____

James Morgan, Director

Recommended to the Graduate Council

Date _____ _____

Laura Kertz, Reader

Date _____ _____

Pilar Prieto, Reader
(Universitat Pompeu Fabra, Barcelona)

Approved by the Graduate Council

Date _____ _____

Peter Weber
Dean of the Graduate School

Curriculum Vitae

Jill Thorson was born in Hanover, New Hampshire, on January 8. She received a Bachelor of Arts in Linguistics and Spanish in 2005 and a Master of Arts in Linguistics in 2007 from the University of Rochester. Her Master's thesis was entitled *A Description of Declarative Intonation Patterns in Puerto Rican Spanish*. Following her Master's degree, she received a Fulbright Fellowship to Spain and conducted research for two years as part of the Grup d'Estudis de Prosòdia at the Universitat Autònoma de Barcelona. She received a Master of Science in Cognitive Science from Brown University in 2012 with the thesis *Syllabic Organization in Children's Early Words: Acoustic and Articulatory Investigations*. She has had her research on prosody and language acquisition published in several peer-reviewed journals. Additionally, she has worked as a teaching assistant in phonetics and phonology, general linguistics, psycholinguistics, and cognition at the University of Rochester and Brown University.

Grants and Awards

Linguistic Society of America Summer Institute Tuition Fellowship	2011
First Year Fellowship, Brown University	2009
Fulbright Research Fellowship (with Grant Extension) to Barcelona, Spain	2007
Acoustical Society of America Student Travel Award	2011
Fellowship to attend the São Paolo School of Advanced Study in Speech Dynamics	2010
Phi Beta Kappa Honor Society	2005
Summer Reach Funding, University of Rochester	2004
Dean's List, University of Rochester	2001 – 2005
Bausch & Lomb Science Award, University of Rochester	2001
Rush Rhees Scholarship, University of Rochester	2001

Publications

J. Thorson and J. L. Morgan. (2015). Acoustic correlates of information structure in child and adult speech. In *Proceedings of the 39th Annual Boston University Conference on Language Development.* Somerville: Cascadilla Press.

J. Thorson and J. L. Morgan. (2014). Directing Toddler Attention: Intonation and Information Structure. In *Supplement to the Proceedings of the 38th Annual Boston University Conference on Language Development.* Somerville: Cascadilla Press.

J. Thorson, J. Borràs-Comes, V. Crespo-Sendra, M.M. Vanrell, and P. Prieto. (2014). The acquisition of melodic form and meaning in yes-no interrogatives by Catalan and Spanish speaking children. *Probus, International Journal of Latin and Romance Linguistics.*

P. Prieto, A. Estrella, J. Thorson, and M.M. Vanrell. (2012). Is prosodic development correlated with grammatical development? Evidence from emerging intonation in Catalan and Spanish. *Journal of Child Language, 39*(2), 221-257.

P. Prieto, J. Borràs-Comes, V. Crespo-Sendra, J. Thorson, and M.M. Vanrell. (2011). Entonación y pragmática en los enunciados interrogativos del español en habla dirigida a niños. Número especial sobre "La interfaz Entonación-Discurso oral en el ámbito hispánico." *Oralia*, *14*, 227-255.

P. Prieto, J. Thorson, A. Estrella, and M.M. Vanrell. (2010). Early Intonational Development in Spanish: A Case Study. *Selected Proceedings of the 13th Hispanic Linguistics Symposium*. Somerville: Cascadilla Press.

J. Thorson. (2007). The Scaling of Utterance-Initial Pitch Peaks in Puerto Rican Spanish: Evidence for Tonal Preplanning. In L. Wolter and J. Thorson (Eds.), *University of Rochester Working Papers in the Language Sciences*, *3*(1), 91-97.

Acknowledgements

I would first like to thank my advisor, Jim Morgan. He has been a great supporter and advocate for me since the first day we met. His ability to find the big picture in my work is truly amazing and he has a way of synthesizing material that is unprecedented. I am grateful for his mentorship and for all that he has done to make me a better writer, presenter, and researcher. I would also like to thank my committee members. Laura Kertz was immediately supportive upon her arrival at Brown University and provided invaluable insight into my research. Pilar Prieto welcomed me into her lab when I emailed her out of the blue nearly eight years ago to sponsor me for a Fulbright. She has been an amazing mentor and collaborator ever since and has served as my intonation guru throughout my dissertation. I thank her for keeping me on my 'prosodic toes' and for all things Catalan that have shaped my life: Moltes gràcies. Thanks also to Katherine Demuth for her guidance during the beginning of my time at Brown, and to Sheila Blumstein and Rachel Theodore for their mentorship.

A heartfelt thank you to my undergraduate and Master's advisor, Joyce McDonough. I took her class *People and their Language* on a whim freshman year and it truly changed the course of my life. Her enthusiasm and dedication to language and its speakers are inspiring and have taught me to remember *why* it is we study language and how essential it is to understanding human culture.

Thanks so much to everyone in the Infant Lab. Lori Rolfe has been an incredible friend and savior on an everyday level and I thank her for her dedication to the lab and

all of its members. Without her, I do now know how I would have been able to get those busy toddlers to complete any of my studies. Thanks to my lab "sisters" past and present, including Glenda, Jie, Elena L., Naomi, Elena T., Erin, and Katherine. Also, thank you to all of the GrEPets in Catalunya who have become both my colleagues and friends.

My amazing friends near and far have been a constant support system over the past decade (and some for decades longer than that). Thank you to my friends at Brown (Glenda, Deanna, Julie, Sara, Kevin) and to those who sent care packages, words of inspiration, and support in every form of communication possible (Jen, Mariah, Carrie, Jen, Rachel, Caitlyn, Carlos, Ed, Hilary). Also, anyone who knows me well knows I am nothing if not defined by my love of animals. A special thanks goes out to my furry four-legged support system.

A big thank you to Paul for keeping me nourished on long days, taking long bike rides and runs with me to relax, and for getting me to fully enjoy any free time I was able to scrounge together. Anyone who is willing to proofread an entire dissertation must either be crazy or the best partner in the world (probably both). Finally, I cannot possibly thank my mom, dad, and sister enough. They have been with me through all of the ups and downs, and not just the ones of the last five years. Also, a big welcome to the new addition in our family, my nephew Spencer, who was born just one month before I defended. Daily pictures and videos of him kept me motivated right up until the end. I love you all and I am lucky to have you as my family.

Contents

List of Tables

List of Figures

Chapter 1

Introduction and Background

1.1 Introduction

Within just a few days of birth, newborn infants are sensitive to their native language prosody and rhythm (Nazzi, Bertoncini, & Mehler, 1998; Nazzi, Floccia, & Bertoncini, 1998). When children begin to speak, they are able to approximate adult-like intonation contour patterns (Prieto, & Vanrell, 2007; Chen & Fikkert, 2007; Frota & Vigário, 2008) and to align these contours with the appropriate semantic meanings and pragmatic intentions (Prieto, Estrella, Thorson, & Vanrell, 2012). However, little research has been conducted on early sensitivities to intonation as it reflects information structure. Furthermore, few studies have analyzed child speech productions both phonologically and acoustic-phonetically within this domain. To redress this previously under-studied aspect of child speech, in what follows I conduct a more detailed investigation into the development of intonation and information structure in order to better understand the mechanisms that lead to successful communication. In this dissertation, I explore the interaction between particular pitch accents and information status in several experiments to show the fundamental role they play in the language acquisition of toddlers, which builds upon previous research to demonstrate that children are able to incorporate these features during early perceptual processing as well as make these distinctions in their own early speech.

Infants must distinguish a number of acoustic features encoded in the speech signal. One of these features is pitch, represented best by the fundamental frequency (F_0). Some aspects of language that employ pitch include stress, tone, intonation, and

affect. Intonation, or the melody of speech, integrates pitch, duration, loudness (or intensity), and segmental quality all together in order to create phonological distinctions (i.e., distinctions in the sound structure of a language that create meaningful differences). A language-learning infant must tease apart the different types of information transmitted by pitch and these other acoustic properties (see §1.2 for a discussion on the definitions of intonation and information structure). Correctly interpreting the acoustic signal and the intonation system hugely influences how the infant or child perceives, produces, and comprehends the correct mapping of intonation to information structure in their own communicative interactions. This mapping is essential for successful word recognition, reference resolution, and word learning, and subsequently, for making semantic and pragmatic distinctions during speech production.

The primary objective of this dissertation is to investigate the role of higher-level linguistic processes during first language acquisition. Specifically, it explores the early development of intonation and information structure from two perspectives: perception and production. For perception, I examine how intonation interacts with different types of information in discourse to guide attention and aid in early word learning. For production, I explore how young toddlers produce information structure distinctions in their own speech, both intonationally and acoustically (i.e., both a phonological and a phonetic account). In order for an intonational analysis to be complete, it is essential that it include both of these components, providing a more complete prosodic analysis

that incorporates phonology and phonetics together. The primary questions that guide this research include,

(1) How does intonation and information structure interact to guide attention?

(2) What role does the isolated acoustic correlate of fundamental frequency have in guiding attention?

(3) Do specific aspects of the intonation and information structure aid in early word learning?

(4) Do young toddlers make information structure distinctions in the same way as adults? How does toddler speech phonologically and acoustic-phonetically compare to adult speech?

This dissertation consists of five chapters. Chapter 1 gives an overview of the terminology as well as provides a discussion of the background literature on the development of intonation and information structure in first language acquisition. The next two chapters concern speech perception. Chapter 2 (Experiment 1) examines how pitch type interacts with information status to guide attention at 18-months. Chapter 3 (Experiment 2) investigates how two-year-olds incorporate pitch accent type with information status during a novel word-learning task. Chapter 4 (Experiment 3) focuses on speech production, using both phonological and acoustic-phonetic analyses to explore how two and a half year olds (in comparison to adults) intonationally mark nouns of different information structure categories. The fifth and final chapter provides a general

discussion of the findings, implications of the research, and future lines of potential inquiry.

1.2 Defining Intonation and Information Structure

Before a full discussion on the development of intonation and information structure can begin, care must be taken to define the foundational elements and the primary theories that surround these topics. Additionally, defining these terms motivates the use of the particular theories that are employed in this set of studies.

1.2.1 Intonation

Intonation is characterized as the melody or cadence of speech. Narrowly, it is defined as the acoustic realization of the fundamental frequency (F_0), or pitch. For the purposes of this dissertation, I employ the broader definition that is similar to the term *prosody*. Not only does this broader definition include pitch, but it also includes the acoustic correlates of duration (timing), loudness (intensity), and segmental quality and reduction (Lehiste, 1970; Bolinger, 1989; Shattuck-Hufnagel & Turk, 1996, and others).

English, like other Germanic languages such as Dutch and German, is considered an intonation-based language and uses pitch to encode affective, semantic, and pragmatic meaning (see Ladd, 1996, for a full discussion on intonational phonology). Other languages, such as Mandarin and Vietnamese, have the additional ability to use pitch to convey lexical differences. Several theoretical and methodological approaches

studying adult language have evolved over the past century in order to study intonation from a phonological standpoint (Palmer, 1922; Cruttenden, 1986; Pierrehumbert, 1980). Only recently has this work begun to expand from adult speech analyses to ones that also include child speech (Snow, 2006; Prieto & Vanrell, 2007; Chen & Fikkert, 2007; Frota & Vigário, 2008; Prieto et al., 2012).

A large portion of the research in intonational phonology has used a theoretical framework of intonation known as the Nuclear Tone approach or the Standard British model (Palmer, 1922; Cruttenden, 1986). Over the last several decades, however, a competing framework has been employed to study intonation: the Autosegmental-Metrical (AM) approach (Pierrehumbert 1980; Beckman & Pierrehumbert, 1986; Ladd, 1996; Gussenhoven, 2004; Jun, 2005; among others). This framework treats intonational units as abstract phonological categories, and keeps them separate from their phonetic implementations. The AM framework is now the most widely accepted approach for studying intonational phonology.

The AM approach proved to be a reliable system for the evaluation of intonational contours with respect to semantic meaning and pragmatic intent (Beckman & Pierrehumbert, 1986). This approach utilizes the Tone and Break Indices (ToBI) annotation system to identify, label, and analyze intonational contours (Silverman et al., 1992; Beckman & Ayers, 1997). Since the creation of the original transcription system, it has been modified for a number of distinct languages (e.g., Gr_ToBI for Greek, Arvaniti & Baltazani, 2000; Cat_ToBI for Catalan, Prieto, Aguilar, Mascaró, Torres-Tamarit, &

Vanrell, 2009; Sp_ToBI for Spanish, Estebas-Vilaplana & Prieto, 2008). Both the Nuclear Tone approach and the AM approach are discussed in further detail in §1.2.1.1 and §1.2.1.2.

1.2.1.1 Nuclear Tone Theory

A brief discussion of the Nuclear Tone system is needed to motivate the use of the AM framework for this dissertation (Palmer, 1922; Cruttenden, 1986). The Nuclear Tone model takes a relatively holistic approach for analyzing intonation, with the final part of the utterance as the primary unit of analysis (i.e., the nuclear contour/tone is associated primarily with the last stressed syllable until the end of the utterance. In the ToBI system it refers to the combination of the final pitch accent, final phrase accent, and boundary tone). This system involves concepts such as *nucleus*, *pre-head*, *head*, *tail* and *tone unit* (versus the ToBI system which does not directly involve these features). The *tone unit* in this model is the basic unit of intonation and has a consistent internal structure. The nuclear tone is obligatorily in a tone unit, while the pre-nuclear positions are optionally included. Thus, it is the nuclear tone in this system that is described generally for its overall rising or falling shape. For annotating such contours, the approach uses descriptions such as *fall*, *rise-fall*, *fall-rise*, *high-rise*, etc.

In the developmental research, Snow (2006) and Balog and Snow (2007) have applied this system to label children's early speech in terms of their overall contour shape and general semantic/pragmatic content. Although this system usefully provides a structure for analyzing the overall direction of the nuclear contour, its more general

descriptive account does not take into consideration the level of detail that may be present in the intonational system of a language. In comparison to the AM framework, the Nuclear Tone model is limited in the ways in which it can describe a contour shape, particularly in reference to whether or not the description includes *pre-nuclear* patterns in addition to *nuclear* ones.

1.2.1.2 AM Framework and ToBI Transcription System

The AM model has had considerable success in accounting for cross-linguistic differences in adult speech. More recently, several groups of researchers have started to successfully apply the AM framework to investigate early intonation patterns in child speech as well (Prieto & Vanrell, 2007, for Catalan; Chen & Fikkert, 2007, for Dutch; Frota & Vigário, 2008, for European Portuguese; Prieto et al., 2012, for Spanish). It is essential that children learn their language-specific inventory of distinct intonation contours in order to sound like native speakers. Given this fact, being able to more intricately describe the organization of intonation aids in understanding which aspects of the prosodic system infants and children are able to understand and eventually produce.

In the AM approach, intonation is analyzed as a sequence of low (L) and high (H) pitch targets that are autosegmentally associated with the prosodic/metrical structure (i.e., with the syllables, words, phrases). The ToBI annotation system follows this framework and labels intonation via a system of monotonal and bitonal pitch accents and boundary tones (e.g., Pitch accents: L*, H*, L+H*, L*+H, etc.; Boundary tones: L%, H%, etc.) (Pierrehumbert, 1980; Beckman & Ayers, 1997). The strengths of

the breaks between words also are marked on a scale from 0 to 4 (e.g., 0 = no break; 4 = intonational phrase boundary). A benefit of this system is that it allows for a more phonologically detailed representation of the metrical structure and intonation contours due to its wide inventory of pre-nuclear and nuclear pitch accents as well as boundary tones.

The AM approach and the ToBI labeling system are not without their limitations. Like all coding systems that rely on human transcribers, there is an intrinsic level of subjectivity that enters into the data. One way that subjectivity is decreased is through the use of detailed guidelines that continue to be updated over time and after further research (Beckman & Ayers, 1997; Beckman, Hirschberg, & Shattuck-Hufnagel, 2004). The ToBI guidelines provide explicit instructions on how to label the different intonational contours and break indices, establishing a baseline for transcribers. In addition, reliability coding by multiple transcribers is typically conducted to ensure the consistency across data. The data in this dissertation were transcribed solely by the author, who is an experienced trained ToBI transcriber. Reliability coding by a second trained transcriber is in progress. Overall, the ToBI system offers three primary advantages for the phonological marking of intonation. The first advantage is that this transcription system allows for a more detailed account of all pitch accents, regardless of utterance position. Second, it is currently the most widely employed system and exists across several languages. This makes its interpretation available to a larger audience.

Finally, the ToBI system has been previously and successfully applied to child speech data (e.g., Prieto et al., 2012), validating its use for analysis in a toddler population.

1.2.2 Information Structure

Information structure refers to how speakers choose to bundle information in a discourse in order to present material to the listener and has been coined both 'information packaging' and 'information structure' in the literature (Chafe, 1976; Clark & Haviland, 1977; Prince, 1981). Vallduví and Engdahl (1996, p. 460) provide the following definition:

> Information packaging (aka communicative dimension, psychological articulation) is a structuring of sentences, by syntactic, prosodic, or morphological means that arises from the need to meet the communicative demands of a particular context or discourse. In particular, information packaging indicates how information conveyed by linguistic means fits into the (hearer's mental model of the) context or discourse.

A more concise definition is found in Büring (2012): "Information structure is simply a cover term for whatever semantic or pragmatic categories such as focus, background, givenness, topic etc. one thinks may influence prosody in this way" (p. 103). Both of these definitions indicate that it is the discourse context that plays a major role in how a speaker will decide to express linguistic information.

When making decisions on how to present information to a listener, speakers have several strategies available depending on their native language. These strategies include manipulations of syntax (e.g., Catalan and Turkish), morphology (e.g., Navajo and Japanese), and, of importance to the discussion here, prosody (i.e., intonation) (e.g., English and German) (Vallduví & Engdahl, 1996). In English, an intonational

manipulation can change the information structure of the utterance. For example, two utterances may be syntactically the same, but a difference in intonation can generate two distinct meanings. See (1.1) and (1.2) for examples of how the intonation can shift which part of a sentence is in focus. The bracketed portion marked with a subscript 'F' denotes the focused component of the sentence.

(1.1) Question: *What are rusty?*

Answer: *[**The pipes**$_F$] are rusty.*

(1.2) Question: *In what condition are the pipes?*

Answer: *The pipes are [**rusty**$_F$].*

The speaker makes these selections depending on the pragmatic context and the listener's state of knowledge and level of attention. Knowledge states are especially important and include what information is shared by both the speaker and the listener in the discourse. Accordingly, the listener is dependent upon the speaker to properly present information in a discourse appropriate manner (Grice, 1975; Vallduví & Engdahl, 1996).

One element that impacts how information is expressed is the degree to which the speaker and the listener are familiar with the information being discussed. Traditionally, a sentence can be partitioned into two primary components: the part that anchors the sentence to the previous discourse (or the hearer's mental model of the world), and the part of the sentence that contributes or adds information to the discourse (or hearer's mental model of the world). This idea of utterances breaking down

into two components dates back to the 19th century, and this dichotomy has been referred to under an array of different names, each with a slightly different meaning depending on the researcher that first identified and defined the terminology. Examples of how this distinction has been labeled in the information structure literature include subject/predicate, known/unknown, theme/rheme, background/focus, and topic/focus, to name just a few (Gabelentz, 1869, as cited in Gundel, 2003; Mathesius, 1928; Halliday, 1967; Bolinger, 1965; Steedman, 1991; Sgall, 1967; among many others). As noted above, information that has not been previously introduced to the discourse is considered *new*, while information that has been previously mentioned is *given* (or *old*). A second common distinction is *topic-focus*, and essentially refers to a similar dichotomy as the *given-new* one.

For production, there have been many conflicting viewpoints as to how the *given-new* distinction is reflected in the speech signal. Some have claimed that only the *new* information in a linguistic expression carries a pitch accent (Halliday, 1967; Prince, 1981; Chafe, 1987). Other researchers have debated whether pitch accent selection and placement is dependent upon factors other than newness alone. Other factors such as the importance, accessibility, or predictability of the information have been argued to lead the selection for what is accented by a speaker in an utterance (Bolinger, 1972; Ariel, 1990; Arnold, 2008; Aylett & Turk, 2004). The argument that *new* information is accented and *given* information is deaccented has continued to be disputed, and a number of elements are now commonly considered to influence whether something is

accented or prominent in the intonation-discourse domain (Venditti & Hirschberg, 2003; Wagner & Watson, 2010).

Many questions persist regarding what guides speakers to accent and deaccent information during a discourse. For example, what exactly is considered *given* information? What factors influence the decision to accent a particular constituent or expression? And, how much of this decision is speaker- versus listener-centered? Prince (1981), Gundel and Fretheim (1993), and Gundel (2003) proposed multi-dimensional answers to these issues by further dividing up the *given-new* space into different levels and hierarchies (see also, Vallduví, 1992, and Kruijff-Korbayová & Steedman, 2003).

The majority of the questions and debates concerning exactly how the *given-new* continuum should be divided up are not addressed in this dissertation. What is important for this discussion is how intonation interacts with the information structure for children acquiring language. This interaction leads to two important considerations regarding 1) what information structure categories are most salient and important to the language-learning infant, and 2) how the *given-new* distinction is reflected in the speech directed to infants. These issues will be further addressed in the three experiments presented in this dissertation.

For the following research, I reduce the *given-new* continuum to a categorical distinction of the two extremes. While specific researchers and studies divide up information structure in different ways, the field has still not come to a consensus on exactly how intonation encodes information structure categories (and what categories

these include). Past studies differ in their terminology, and may also introduce other expressions such as narrow versus broad focus (i.e., what part and how much of the utterance is considered *new*), contrastive versus non-contrastive stress, and emphatic stress.

For my own working definition, *new* information is information that is has not been previously established in the discourse, and it has not been previously introduced linguistically or visually. In contrast, *given* information *has* been previously established in the discourse. *Given* information is linguistically or contextually old and the infant has either heard or seen it before in the discourse context. In the following section, intonation and information structure are detailed concurrently in order to begin to understand how the mapping between the two is acquired by first language learning infants and children.

1.3 Background: The Interaction of Intonation and Information Structure

1.3.1 Early Sensitivities

Some of the first facets of human speech that we are exposed to are rhythm and pitch. Even in the womb, low-pass filtered speech is perceivable and exhibits these fundamental prosodic properties. Considerable evidence points to the influence that this pre-birth exposure to speech has on an infant's perception of language. For example, it has been

shown that infants of only a few days old have a preference for the speech of their mother's voice over a stranger's voice, their native language over a non-native language, and a familiar story that they were exposed to while in the womb over an unfamiliar one (Mehler, Bertoncini, Barriere, & Jassik-Gershenfeld, 1978; DeCasper & Fifer, 1980; DeCasper & Spence, 1986; Mehler et al., 1988).

It has been further hypothesized that it is the prosodic properties of speech, which are embedded partially in the pitch of the language, that create these early perceptual preferences in young infants. Nazzi, Bertoncini, and Mehler (1998) demonstrate that infants are born with the ability to discriminate languages by their broad language-class rhythmic patterns (i.e., syllable-, stress-, and mora-timed varieties). Also, by using a high-amplitude sucking procedure (which monitors nonnutritive sucking behavior in response to auditory stimulation), Japanese newborns were shown to be able to discriminate phonetically identical words that only varied by their pitch contours: low from high, and high from low (Nazzi et al., 1998). These studies provide clear evidence for an early sensitivity to pitch and rhythm, two fundamental components of prosody and intonation.

Along with learning their native rhythm and pitch patterns, an infant must also learn the complex systems for how semantic and pragmatic meanings map onto intonational contours. Developmental research in this area ranges from a focus on specific aspects of intonation, such as the acquisition of contrastive focus, to a broader overview of the acquisition of all possible intonation types (e.g., Hornby & Hass, 1970;

Wieman, 1976; Baltaxe, 1984; Behrens & Gut, 2005; Patel & Brayton, 2009; Prieto et al., 2012). Even so, there is still much that is unknown regarding prosodic development and how children acquire the mapping between intonation and meaning. The area of study that is relevant to this dissertation is how intonation is used to relate components of the information structure, and how children begin to encode this information as well as decode the relevant and related categories.

Initial accounts for how intonation correlates with information status were based partially on a biological approach. Bolinger (1982) identified what he called tensions and relaxations of the body and claimed that these controlled intonation. Later, Gussenhoven (2002) articulated a biological reasoning more succinctly by introducing three core 'biological codes': the Frequency Code, the Production Code, and the Effort Code. The Production Code, which is defined as the generation of energy that is associated with the exhalation phase of the breathing process (i.e., Lieberman's breath group theory (1967)), correlates the most with informational status. Traditionally, a high pitch marks the beginnings of utterances or phrases and a low pitch marks the ends. Meanwhile, the Effort Code measures the amount of energy exerted for a particular part of an utterance and has been hypothesized to be the mechanism behind the expression of *focus*. These 'biological codes' imply that infants follow the breath group theory and produce physiologically less complex expressions before ones that require more articulatory control. This theory has been challenged in recent literature, which shows that infants use a variety of pitch contours, and that the distribution of early utterance patterns may

be linked more to the frequency of the contours and accents in their language rather than to a physiological tendency (Prieto et al., 2012).

Accordingly, we cannot solely rely on a purely biological account for how or why words are accented to relate details of the information structure. Instead, we must continue to examine how the language-learning infant acquires pragmatic and discourse structures. By looking at the different ways that an infant perceives, produces, and comprehends the information structure from the adult speech input, we can begin to understand how infants encode these unique structures within the intonational system. Discovering how infants create the mapping between intonation and information structure is crucial to developing an account that can explain how the infant learns the skills necessary to communicate in a pragmatically motivated discourse.

1.3.2 Encoding and Decoding the Information Structure

Intonation is just one of the ways in which the information structure can be encoded in language. As defined earlier, information structure refers to the ways in which speakers package discourse elements in order to express what is pragmatically relevant or most prominent. The most general aspect of this packaging is how speakers choose to express what is *given* and/or *new* to a discourse context. Along with intonational marking, English-speaking adults have a number of different strategies available to communicate informational status. Some of these options include grammatical manipulations as well as various sentence constructions (e.g., clefts). These informational cues reflect what the speaker and the listener have and do not have in their shared knowledge space.

Consequently, the speaker and the listener use these features to focus on important portions of an utterance and build up the relevant discourse knowledge.

To elucidate the number of levels that may exist in this continuum and the complexity that it would pose for the intonation system, an example is first briefly discussed. Gundel, Hedberg, and Zacharski (1993) propose a six-level hierarchy designed to reflect six distinct cognitive statuses that a speaker/listener may access when discussing a referent. Gundel and colleagues break down the information structure relationship beyond *given* and *new*, and propose the Givenness Hierarchy, consisting of a continuum where full indefinite NPs are used for a 'type identifiable' referent on one extreme, and where pronouns for 'in focus' referents are used on the other (1.3). The Givenness Hierarchy in full, as shown in (1.3), demonstrates each of the levels of cognitive statuses along with which determiners are associated to each level (for a full discussion, see Gundel, 2003).

(1.3) *The Givenness Hierarchy* (Gundel, Hedberg, & Zacharski, 1993)

in focus	>	activated	>	familiar	>	uniquely identifiable	>	referential	>	type identifiable
it		*this/that/this* N		*that* N		*the* N		indefinite *this* N		*a* N

Thus, grammatical substitution or simplification is one of the ways in which the informational state of a referent can be signaled to the listener. In the case of Gundel et al.'s (1993) Givenness Hierarchy, no explicit reference was made to what the intonation might look like for each of these levels. In addition, Gundel's hierarchy is based on the 'cognitive status' or givenness of a referent in discourse (e.g., 'in focus' or 'identifiable'),

but other theories explore the status of a referent in different manners, such as via 'accessibility' (see Ariel (1990) for a discussion of an approach based on the 'accessibility' of a noun phrase/referent). If applying Gundel's theory to acquisition, it is unknown how sensitive children are too the different levels of givenness. Even with a simplification of the continuum, are children able to distinguish *given* from *new* information in discourse? This dissertation explores how the *given-new* distinction affects reference resolution and word learning, as well as how it is produced in early child speech.

With respect to argument realization[1], children behave in accordance with a given discourse-pragmatics. Many researchers have shown that children are much more likely to overtly realize arguments when they are informative or *new* than when they are *given* (for a more complete discussion of argument realization by children, see Allen, Skarabela, & Hughes, 2008). Earlier research shows that even at the one-word stage, before children begin producing multi-word utterances, they are most likely to produce the most informative or *newest* word, omitting others that would be considered *given* (Greenfield & Smith, 1976). This is evidence that children are indeed sensitive to information structure and reflect that knowledge in their word selection. Further support for children's processing of discourse-level information is found in the fact that pronoun interpretation by three-year-olds is affected by the contextual prominence of the referent, a process that is similar in adults (Song & Fisher, 2005). Since children can distinguish *given* from *new* information based on argument structure or pronoun realization, this

[1] Argument realization is defined as overtly stating a particular argument of a verb, which is defined by that verb's argument structure. For example, verbs can call for different argument

leads to the question of how the information structure is specifically encoded by the intonational system during language acquisition. The first two studies proposed in this dissertation address this question by merging information structure with intonation in order to see how this interaction influences early attention and word learning.

Infants have been shown to be sensitive to the *given-new* distinction and are able to respond accurately given the discourse context via either language or action (Greenfield, 1979; O'Neill & Happé, 2000). O'Neill and Happé (2000) showed that typically-developing children at 22-months-old were able to attend to what was considered to be the most contextually salient and relevant items in a discourse, and then use this information to communicate effectively. Thus, children typically attend to the most 'prominent' information in a discourse. My research measures the extent to which children are able to intonationally encode the *given-new* distinction in their own speech. Past studies have already begun to look at how and at what point children produce different aspects of the information structure (Hornby & Hass, 1970; Wieman, 1976; Behrens & Gut, 2005; Chen, 2010).

1.3.3 Perception, Production, and Comprehension

The primary aim of this dissertation is to incorporate complementary viewpoints in order to provide a more succinct and compelling argument for how the interaction between intonation and information structure affects early language acquisition. This interaction can be studied from three primary perspectives: 1) perception, 2) production, and 3) comprehension. While many researchers focus their experiments on one of these

areas (e.g., Wieman, 1976; Behrens & Gut, 2005; Hornby & Hass, 1970), others have dedicated their work to creating links between these approaches in order to determine a developmental timeline for each (e.g., production vs. comprehension, Cutler & Swinney, 1987; Chen, 2010). Also, the studies discussed in this section explore various intonational realizations of the information structure including emphatic stress, prominence, contrastive/corrective stress, and the *topic-focus* distinction.

Wieman (1976) looked at stress production by children during the language-acquiring years. The study analyzed the spontaneous speech of five children at the onset of the two-word period between the ages of 1;9 and 2;5 (years;months). Her main findings showed that children chose to stress *new* material from nearly the onset of the two-word period, in particular, regardless of word order or the lexical category of the word. For example, after a child utters '*MAN*,' they will accent the *new* information being introduced in an adjective plus noun two-word combination, placing primary stress on the adjective and not the noun (e.g. '*BLUE man'*).

One drawback to this study was that the data set was limited with only five children studied and a finite set of only a few utterances to analyze that met criteria. A second drawback was that since this study was a purely perceptual analysis, no acoustic or more detailed accounts of the data were conducted. A third issue was that child speech is often difficult to analyze at the two word period for stress placement on *new* versus *given* information. This difficulty arises due to the fact that it is common for a child to accent both words in a two-word utterance as well as to also produce a short

pause between the words, making it difficult to have a large body of evidence for accentuation during this age period (Behrens & Gut, 2005; Chen & Fikkert, 2007). Critically, although lacking a detailed acoustic account, this study showed that these children did in fact stress *new* information.

Behrens & Gut (2005) studied the speech of one monolingual German-acquiring child from age 2;0 to 2;3. During this age range, the child passed through a phase where he would accent both words in certain two-word combinations, specifically when both words were nouns (Noun+Noun) (Behrens & Gut, 2005). Interestingly, the authors found that not all two-word combinations exhibited the same types of accentuation and that different semantic combinations developed at unique rates during this time period. That is to say, this child exhibited more variability in his accentuation and duration between words at the beginning of this window of analysis. As is the case with many studies using spontaneous speech, the authors decided to conduct a succinct analysis on a small set of data, looking at only certain two word combinations pre-determined by their semantic relationship. Important to note is that the data were from only one child. Together, Wieman (1976) and Behrens and Gut (2005) demonstrate that while intonation patterns vary during early speech development, discourse newness is one of the primary elements attracting accentuation.

Other research has looked at how contrastive information is produced by young children. Hornby and Hass (1970) showed that four-year-olds stressed the contrastive element in their productions. Their study asked twenty children to describe pairs of

pictures where the second picture contrasted with the first in one of three ways: by action, agent, or object. For example, the first picture might depict 'a girl petting a *cat'* and the second 'a girl petting a *dog*,' where the object is in contrast in this case. Children were reported as using contrastive stress on the object when describing the second picture. Hornby and Hass (1970) took a step forward with this study by realizing that a controlled experiment was needed to look more fully at the relationship between what they called *topic-comment* (or *given-new*). One critique of Hornby and Hass (1970) is that they did not clarify the criteria that they used for labeling contrastive stress. The scoring was performed perceptually, and no acoustic analysis or formal intonational transcription was utilized. This implies that while their central findings are correct, there still remains an opportunity to study the nuances and complexities of intonation within the relationship between *given* and *new*.

Hornby (1971) extended Hornby and Hass (1970) and analyzed the speech of six-, eight-, and ten-year-old children. This study asked children to verbally correct a picture that they had been presented. It found that children preferred to use a contrastive accent for the majority of responses, especially to other methods for marking contrast (e.g., clefts or passives). Like previous work, analyses were perceptual in nature. Crucially, like Wieman (1976) and Behrens and Gut (2005), Hornby and colleagues showed that children are able to use accentuation successfully to mark the informational status of a word in their early speech. These two studies analyzed the speech from

children aged four and older. Studying younger populations in future research is necessary in order to learn when children begin to stress *new* information in general.

MacWhinney and Bates (1978) conducted a series of production studies and looked cross-linguistically at how information status is produced and analyzed in English, Hungarian, and Italian. MacWhinney and Bates (1978) examined the speech of three-, four-, five-, and six-years-olds, as well as an adult control group. Their procedure consisted of subjects describing a series of pictures where certain items varied in their newness relative to the discourse setting. They found a significant increase in the use of what they called emphatic stress, which varied as a function of age and language. Overall, English speakers used emphatic stress most often to signal newness. In contrast, Hungarian and Italian speakers after age three- and four-years-old resorted to methods besides stress to express *new* information (e.g., word order). For the English children, the use of emphatic stress on *new* material increased with age, with six-year-olds behaving most like the adult controls.

One prediction of MacWhinney and Bates (1978) was that emphatic stress would be used more reliably and consistently for information that was considered 'newer', with newness varying along an eight level continuum (similar to Gundel et al.'s (1993) Givenness Hierarchy). The authors identified two criteria for marking whether emphatic stress was present or not on a particular element in the discourse. These two criteria for emphatic stress were: 1) intonationally the element was more stressed than other items in the utterance; and 2) the amount of stress on that item was considered more than it

would have been in a neutral context. As with earlier studies by Hornby and Hass (1970), Wieman (1976), and Behrens and Gut (2005), these criteria were met via a perceptual threshold. The relative increase of emphasis in newness was measured using scorer-dependent criteria and was not verified using acoustic or phonetic evidence. Thus, children showed differences in the accentuation of *new* material, but the specific types of intonation contours that they were using remain unknown with this type of analysis.

Another study that demonstrated the viability of pursuing research on contrastive stress was Baltaxe (1984). This study conducted an experiment to look at the production of contrastive stress in typically- and atypically-developing (due to aphasia or autism) populations. Of interest to this discussion are the results from the typically developing population, who ranged in age from 29 to 48 months (2;5-3;0). In contrast to MacWhinney and Bates (1978), Baltaxe (1984) examined a younger population and demonstrated a similar pattern in the consistency of contrastive stress usage. Children showed more stable productions when the subject was in contrast, versus the verb or the object. These results are also similar to Hornby and Hass (1970), and indicate that it is the subject that occupies the most stable position for reliable accentuation. Even though this study did not label the specific intonational contour, stress was perceived as prominent (assessed perceptually as with previous experiments). Still, a significant issue with perceptual studies such as Baltaxe (1984) and MacWhinney and Bates (1978) is that it is difficult to know exactly which features of the speech signal drove perceptual

prominence, suggesting that further research along these lines hold promise as nuanced extensions of their previous work.

A substantial amount of recent research has cast doubt on what exactly perceptually distinguishes prominence. One critique of this method is that a listener may be indirectly biased in a certain direction when making perceptual judgments. The Multiple Source Model (Watson, 2010) identified the features that are critical in determining prominence, and demonstrates that there are other acoustic cues to prominence besides intonational accent alone that influence perception. This model highlights the fact that it is difficult to know what specific features in the speech signal are influencing perceptual judgments. A second important critique is that most studies analyze narrow focus constructions (i.e., contrastive or emphatic stress). Many of the experiments presented did not inspect *new* information when it was produced in broad focus (i.e., where the entire utterance is in focus, versus narrow focus, where only part of the utterance is in focus), nor did they look at specific intonational patterns or the effect of deaccentuation, all of which play a vital role in information packaging in West Germanic languages.

Herbst (2007) took advantage of the fact that these aspects had yet to be analyzed and looked at the speech of German-acquiring children with a mean age of 5;8. Her study analyzed the production of referents at one of three levels: *new*, *given,* and *accessible*. *New* referents were completely new to the discourse and occurred for the first time. *Given* versus *accessible* referents were determined by their distance between the

first and second mention, and were tested in the *immediate* and *distant* conditions, respectively. *Given* referents were considered in the *immediate* condition if after the referent was introduced in one picture, it would immediately be presented in the following picture as the target referent. In the *distant* condition, *accessible* referents were introduced at the beginning of a story and then re-introduced as the target referent after four or five intervening pictures had been displayed. Unlike previous studies, productions were acoustically analyzed using Praat (Boersma & Weenink, 2014) and labeled according to the GToBI guidelines for German.

Results from Herbst (2007) showed that for *new* versus *immediate* target productions, the target in the *new* condition carried a pitch accent in over 90% of cases. In comparison, the target referents in the *immediate* condition were deaccented over 60% of the time. For both the *new* productions and for when the *immediate* (or *given*) targets were produced with an accent, the most favored pitch accent was the H* in the GToBI transcription system.

These results provide clear evidence that five-year-olds intonationally differentiate *given* from *new* during production. Furthermore, the results from *given* versus *accessible* referents also show that children at this age can discriminate these two different levels during production. The *distant* (or *accessible*) referents, like in the *new* condition, were accented over 90% of the time. In comparison, the *immediate* condition targets only received an accent about 40% of the time. This raises the question of whether children have an intonational category for the *accessible* (or *distant/semi-active*)

referential level. Herbst (2007) supports the hypothesis that even though the production of intonational contrasts may still be developing through the pre-school years, children at this age can minimally and intonationally differentiate *given* from *new* information.

Grassmann and Tomasello (2010) investigated how prosodic stress interacts with contextually *new* referents to guide the attention of German-acquiring two-year-olds. The experiment analyzed toddler looking time responses in one of three experimental conditions designed to pull apart and look at the effect of *newness* and *stress*. The three conditions were: *newness only*, *stress only*, and *newness and stress* combined. For the *newness only* condition the infant was presented first with an image of a duck while a live experimenter told the child '*Look a duck. Look the duck*'. Next, a blank screen appeared and the experimenter referenced both the contextually old material (*duck*) as well as a new referent (*apple*) ('*The duck has the apple*$_{New/Untressed}$'). In this case, no emphasis or stress was placed on the discourse new referent (*apple*).

The *stress only* condition differed from the *newness only* condition in that during the initial frame, both referents were displayed while the experimenter identified them both (*'Look the duck and the apple'*). Before the target images reappeared in a new configuration on screen, the experimenter would produce an utterance with stress on only one referent ('*The duck has the APPLE*$_{Given/Stressed}$'). Finally, the *newness and stress* condition would initially introduce only one referent (*duck*). Then, the test screen would display both the old referent (*duck*) and a new one (*apple*) while stressing the new one ('*The duck has the APPLE*$_{New/Stressed}$'). The analysis measured looking time to the target

during the final test display containing the two referents. Results showed that neither *newness* nor *stress* alone was sufficient to guide attention to a referent, but rather a combination of both *newness and stress* was needed. This study was one of the first to show the communicative function of stress in the perceptual processing of young children, an important predecessor to the study presented in Chapters 2.

Importantly, all of the auditory stimuli in Grassmann and Tomasello (2010) were produced by a live experimenter. The rationale in using this method was to create a more realistic and natural setting. The drawback of this technique is that even with training, there is a considerable amount of variability introduced into the stimuli and data (see §2.1 for a continued discussion of Grassmann and Tomasello (2010)). In general, there has been insufficient focus on the perception and comprehension of *given* versus *new* information when introduced in broad focus in discourse. However, there are several studies that have investigated what is referred to as the 'production-comprehension asymmetry'.

The earliest and most cited supporter in favor of the production-comprehension asymmetry is Cutler & Swinney (1987). This study consisted of a series of four reaction time experiments testing how long it took for three- to eight-year-old children to recognize a target word when the preceding context varied the focus or accentuation of the target. In the first two experiments, Cutler and Swinney (1987) tested whether children displayed a reaction time advantage for previously accented words in a word recognition paradigm. Results showed that the youngest group of four- to six-year-olds

did not show an increase in processing time, but that the older children and adults did for stressed words. A decrease in reaction time by the four- and six-year-olds was found only in the cases where stressed words were embedded in well-formed sentences (experiment 1), but it did not show up when the target words were presented in list form as syntactically ill-formed sentences (experiment 2). Cutler and Swinney (1987) suggested that four- to six-year-olds could not both semantically and prosodically interpret sentences of this type like the older children, since stress only aided processing when there was no semantic load (exp. 2).

Experiments three and four followed up on the results from the first two, and asked whether or not a word in focus aids in comprehension (Cutler & Swinney, 1987). In this case, only the youngest children (3;0 to 4;0) did not show an advantage for words that had been in focus. This led the authors to conclude that children under the age of six do not have an underlying discourse representation, as indicated by their delay in processing. Since previous literature pointed to children producing these distinctions in their own speech at an early age (Hornby & Hass, 1970; Wieman, 1976; MacWhinney & Bates, 1978), the authors proposed an innate hypothesis to account for this paradox in production versus comprehension. They cited Bolinger's (1982) hypothesis that prosodic productions are innate and that this is why production skills do not require the same amount of time to develop as comprehension skills in order to become adult-like. Cutler and Swinney (1987) stated that children's early productions are innate responses and only later in development do they become adult-like in the sense that they are true to

the discourse planning models and representations. Several aspects of this study call for further discussion, including how the authors established newness in discourse and how exactly they designed their experiments. I address and clarify both of these points in the experiments of this dissertation.

There are four primary concerns with Cutler & Swinney (1987). First, stress did not always correlate with new material, varying perhaps in a non-naturalistic way and confusing younger language learners who are sensitive to prosody. Second, a preceding question in the discourse was used to set up the contextual framing that created narrow focus (not the intonation itself). Thus, focus was defined only by the syntax, with no particular intonation contour on the focused word, further complicating the task for young learners. Third, the task administered could be considered inappropriate for the age groups tested. As a speeded word recognition task, the issue arises of what exactly is considered comprehension. In terms of *given* versus *new* material, the target words varied in how many times they were repeated during the experiment, making it difficult to draw correlations to children's productive abilities. Fourth, this study did not have its own production correlate, making it still more difficult to use in support of the production-comprehension asymmetry.

A later set of experiments continued to explore the connections between production and comprehension. Wells, Peppé, and Goulandris (2004) conducted a study to look at the production and comprehension of intonation by five- to fourteen-year-olds. Their study administered a series of tests designed to look at the functional use and

understanding of intonation (the task administered was the PEPS-C, or Profiling Elements of Prosodic Systems – Child version; Wells & Peppé, 2001, 2003; Peppé & McCann, 2003). The tasks were constructed to look at four communicative areas that use intonation, including *chunking* (or prosodic phrasing), *affect*, *interaction*, and *focus*. Results from these tasks demonstrate that most functional intonational production is developed by age five with some of the more complex prosodic contrasts not fully developed until around age eight. In comparison with the production component, the authors showed more variability with the comprehension part of the study. The comprehension abilities develop more gradually over the time span analyzed, with some aspects not fully developed until almost age eleven. However, this battery of tests does not directly test the *given-new* distinction. The one task in the set that most directly relates to this dichotomy is the portion on *focus*.

The *focus* task in the PEPS-C had children both produce and interpret corrective narrow focus (similar to *contrastive focus* as elicited in my Experiment 3) (Wells et al., 2004). The experimenter asked a question such as '*How about a green bike?*' and the child had to respond in order to acquire the picture that they actually needed (e.g., 'I want the *WHITE* bike'). For the comprehension of *focus*, a recorded version of either 'I want *CHOC*OLATE and honey' or 'I want CHOCOLATE AND *HON*EY' was played for the child. The child then had to decide what food the speaker wanted in return. This task clearly analyzed the production and comprehension of this particular type of corrective narrow focus, but it did not look at how children might produce and understand other

intonational categories that reflect information structure. Wells et al. (2004) showed an asymmetry between prosodic production and comprehension, with comprehension abilities continuing to develop up until fourteen years of age. Despite this fact, this study only looked at a limited aspect of prosodic development. Recent work has called into question the existence of this asymmetry. Since this study was testing a speech therapy task, it looked at a larger set of focus constructions. My research aims to both simplify the experimental paradigm as well as to collect a more natural speech sample in order to examine toddlers' production of *contrastive focus* and *newness*.

The main purpose of Arnold (2008), by contrast, was to look at the comprehension and interpretation of accented and unaccented noun phrases by adults and four- and five-year-old children using an eye-tracking paradigm. The primary goal was to observe how children comprehend referential expressions online in comparison to adults. The experiment presented subjects with a screen displaying four items, two of which were linguistic cohort competitors that overlapped in their initial segments (e.g., ba<u>gel</u>/ba<u>con</u>). During the test phase, the subject heard an utterance such as '*Put the <u>bagel</u> above the star. Now put the <u>bagel (it</u>) on the square.'* Since accented information is associated with *new* information and *given* with unaccented, this experiment manipulated the accentuation of the second production of the target word.

Arnold (2008) showed that both adults and four- and five-year-olds exhibited a bias to the previously mentioned object when the target in the second utterance was an unaccented noun phrase or pronoun (about 300ms from the onset of the unaccented

word). Accented words in the second utterance did not show any observable bias. This study demonstrates the effect of deaccented *given* information in discourse and how it aids processing. The near adult-like performance by this age group of children in their pronoun interpretation, aided by the deaccentuation of *given* referents, lends support for early comprehension for this particular aspect of the *given-new* contrast by exploiting the prosodic structure of discourse. Arnold (2008) shows that an investigation of prosodic cues in early child language acquisition is viable for children as young as four years old. Whereas the work by Wells and colleagues suggest that the comprehension of prosodic structure continues to develop until fourteen years, Arnold's (2008) findings indicate that with the use of a more age-appropriate paradigm, reference resolution effects can be seen as in a younger populations. Experiment 1 of this dissertation explores the effect of pitch type and *newness* on reference resolution with even younger toddlers at 18-months of age.

The evidence in favor of a production-comprehension asymmetry in development found by Cutler and Swinney (1987) and Wells et al. (2008) does not include a range of informational statuses and intonational representations. These studies focused on specific and more complex notions of information structure. When Arnold (2008) employed a more simplified and appropriate paradigm to look at the comprehension of *given* versus *new* information, such an asymmetry was not found. If testing were to be completed with younger children using a more an age-appropriate way to demonstrate the *given-*

new distinction, the question remains of whether or not these children will still have difficulty in comprehension.

Recent work by Grassmann and Tomasello (2010) and Chen (2010) also contradict research showing that comprehension develops after production. Chen (2010) argued that these earlier studies found this asymmetry due to a possible effect of experimental design and a variation in methodological testing. In her study, Chen (2010) used the word *focus* to refer to words with accents that express *new* information, both in narrow and broad focus conditions. Her study looked at the production and comprehension of the focus-to-accentuation marking for *narrow non-contrastive focus* in Dutch-speaking adults and four- to five-year-old children. For her production experiment, children produced utterances where the portion in focus was an argument of the verb (the subject or the object) as well as a full noun phrase.

The experimental paradigm in Chen (2010) was designed as a game that the children played with an experimenter and a robot. The child watched the experimenter present and describe one picture at a time, and then the experimenter asked a question to the robot about something in the picture (e.g., '*What is the cleaning lady picking up?*'). At this point, the robot would answer the experimenter's question but with an abnormal prosody that was robot-like in nature, lacking any distinct prosody on the words or overall utterance. The child then had to reproduce the robot's utterance using his or her own intonation in order for the experimenter to find the correct picture in a matching game. In example (1.4) below, the utterance elicited contained *non-contrastive*

focus on the subject. (*Focus* and *topic* are analogous to *new* and *given* in the current discussion).

(1.4) Experimenter: *Kijk! Een biet. Wie eet de biet?*
"Look! A beet. Who is eating the beet?"

Child Subject: *[Een poetsvrouw]$_{focus}$ eet [de biet]$_{topic}$.*
"A cleaning-lady is eating the beat.

Results from the first part of Chen (2010) suggested that four- and five-year-olds can produce *non-contrastive narrow focus*, but they do not yet exhibit complete adult-like control. The child pattern matched the adult frequency pattern for how often they accent a focal noun (93% for children and 98% for adults). For non-focal nouns, children tended to use accentuation more often than adults, particularly when the words were in sentence-final position. This was hypothesized to be due to a tendency for children to seek confirmation from adults by placing rising stress on the final word of a sentence.

The second part of Chen (2010) looked at four- and five-year-olds ability to comprehend the focus-to-accentuation mapping. Just as felicitous intonation has been shown to make comprehension easier and faster, infelicitous intonation has been shown to do the opposite for comprehension. With evidence from reaction time experiments demonstrating this processing effect in adults (Birch & Clifton, 1995), Chen (2010) adapted their technique to test comprehension of newly accented material in children. The experiment consisted of a game where the child participant would listen to dialogues between a boy and his pets. The boy in the game would ask questions to the pets about a picture and the participant had to press one of two buttons depending on whether the

answer the pet gave was correct or incorrect. *Focus* was determined by the preceding question type in the game dialogue (either a WHO- or WHAT-interrogative). The experimental stimuli were manipulated in two ways, by accent placement (e.g., pragmatically appropriate or inappropriate) and by the focus location (e.g., subject or object) in the answer sentence. Incorrect and correct judgments as well as reaction times for these responses were collected for each subject.

Unlike previous work by Cutler and Swinney (1987) that show an asymmetry in production and comprehension, the results from this study showed that children (although slower overall than adults) and adults both exhibit faster reaction times when the utterance was pragmatically appropriate versus when it was inappropriate. This indicated that children were correctly making the focus-to-accentuation mappings in the case of *non-contrastive narrow focus*. These results are notable as they provide evidence contrary to the production-comprehension asymmetry. By using a simplified methodology that was appropriate for the age range tested, Chen (2010) showed that children and adults have similar patterns for the production and comprehension, even if children are not yet fully adult-like. Experiment 3 of this dissertation concerns early speech production and analyzes the subtle phonetic differences in how young children and adults make information status distinctions. Importantly, her findings also indicated that processing is slowed for pitch accents that are contextually inappropriate. I further explore how pitch accent felicitousness affects novel word learning in Experiment 2.

Chen (2011) analyzed specific pitch accent realizations of *topic* and *focus* in Dutch-speaking children. The main finding of this study was that children as young as four and five are adult-like in the marking of accent for *topic* and *focus*, but not yet in their accent type in sentence-final position. The marking of the H*L pitch accent, common for sentence-final focus in Dutch, continues to develop from five to seven years old. Like in her previous work, the tendency for younger children to place rising pitch accents in sentence-final position is hypothesized to be due to a confirmation-seeking interpretation. Thus, while Dutch children can produce the *topic-focus* distinction as young as age four, pragmatic marking may still affect accent selection until later in development. This study illustrates how subtle pragmatic effects can affect pitch accent selection. Thus, it is important to consider context in production, as well as to label these differences during analysis. For Experiment 3 of this dissertation, an effort is made to analyze only pragmatically similar utterances together, but it is evident that this factor is difficult to entirely control (see Chapter 4).

Many of the experiments discussed have studied comprehension in children age three-years and older. Even cases where an asymmetry was disputed, children have already reached pre-school age (4;0 or 5;0). Recall that Grassmann and Tomasello (2010) looked at the processing abilities of two-year-old children and showed that they were able to use *newness* and *accentuation* together in order to aid in word recognition and reference resolution. Typical comprehension experiments may not always be ideal with younger populations due to task complexity. In these cases, perceptual processing can

reveal substantial information in how young children are using intonation and information structure in areas such as reference resolution and word learning.

This dissertation addresses the primary critiques and concerns brought up throughout this literature review including issues of perceptual and subjective coding techniques, implementation of age-appropriate methodologies, and the testing broad focus aspects of the *given-new* contrast. I address concerns regarding the coding analysis by using both a phonological and phonetic approach, which provides a more complete picture of the development of intonation with information structure. Also, I correct for methodological issues through the use of more sensitive testing techniques (e.g., automated eye-tracking) and simplified procedures. Finally, I expand on the *given-new* distinction to test how its broad focus realizations are processed and produced by young toddlers.

My research begins earlier in development than past experiments and examines both perceptual and productive abilities in young toddlers age 1;6 to 2;6. Perceptually, it addresses how attention is mediated by the interaction of intonation and information structure. Additionally, it further probes production abilities in order to analyze how toddlers intonationally express *given, new,* and *contrastive* information. Finally, the experiments presented in this dissertation offer not only a phonological viewpoint and analysis, but an acoustic one as well.

1.4 Acoustic Underpinnings

An important approach to understanding how infants and children can encode and decode the information structure as it is packaged by the intonational system is by identifying what the most important and reliable acoustic parameters are in cueing these differences. Substantial research has been conducted looking at how *given* and *new* information are acoustically realized. More recently, this area of study has been expanded to include what acoustic correlates are typically associated with particular functions of the information structure (e.g., narrow vs. broad focus, contrastiveness). The present set of experiments analyzes the acoustic correlates of intonation and information structure in order to explore the perceptual and productive abilities of toddlers.

Discourse pragmatics and informational status have been shown to influence the ways in which adult speakers choose to produce the mention of a referent (Jackendoff, 1972; Steedman, 1991; Pierrehumbert & Hirschberg, 1990; Vallduví & Engdahl, 1996; Shattuck-Hufnagel & Turk, 1996). A first mention is typically stressed or accented by the speaker (Prince, 1981; Chafe, 1987). Within the AM approach, *new* information is produced using an H* accent (where there is a high peak on the stressed portion of the target word), and *given* information is often deaccented with no pitch accent present on the word (in Western Germanic languages) (Ladd, 1996; Féry & Kügler, 2008). In addition, *contrastive* information is phonetically distinct from *discourse-new* elements, with *contrastive* information showing greater duration, relative intensity, and F_0

movement (Katz and Selkirk, 2011). Focused elements have also been shown to exhibit higher relative F_0 than non-focused ones (Müller, Höhle, Schmitz, & Weissenborn, 2006).

In general, acoustic analyses of multiple word repetitions have shown that *given* information is acoustically reduced by the speaker (Fowler & Housum, 1987). This reduction in stress over the course of repetitions is often referred to as the 'given/new contract', indicating that the choices of the speaker influence the way in which the listener receives and interprets this information (Halliday, 1967; Chafe, 1987; Prince 1981). This has been the case for adult-directed speech, but some claim that it also persists in speech directed to infants, an issue that has recently been called into question (Fisher & Tokura, 1995; Bortfeld & Morgan, 2010; see §4.4 for further discussion on the impact of child-directed speech).

Adults are precise in their productions of pitch accents, aligning F_0 peaks and valleys consistently even as speech rates change (Arvaniti, Ladd, & Mennen, 1998; Xu, 1998; Ladd, Faulkner, Faulkner, & Schepman, 1999; Ladd, Mennen, & Schepman, 2000; Ladd & Schepman, 2003; Dilley, Ladd, & Schepman, 2005). In perception, on the other hand, this type of precision is not as important. In fact, as adults, we have the ability to adapt to new speakers and situations as well as perceive pitch accents and meanings of varying productions (Watson, Gunlogson, & Tanenhaus, 2008; Kurumada, Brown, & Tanenhaus, 2012). Still, there has been little research to date on how children use fine-grained intonational information and how it is linked to patterns in language

development. This suggests the potential for future studies that look particularly at the acoustic correlates of intonation in the speech of children and adults.

Pitch and rhythm are first exposed to the infant when he/she is still in the womb, so it is not surprising to learn that children are nearly adult-like in their selection and production of nuclear intonational contours from nearly the onset of speech (Prieto et al., 2012). Detailed work in a variety of languages has shown sophisticated F_0 patterns in children as young as 12 months old (Prieto & Vanrell, 2007, for Catalan; Astruc, Payne, Post, Vanrell, & Prieto, 2013, for Catalan and Spanish; Frota & Vigário, 2008, for European Portuguese; D'Odorico & Carubbi, 2003; D'Odorico & Fasolo, 2007, for Italian). Children do not necessarily acquire all of the intonational features seamlessly, but these studies showing early F_0 control propose that the acquisition of the productive abilities in this area arrive relatively early in development.

A recent study addresses these issues in adult speakers and directly outlines the acoustic underpinnings of specific information structure categories in English. Breen, Fedorenko, Wagner, and Gibson (2010) investigated three questions concerning the study of acoustics and information structure: 1) Do speakers mark information structure using prosody? 2) If so, then what are the correlated acoustic features? and 3) How well do listeners retrieve this prosodic information from the acoustic signal? The authors conducted three studies to test these respective questions and found that speakers mark information structure using particular features of the acoustic signal. To answer their first question, the experiments showed that participants reliably provide acoustic cues to

distinguish focus location (subject, object, or verb), focus type (contrastive vs. non-contrastive), and focus breadth (narrow vs. wide). In response to their second question, the four best acoustic correlates to these prosodic distinctions were word duration (plus silence after the word), mean F_0, maximum F_0, and maximum word intensity.

Specifically, focus location was correlated with greater intensity, longer duration, and a higher mean and maximum F_0. In the case of focus type, contrastive stress exhibited a greater intensity, longer duration, and a lower mean and maximum F_0 than non-contrastive stress. Finally, in relation to focus breadth, objects in narrow focus were distinguished by a greater intensity, longer duration, and higher mean and maximum F_0, while wide focus exhibited higher intensity and F_0 as well as longer durations on pre-focal words. To answer their final question, Breen et al. (2010) ran two perceptual experiments testing the saliency and discriminability of these acoustic features with these associated meanings. Listeners proved to be able to successfully distinguish utterances along these dimensions.

Breen et al. (2010) discussed their results in terms of the *direct-relationship* approach, but note that the results are also compatible with the *indirect-relationship* approach. The *direct-relationship* describes an approach where the acoustics are directly associated with particular meanings (Xu & Xu, 2005; among others). The *indirect-relationship* approach, or the intonational phonology framework, is where the mapping between acoustics and meaning is mediated by means of phonological categories (Gussenhoven, 1983; Ladd, 1996; Pierrehumbert, 1980). Since intonational phonology

posits pitch accent categories that are phonetically mediated by the acoustics, Breen's (2010) study demonstrates how acoustics relate to informationally structured categories. These results support an analysis like the one used in this dissertation, where intonation plays an integral role in the information structure and demonstrates that there exist clear acoustic correlates to these categorical distinctions.

Shattuck-Hufnagel and Turk (1996) identify the primary acoustic correlates of prosody as duration, fundamental frequency, amplitude (or intensity), and segmental quality and/or reduction. Since acoustics vary considerably due to segmental qualities, they propose an integration of both phonology and acoustics in the study of prosody. They define prosody "as both (1) acoustic patterns of F_0, duration, amplitude, spectral tilt, and segmental reductions, and their articulatory correlates, and (2) the higher-level structures that best account for these patterns" (Shattuck-Hufnagel & Turk, 1996, p. 196). This definition supports the rationale for including both phonological and acoustic analyses to study prosody. By studying both levels, we can provide a more complete picture for how intonation relates to the information structure, and how it develops during infancy and early childhood.

1.5 Summary

The literature to date has varied considerably in how information structure is implemented by the intonation system. A variety of aspects have been analyzed, with early work by Hornby and Hass (1970) and Wieman (1976) looking at the production of

corrective focus, and more recent work analyzing the *topic-focus* distinction in an effort to address the production-comprehension asymmetry (Chen, 2010, 2011; Cutler & Swinney, 1987). Until recently, almost no work has examined how information structure interacts with intonation during early perceptual processing (Grassmann & Tomasello, 2010). There are a number of concerns with these studies though, leaving open many areas for future research.

The biggest critique of past work is that it has been largely perceptual in nature, and has failed to take into consideration a full phonological and phonetic account of the data. It also has differed in terminology and in what type of information category has been analyzed (e.g., *topic* vs. *focus*, *contrastiveness*, *emphasis*, *prominence, narrow focus*), with a gap in broad focus analyses. In addition, most of this work has focused on production and comprehension in children, with a tendency to study older age groups. The research presented in this dissertation emerges from these critiques and aims to provide a more structured and detailed account of the development of information structure and intonation.

As previously mentioned, I include a phonetic component along with a phonological one for the intonational analysis in this dissertation. In the perceptual studies, I am careful to identify the acoustic correlates of the pitch accents. For example, in Experiment 1 I fully isolate F_0 to study its specific role in reference resolution. In the production study in Experiment 3, I analyze the pitch accents following an AM approach as well as analyze the corresponding acoustic correlates for both adult and child speech. I

am also very careful to use consistent and explicit definitions of *given*, *new*, and *contrastive* across my set of experiments (as established in §1.2.3), with attention on both *broad* and *narrow focus* uses (Experiment 1 and 2) and investigations (Experiment 3). Consistency in terminology strengthens the connections between the three experiments. Finally, I study the interaction of intonation and information structure in younger populations (1;6 to 2;6) by using methodologies and novel procedures that are more appropriate for this age period.

In line with work by Chen (2010) and Arnold (2008), I developed studies that are suitable for the ages analyzed. I successfully apply the AM approach to analyze perception and production, which provides corollary support of the application of this model in other languages such as in Catalan, Castilian Spanish, and European Portuguese (Prieto & Vanrell, 2007; Prieto et al., 2012; Frota, & Vigário, 2008). Also, I follow recent prosodic research and include an important acoustic component (Shattuck-Hufnagel, 1996; Breen et al., 2010). My work supports the conclusions that children can exploit information structure categories and specific pitch accents (e.g., H* and L+H*) during reference resolution and word learning, that they are adult-like in their production with respect to nuanced intonation, and that experiments that combine acoustics, intonational phonology, and age-appropriate contextual situations are necessary to accurately map out these capabilities of children with respect to language acquisition.

Chapter 2

Experiment 1:

Attention

2.1. Introduction

This dissertation focuses on both perception and production in early language acquisition. In this chapter, I begin by investigating early speech perception abilities at 18-months-old and how attention is mediated by intonation and information structure. Establishing how these factors interact to affect early attentional processing lays the groundwork for how this impacts subsequent early word learning, which is discussed in Chapter 2.

Infants as young as five days old are sensitive to native language rhythm and pitch patterns. They can discriminate languages by their broad rhythm-class patterns (Nazzi, Bertoncini, & Mehler, 1998). That is, a newborn infant will be able to distinguish a syllable-timed language such as French from both a stress-timed language like English or a mora-timed language like Japanese. Additionally, young infants can also differentiate low-high from high-low pitch contours, and strong-weak from weak-strong disyllabic words (Nazzi, Floccia, & Bertoncini, 1998; Jusczyk & Thompson, 1978). Then, from the onset of speech production at around one year of age, babies are able to approximate adult-like intonation contours (Prieto & Vanrell, 2007; Chen & Fikkert, 2007; Frota & Vigário, 2008) and align these contours with felicitous semantic meanings and pragmatic intentions (Prieto, Estrella, Thorson, & Vanrell, 2012). However, little research has been conducted on the early *comprehension* of contours as they reflect information status. Previous research suggests that toddler attention to referents is mediated by both intonation and information structure in discourse (Grassmann &

Tomasello, 2010). In turn, attention to referents is essential for making the correct word-to-object or word-to-action mappings necessary for early word learning (Tomasello & Akhtar, 1995; Grassmann & Tomasello, 2007). The motivation for the current study is to investigate how American English-acquiring 18-month-olds are guided by mappings from intonation to information structure during on-line reference resolution in discourse.

Previous research by Grassmann and Tomasello (2010) claimed that German-acquiring two-year-olds attend to a referent of a familiar word if and only if the word is both *stressed* and *new* in the discourse context. Their experiment consisted of three conditions where the target referent could be introduced in a brief discourse context with either (a) *stress* only, (b) *newness* only, or (c) both *stress* and *newness*. Grassmann and Tomasello (2010) used a live speaker to present stimuli to each subject in order to ensure a level of intentionality on the part of the speaker and measured looking time and pointing to a referent as their dependent variables. Children in their experiment looked equally as long at the distractor referent as at the target when it was *new* only (and not stressed) or *stressed* only (and not new). Critically, they found reliable looking to the target referent in the third condition only, where it was both *new* and *stressed*.

This experiment expands upon the work by Grassmann and Tomasello (2010), introducing a number of methodological changes. First, I ask whether specific pitch movements more systematically predict patterns of attention to referents, rather than using one collapsed "stressed" category (similar to the H* pitch accent in the ToBI transcription system). Second, I add a flat fundamental frequency condition to the

design to act as a control against which the pitch movement (or 'stressed') conditions may be compared. Third, because a live speaker may have produced varying intonation contours, I control for speaker variations and possible experimenter bias by using pre-recorded stimuli. This in turn allows the ability to isolate the role of pitch in guiding attention while holding duration and average intensity constant.

Instead of one 'stressed category', this study explores how unique pitch movements affect toddler attention. I focused on two pitch accent movements in American English, the simple monotonal H* and the complex bitonal L+H*. Previous work in adult speech perception shows that the simple H* pitch accent can be associated with either new or contrastive information in discourse, while the complex L+H* accent is more typically associated with only a contrastive interpretation (Watson et al., 2008). Although I only manipulate pitch (F_0) in this experiment, I exploit these mappings in American English in order to test how these specific pitch accent movements interact with referential newness and givenness during early attentional processes. For this study, the three pitch type movements selected were a simple monotonal (˜H*), a complex bitonal (˜L+H*) and a flat F_0 control.

Information structure in this experiment is used to describe the simple dichotomy between *new* and *given* information. For each of the short discourse contexts presented, a referent is considered *new* if it has not been previously mentioned or seen by the participant. A referent is *given* if it has been previously mentioned and seen during the discourse context prior to the test phase.

Finally, past research has emphasized the role of social pragmatics and intentionality in achieving successful word learning (Tomasello & Akhtar, 1995; Akhtar, Carpenter, & Tomasello, 1996). Importantly, this study tests outcomes when live interactions are removed from the experimental design and any degree of intentionality is only accessible through the utterances themselves and their corresponding prosody.

The primary research question for this study concerns how the mapping between information structure and intonation guide toddler attention in a controlled discourse context. It expands upon past research in order to test how sensitive 18-month-old toddlers are to specific pitch accent movements in American English, depending on the discourse context. The predictions for this experiment are three-fold. First, I predict that newness will be sufficient to guide attention to a referent, regardless of pitch accent type. Even when a target is deaccented when introduced, the novelty of the referent may still attract the attention of the toddler. Studies of visual attention in the absence of discourse have consistently found that infants and toddlers prefer to look at novel patterns or objects, and I would expect this bias to influence looking during discourse as well (Fantz, 1964; Rose, Gottfried, Melloy-Carminar, & Bridger, 1982).

Second, I predict that a semantically and pragmatically appropriate pitch accent movement will guide attention to both *new* and *given* referents. As previously mentioned, the pitch accent movements that will be used in this study have been shown to guide attention to presentationally new (H*) and contrastive information (both H* and L+H*) (Watson et al., 2008). Consequently, these two pitch accent movement types

should guide attention to a referent that is contrastive, even if it is *given* in the discourse.

Finally, the third prediction is that the complex pitch movement will show increased looking to the target referent over the simple pitch movement, specifically in the case where the referent is *given* in the discourse context. Such an increase in looking is expected due to two potential influences. First, recent studies have shown that a bitonal rising pitch movement is more prominent than a monotonal pitch movement (Baumann & Roth, 2014; Turnbull, Royer, Ito, & Speer, 2014). Second, the bitonal pitch movement is a common feature of child-directed speech, a preferred register type versus adult-directed speech for children at this age (Fernald & Simon, 1984; Fernald, 1985; Yu, Khan, & Sundara, 2014). The predictions here depart significantly from what has been reported in research by Grassmann and Tomasello (2010) for how intonation interacts with information structure to affect attention during online reference resolution.

It is also important to explore attentional outcomes when no label is provided for an object, which is included as a control condition in this study. The control scenario examines infant attention in the absence of a referent label, addressing the broader question of how attention is affected by the presence or absence of a minimal linguistic label. As previously discussed, novelty in the visual domain has been shown to increase attention to a novel target over one that is familiar (Fantz, 1964; Rose et al., 1982). In addition, infants as young as nine months old prefer to look at an object that is

presented with its corresponding label (Oviatt, 1980), and they look longer to a target that is named over one that is not (Baldwin & Markman, 1989). Thus, these studies demonstrate the attentional facilitation of a linguistic label. The experimental conditions of this study manipulate the type of pitch accent used to label an object, while the control condition removes the label entirely, leaving only a visual *new/given* manipulation during the test phase.

2.2 Methods

2.2.1 Participants

Data were analyzed for 48 American English-acquiring 18-month-old toddlers (22 female). The age of participants ranged from 529 to 589 days, with a mean age of 551 days. Data from ten additional participants were discarded due to fussiness (8), or equipment malfunction (2). All participants were from Providence, RI, and surrounding areas. A second group of 12 American English-acquiring 18-month old toddlers (3 female) were tested as the control group (as discussed in §2.1). The age of this group of participants ranged from 535 to 579 days, with a mean age of 558 days. Data from four additional participants were discarded from this group due to fussiness (2), inattentiveness (1), or equipment malfunction (1). All participants received a t-shirt, book, or small gift at the time of their visit.

2.2.2 Stimuli

The same female speaker produced all target and distractor utterances. To create the three different pitch types, the speaker first produced carrier sentences with H* accents on the target words using careful (slow and clear) but not child-directed speech. The pitch contour of the target word only was then digitally manipulated in Praat (Boersma & Weenink, 2014) in order to create the two pitch movement versions, the simple monotonal and the complex bitonal. For the flat F_0 versions, deaccented productions were spliced into the same carrier phrase as the two pitch movement types (*simple* and *complex*) and then matched for duration and average intensity. All test stimuli were

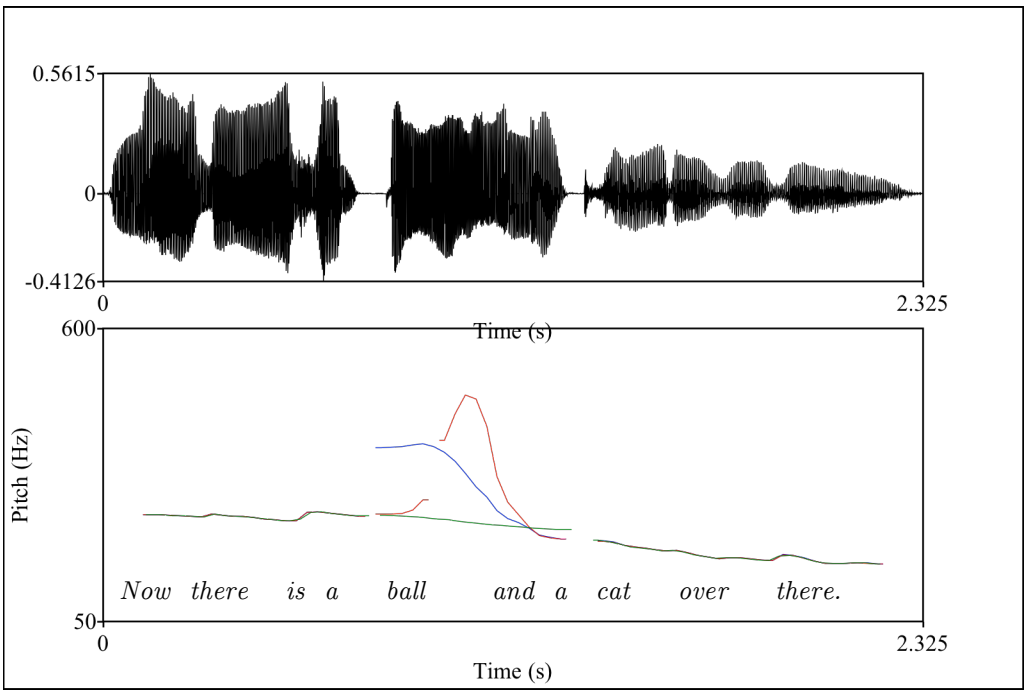


Figure 2.1. *Waveform and fundamental frequency contours for three types of pitch movements: flat F_0 (green line), simple monotonal (blue line), and complex bitonal (red line). This example utterance is "Now there is a ball and a cat over there", where "ball" is the target word being manipulated.*

resynthesized. Naïve listeners judged the resynthesized target speech sounds as sounding natural, although sometimes a bit faster than the rest of the utterance (particularly in the case of the complex bitonal pitch movement). Figure 2.1 shows the F_0 contours for each of the three pitch movement types for an example set of test utterances.

Six (C)CVC monosyllabic target words and 18 (C)CVC monosyllabic distractors/fillers were used in the procedure (see Table 2.1 for a full list of stimuli by trial). All words were phonologically distinct within a trial, and target words were primarily sonorant in nature in order to ensure a more reliable F_0 track. All of the target and distractor/filler words used are commonly known by 18-month-olds and conformed to the MacArthur Communicative Development Inventory (MCDI) norms (Fenson et al., 2007). Previous knowledge of the words was confirmed by a vocabulary questionnaire completed by the caregiver. If the toddler was not familiar with a particular word, then that particular trial was eliminated during analysis (this was not a common occurrence and affected only one trial for six different participants, for a total of six trials overall).

Table 2.1. *Stimuli for Experiment 1.*

Trial Number	Target Word	Distractors/Fillers
Practice	spoon	cake
		bear
		fish
1	ball	sock
		lamb
		cat
2	moon	dog
		shoe
		book
3	cow	pig
		tree
		duck
4	doll	bus
		cup
		sun
5	plane	star
		truck
		dress

2.2.3 Design and Procedure

This study employed a 2x3-mixed design to test toddlers' responses to variations in information status and intonation, isolating the specific role that pitch plays in directing attention to *new* or *given* referents. The two independent variables were Information Structure and Pitch Type. Information Structure was manipulated within-subjects and consisted of two levels: *new* versus *given.* Pitch Type was manipulated between-subjects and consisted of three levels: *flat F_0, simple monotonal,* and *complex bitonal.*

Each trial consisted of a Context Phase and a Test Phase. The Context Phase consisted of two parts (or slides), and established what information would be *new* or *given* at test (see Figure 2.2). Each slide was displayed for six seconds. The following Test Phase had three components. First, there was a center fixation point presented in

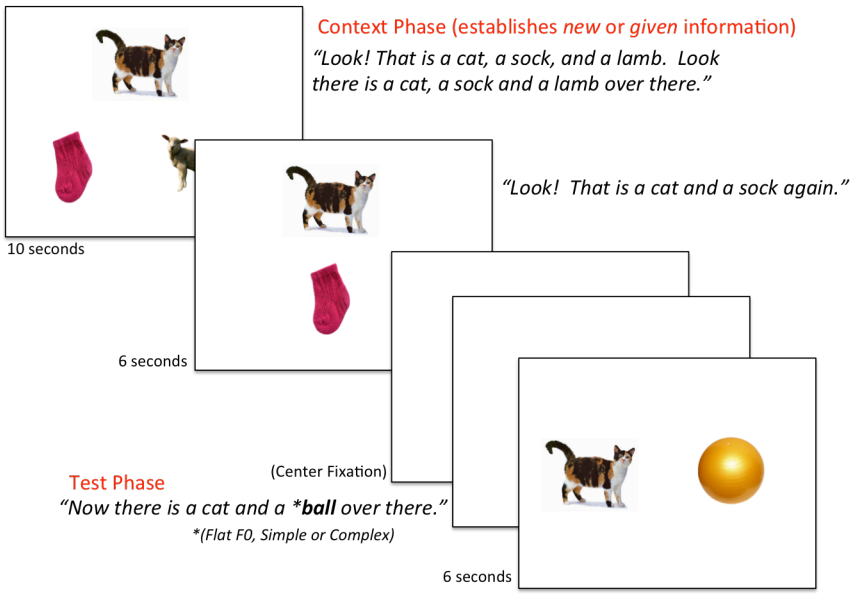


Figure 2.2. *Example new trial from Experiment 1. The target referent presented as deaccented, simple or complex depending on the condition. In a given condition, the target (e.g., 'ball') would have also been present in the first slide during Prior Discourse Context (e.g., in place of the 'lamb').*

order to ensure that each test trial began with a neutral look to the center, eliminating a left or right side bias once the images were displayed. Second, after the participant looked at the center fixation point for a period of 500ms, the test utterance began playing. The final test slide containing the two images of the referents (the target and the distractor) appeared after both the target word and the distractor word were finished being auditorily presented in the carrier phrase, and just as the utterance was finishing (synced to the onset of '*...over there*'). While the final test slide was shown for a period of six seconds, the *proportion of looking time (PLT) to the target* was collected using a SMI iViewX™ RED eye tracker. Importantly, the target item in the test slide was either *new* or *given* in the discourse context, as well as in contrast to a referent in the previous slide, making both the simple (˜H*) and the complex (˜L+H*) movements acceptable in this context.

There were five warm-up trials and two test blocks per condition. The warm-up trials presented and labeled familiar objects in the same configurations as would appear during the test trials (in groups of two or three). The warm-ups served as an opportunity for the child to become comfortable with the procedure and the testing environment. The warm-ups also provided time to verify that the eye tracker was reliably recording eye data. Each of the two test blocks tested one of the within-subject Information Structure levels: one block for *new* and one block for *given.* Pitch Type was tested between-subjects. Each test block included one practice trial and five test trials. With two blocks per condition, there were a total of two practice trials and ten test

trials (twelve total trials) for each condition. Trial order within a block was randomized and block order was counter-balanced across participants. The location of the target item on the screen (left or right) was also counter-balanced within and across conditions.

A control condition was also included which eliminated the use of a test utterance during the Test Phase. Instead of hearing a test utterance such as "*Now there is a cat and a ball over there*", the participant only heard an attention getter word (e.g., "*Look!*") before the final test slide appeared. All timing remained the same as the other conditions. Thus, the independent variable of Information Structure (*new* vs. *given)* was still present, but the Pitch Type factor was eliminated during the Test Phase. Important to note is that the utterances during the Context Phase were still present in this control condition and kept consistent across all conditions by trial.

2.3. Results

The *control* condition produced results similar to that of the *flat* F_0 condition. Figure 2.3 illustrates the effect of removing the test utterance stimuli from the experimental design in comparison to the *flat* F_0 pitch type conditions. Paired t-tests comparing the group means between of the *flat* F_0 condition with the *control* condition revealed no significant differences between these two groups (*new* conditions: $t(8) = 0.155$, $p = .88$; *given* conditions: $t(8) = 0.185$, $p = .86$).

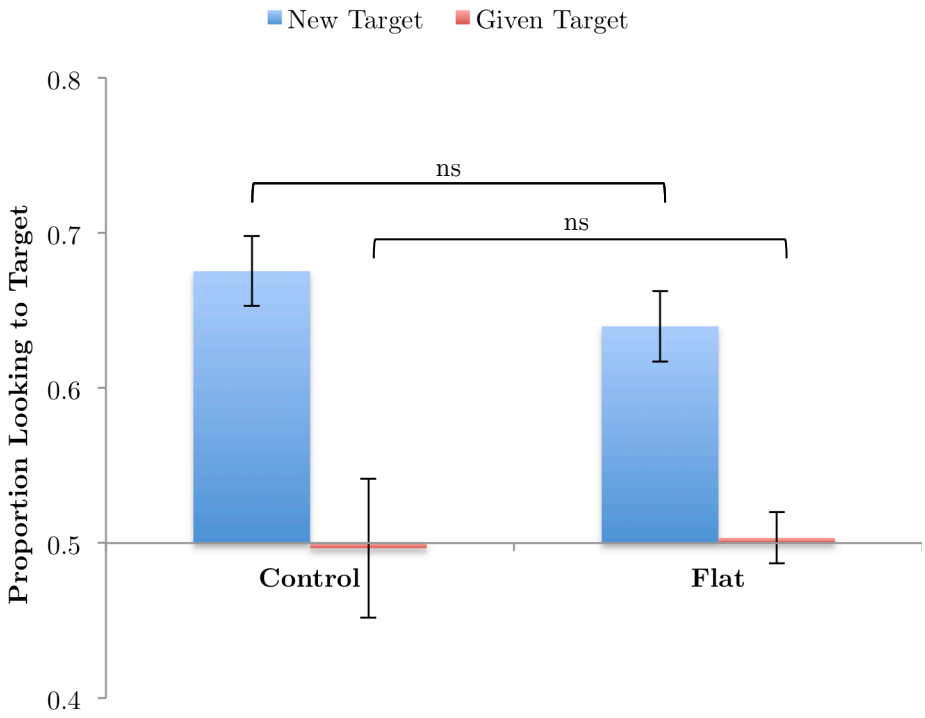


Figure 2.3. *Bar graph comparing proportion of looking to target in the control condition versus the three pitch type conditions. Error bars show standard error. ns = not significant.*

For the conditions that did receive a linguistic label at test, a 2x3 repeated measures mixed ANOVA showed a significant main effect of Information Structure ($F(1,45) = 28.26$, $p < .001$, $\eta_p^2 = .386$) and a significant main effect of Pitch Type ($F(2,45) = 5.03$, $p = .011$, $\eta_p^2 = .183$). The two-way interaction of Information Structure by Pitch Type was also significant ($F(2,45) = 4.33$, $p = .019$, $\eta_p^2 = .161$).

Planned comparisons between groups revealed more looking to a target than a distractor when the referent was *new* to the discourse context, regardless of pitch accent type. In the *flat* F_0 condition, a paired t-test showed that there was significantly longer looking to the target over a distractor only when the referent was *new*, not when it was *given* ($t(15) = 5.05$, $p < .001$) (see Figure 2.4). Additionally, both types of accentuation

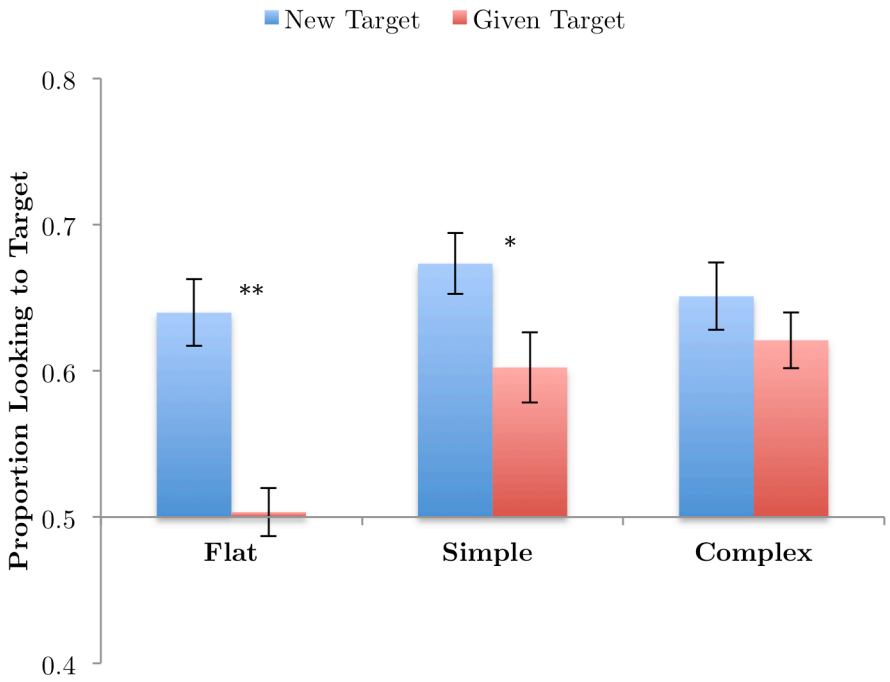


Figure 2.4. *Bar graph of proportion looking to a new or given target referent for each pitch type condition. Error bars show standard error. *: $p < .05$, **: $p < .01$.*

(*simple* and *complex*) guided more looks to the target than to the distractor when the target referent was either *new* or *given* to the discourse context. In the *simple* condition, there was significantly longer looking to the target when the referent was *new* to the discourse context than when it was *given* ($t(15) = 2.79$, $p = .014$). This difference in proportion of looking to *new* and *given* target referents was not significant in the *complex* pitch type condition ($t(15) = 1.21$, $p = .244$). Crucially, there were more looks to the target referent than the distractor in all conditions except when the target had a *flat* F_0 and was *given* in the discourse.

2.4. Discussion

Contrary to Grassmann and Tomasello (2010), but consistent with literature on visual attention and novelty, *newness* is sufficient to draw 18-month-olds' attention to referents in discourse, even without accentuation or linguistic labeling (i.e., the *flat* F_0 and the *control* conditions). A preference for the novel (or *new*) stimulus item over a familiar (or *given*) one suggests a more mature level processing by 18-month-olds (Fantz, 1964; Rose et al., 1982). Toddlers prefer to look at the more prominent or salient item, where salience in this case is achieved through pitch movement and visual novelty effects.

The *control* condition revealed the effect of removing the test utterance from the experimental design and how this impacted attentional processing. In this condition, the removal of the test utterance produced results comparable to the *flat* F_0 condition. This further teases apart the roles of visual discourse newness from pitch movement effects on discourse processing. Since there is no difference between the *control* and the *flat* F_0 conditions, it is reasonable to suggest that the novelty effect of the referent in the *new* condition is driving attention in these two conditions, and not the linguistic information.

As predicted, even in the case of a target referent that is *given* in the discourse context, both *simple* and *complex* pitch movements guide attention to the target. Thus, the presence of either *newness* or a *pitch movement* shifts attention to a target referent, regardless of pitch movement type. Again, this suggests that the more salient or prominent the stimulus item, the longer the toddler will look. For *given* information, the pitch movement is driving attention to the referent. When there is both newness and a

pitch movement, this results in even greater looking to the target. This was especially evident in the *simple* pitch movement condition, where there was significantly longer looking to the target when it was *new* than when it was *given* in the discourse context.

I also predicted that there would be an increase in looking time to referents introduced with a *complex* pitch movement. While there was greater looking to *given* referents produced with a *complex* pitch movement over ones produced with *simple* movements, this difference was not significant. One explanation for why the *complex* pitch type results were not as robust as predicted is that the only acoustic cue available was F_0. If more acoustic correlates of intonation were present (i.e., increased duration and intensity), I anticipate that this would lead to an increase in saliency and thus greater looking in the case of the *complex* bitonal pitch movement.

These results reflect only pitch movement variations on the target word. Since duration and average intensity were held constant across conditions, all effects were due to either the information status of the referent or its pitch movement type (i.e., F_0 movement only). An extension of this study will test what effects (if any) the other primary acoustic correlates of intonation have on toddler attention (e.g., duration, intensity). The question remains of how much each of the acoustic correlates of intonation weighs in directing attention both apart and together.

2.5. Conclusion

Experiment 1 tested how attention to a referent in a controlled discourse context is modulated by pitch movement type and information status. The three pitch movement types analyzed were a *simple monotonal* pitch movement, a *complex bitonal* pitch movement, and a *flat* F_0, and the information status of a referent was either *new* or *given*. Results showed that 18-month-olds' attention to a referent during discourse was driven by the presence of either discourse newness or a pitch movement on the target word. Critically, the methodological differences introduced changed the complexion of results in comparison to previous research by Grassmann and Tomasello (2010), but are consistent with the research on visual salience and novelty. In particular, I controlled for acoustic variation by using pre-recorded stimuli, isolated the specific role of fundamental frequency on attention, tested three pitch types, and collected more fine-grained fixation data using an eye-tracking paradigm. Furthermore, a *control* condition showed that visual discourse newness is sufficient to guide attention to a *new* referent in the absence of any linguistic cues. An ongoing extension of this study tests what happens when the full range of acoustic correlates to a pitch accent are employed (i.e., pitch, duration, and intensity vary naturally with each pitch accent type). Future work will also extend analyses to other types of pitch accents as well as other discourse contexts. This experiment lays the framework for testing how information structure and intonation interact to aid two-year-olds' in early word learning. A novel word learning task that

explores the effects of pitch accent type and contrastiveness on learning is presented in Experiment 2 in Chapter 3.

Chapter 3

Experiment 2:
Word Learning

3.1. Introduction

Experiment 1 demonstrated how 18-month-olds' attention to referents is mediated by both the intonation and the information structure of discourse. This chapter extends the attentional processing findings to investigate how intonation interacts with information status during a novel word learning task (Experiment 2).

Many factors combine to facilitate early word learning by infants and toddlers. Both cognitive and linguistic elements impact the ability of infants to learn the names of objects and actions in their environment. In the cognitive domain, some have argued that attention and memory lie at the root of successful word learning (Samuelson & Smith, 1998). On the linguistic side, others have argued that the manner in which children receive speech input is essential to this fundamental language process, such as through discourse novelty (Diesendruck, Markson, Akhtar, & Reudor, 2004) or word accentuation (Grassmann & Tomasello, 2007). Word learning entails association of linguistic labels with real-world referents, and thus it is a combination of both linguistic and cognitive processes working together that guides early word recognition and learning.

Successful word learning is dependent on three components: 1) *Segmentation* of the word from the speech stream, 2) *Identification* of the real-world referent, and 3) *Mapping* of the linguistic form (i.e., word) to the real-world referent (or concept) (see Waxman & Lidz, 2006, for a full review of early word learning). In order to create the mapping necessary between a word and its referent, a child must first be able to extract

that word from the speech signal as well as identify the corresponding real-world referent. Infants have been reported to be able to segment words from as early as 7.5-months-old (Jusczyk & Aslin, 1995). Since pauses do not necessarily indicate the beginning and ends of words in fluent speech, particular aspects of the speech signal are helpful for early word segmentation. For example, words that are at the end of an utterance are more easily recognized (Fernald, Pinto, Swingley, Weinberg, & McRoberts, 1998) and segmented (Seidl & Johnson, 2006). Also, words with specific rhythmic patterns are more readily segmented (Jusczyk, Houston, & Newsome, 1999), and highly familiar words serve as anchors for segmentation (Bortfeld, Morgan, Golinkoff, & Rathbun, 2005). Infant-directed speech prosody (e.g., exaggerated pitch range and specific intonational contours, including L+H*) has been shown to facilitate seven-month-olds' word segmentation in comparison to adult-directed speech prosody (Thiessen, Hill, & Saffran, 2005). In addition to word segmentation, children must identify the real-world referents to these linguistic units.

Linguistic, conceptual, and social processes all aid in word learning and specifically in the identification of real-world referents. For example, joint attention, eye gaze, and gesture (e.g., pointing) are some of the ways in which an adult can help an infant in locating and identifying objects in their environment (Samuelson & Smith, 1998; Brooks & Meltzoff, 2002; Moore, Angelopoulos, & Bennett, 1999; Woodward, 2003). Crucially, both segmentation and identification are necessary in order to accomplish the final step of creating a map between a linguistic element and its

conceptual referent. See Waxman and Lidz (2006) for a thorough review on how linguistic, conceptual, and social processes interact to support early word learning.

Of the elements that are essential for early word learning, this experiment is specifically interested in intonation and information structure. These two aspects are involved in both word *segmentation* (intonation) and referent *identification* (information structure). In particular, I investigate the role that specific pitch accents (H* and L+H*) play in situations where toddlers are learning *new* or *contrastive* information.

Several studies have shown that children are responsive to the *given-new* distinction in human speech and that they can appropriately use this information to respond to the discourse-pragmatics (Greenfield, 1979; O'Neill & Happé, 2000; Song & Fisher, 2005). A series of studies by Akhtar and colleagues investigated the interaction of intonation with information structure, and how the accentuation of *new* material aided in joint attention, word-to-object mapping, and word learning. Tomasello and Akhtar (1995) explored how the social-pragmatics of a specific context affected word learning and word-to-object or word-to-action mapping.

In the first of two experiments, Tomasello and Akhtar (1995) tested 27-month-olds in play sessions where the experimenter introduced a novel word while a nameless object was engaged in a nameless action. Results from this experiment clearly demonstrated that the children map the novel word either to the object or action depending on which was *new* to the discourse. While the first experiment only varied the pragmatic cue of discourse, their second experiment relied on the actions of the

experimenter to guide children in assigning the novel word to either a new object or action. In their second experiment, the two conditions were referred to as Action-Highlighted or Object-Highlighted. The conditions differed in how often the experimenter performed the nameless action upon the nameless object during a free-play session.

In the procedure for Action-Highlighted, the experimenter alternated turns with the child, each time demonstrating the nameless action with the nameless object while saying the novel word (e.g., '*Widget, Jason, widget*') (Tomasello & Akhtar, 1995). In the Object-Highlighted condition, the experimenter demonstrated the initial nameless action sequence only once and then used only language during the following turns instead of continuing to demonstrate the object performing the action. As with the first experiment, children mapped the novel word to either the new object or the new action depending on the experimental context. Results showed that the novel word was mapped onto the action in the Action-Highlighted condition, and onto the object in the Object-Highlighted. This study clearly indicates the power and importance of both discourse newness and social-pragmatic cues when a child is taught a novel word. While a child may exhibit a bias for mapping novel words to objects, this can be manipulated by altering the discourse setting.

Follow up work by Akhtar, Carpenter, and Tomasello (1996) further investigated the role of discourse novelty in word learning, and demonstrated the power of newness in a communicative environment. In this study, the experimenter would exclaim excitedly,

'There's a modi. Look, there's a modi!' to the child while gazing in the direction of four objects. The child had previously played with all four objects, but the experimenter was only present to play with three out of the four of them. Children accurately mapped the new word to the object that was new for the experimenter, indicating that children take into consideration newness from different individuals' perspectives, and that speakers like to talk about new things. A limitation of these experiments by Akhtar and colleagues is that they do not address in what manner the utterances were spoken to the participants. Since these were live play sessions, a substantial amount of variability is automatically introduced into the data. In reference to the AM approach and ToBI transcription system, these studies did not expressly state what types of intonational contours were used for the utterances or what types of pitch accents were assigned to the *new* discourse information.

Manipulating *newness* in discourse contributes to what the infant or child attends to, and to what they learn from in their environment. A counter-argument to the claims made by Akhtar et al. (1996) is that memory and attention serve as the basis for word learning, not the social-pragmatics and joint attention (Samuelson & Smith, 1998). Samuelson and Smith (1998) argued that word learning is based in rudimentary cognitive processes. To support this claim, they altered the experiment of Akhtar et al. (1996) by introducing the fourth object to the setting in a specific, intentional, and distinct way from the other three objects (as opposed to introducing it as *new* to the experimenter). When two-year-olds matched the novel word during the test phase with

the fourth target object, they claimed this was due to "mundane memorial and attentional processes" and not the linguistic events (Samuelson & Smith, 1998, p. 100).

A different interpretation of Samuelson and Smith's (1998) cognitively-driven explanation is that an intentional manipulation by the experimenter is needed in this type of experiment, in contrast to an accidental one, in order for the child to make the necessary word-to-object mapping (Diesendruck et al., 2004). This lends support to the explanation that children are able to use intentional information from the discourse-pragmatics when deciding where to focus their attention and learn a new word. Furthermore, it supports learning in scenarios where the child is sensitive to contextually or linguistically *new* information. This sensitivity is an important part of directing joint attention and consequently word learning. For an infant, these cognitive and linguistic processes must come together in order for word recognition and word learning to occur.

The studies discussed so far have demonstrated that children are sensitive to changes in discourse newness. Grassmann & Tomasello (2007) tested two-year-olds' word learning abilities by manipulating the accentuation of novel words during the learning phase. Four conditions were created to test the interacting roles of sentence stress (on either the noun or the verb) and discourse newness (either the object or the action is *new* to the discourse, as demonstrated by the experimenter). See Table 3.1 for a list of each condition with example novel linguistic expressions.

Table 3.1. *Grassmann & Tomasello's (2007) experimental conditions.*

CONDITION	EXAMPLE
Object new – Noun accent	*Der **FEKS** miekt*
Object new – Verb accent	*Der Feks **MIEKT***
Action new – Noun accent	*Der **FEKS** miekt*
Action new – Verb accent	*Der Feks **MIEKT***

In the test examples (Table 3.1), the first word is a German determiner followed by a nonce noun and then a nonce verb (each conforming to the phonotactics and syntactic patterns of German). While earlier studies have shown that children will 'fast map' novel names with novel nouns (Golinkoff, Hirsh-Pasek, Bailey, & Wenger, 1992; Mervis & Bertrand, 1994), Grassmann and Tomasello (2007) took this one step further and empirically demonstrated that children can specifically use prosodic highlighting of *new* material in a live discourse scenario to match a nonce word with a new referent in the environment. Another noteworthy aspect of this study is that children were able to match the nonce word with the new object when the noun was accented, but they were unable to do this in the 'Action new-Verb accent' condition as the authors' predicted. Why this particular inability occurs may be partially due to the increased difficulty in learning verbs and the ways in which verbs are tested during comprehension (an arguably more difficult task than testing object comprehension). Since the earlier studies by Akhtar and colleagues did not take into consideration prosodic highlighting as a factor, Grassmann and Tomasello (2007) demonstrates the important role that prosody plays in relaying intonational cues to what is *given* and/or *new* information in discourse.

Born out of the results of their 2007 study, Grassmann and Tomasello (2010) tested the hypothesis that newness and stress direct attention during word recognition. They showed that children could be reliably guided to a new target in a scene, but only when the accompanying speech accented the new referent. Though there were limitations with their 2010 study (discussed in Chapter 2), it is one of the first to look at how both newness and stress of a particular word influence toddler orientation to referents. If the presence of stress impacts word learning, then a further step would be to explore how specific types of contours and pitch accents affect this process. One prediction that emerges from this line of inquiry is that pragmatically appropriate pitch accents facilitate word learning by facilitating word segmentation as well as conditioning expectations for referent identification.

My experiment extends the findings from Experiment 1 to investigate how information structure interacts with intonation during early word learning. It follows that if these two factors (information structure, intonation) mediate attention, and attention is essential for mapping a word to its real-world referent, then they should also impact the word learning process. Experiment 1 showed that the story is likely more complex than previous research has suggested, and that different pitch accents as well as different aspects of the information structure each play an interacting role during reference resolution.

The current experiment explores how *contrastiveness* (contrastive, noncontrastive) interacts with *pitch accent type* (deaccentuation, H*, and L+H*) during

a novel word learning task. Contrastiveness here is defined as the introduction of a referent in direct opposition to a previously mentioned referent in the discourse. For example, if a bear and a sheep are introduced first in one scene (a slide), and then a bear and a pig in the next, the pig is in contrast to the sheep and the bear remains constant across the two scenes. If the second scene presents a pig with a duck instead, there is no contrast present, and the pig is simply new to the discourse in this case. As in Experiment 1, the two pitch accent types of primary interest are H* and L+H*, with a deaccented control condition as well. Watson et al. (2008) demonstrated that these two pitch accents overlap in speech perception and production, with H* used for both *new* and *contrastive* information and L+H* primarily reserved for *contrastive* information. Katz and Selkirk (2011) additionally showed that contrastive elements exhibit longer duration, higher F_0, and more F_0 movement than discourse-new elements.

In contrast to Experiment 1, which only manipulated fundamental frequency, this study employs the full range of acoustic cues to differentiate pitch accent types (i.e., duration, pitch, intensity, and segmental quality). In the discussion of Experiment 1, I predicted that there would be increased looking to the target in the complex bitonal (~L+H*) condition if additional cues had also been present, even when the target was *given*. Since this experiment does not isolate individual acoustic cues, the L+H* pitch accent is more salient (i.e., prominent) than the H* pitch accent. This follows recent research by Turnbull et al. (2014), who demonstrated that the rising bitonal pitch accent (L+H*) is more prominent than a monotonal one (H*). Thus, with the complete range

of acoustic cues available during attentional processing, there should be an increase in learning for the more prominent elements.

My primary prediction for this study is that the more prominent a novel word is during learning, the better a child will learn the word-object association. As Experiment 1 showed, attention is mediated by both the information structure and the intonation. If attention is increased during the more salient (and felicitous) learning contexts, this will aid in extracting the label from continuous speech and making the correct word-to-object mapping. In particular, the contrastive context and the complex L+H* pitch accent are more prominent than the noncontrastive context and the H* or deaccented productions (Katz & Selkirk, 2011; Turnbull et al., 2014). See §3.2.4 for a more detailed list of predictions by condition. The overall aim of this study is to investigate how *contrastiveness* and *pitch accent type* affect the learning of a novel word.

3.2 Methods

3.2.1 Participants

Data were analyzed for 36 American English-acquiring 24-month-old toddlers (17 female). The age of participants ranged from 712 to 799 days, with a mean age of 743 days (2;0.13). Data from fourteen additional participants were discarded due to fussiness (6), inattentiveness (5), or equipment malfunction (3). All participants were from

Providence, RI, and surrounding areas. All participants received a t-shirt, book, or small gift at the time of their visit.

3.2.2 Stimuli

The same female speaker who recorded the stimuli for Experiment 1 produced all utterances using careful (slow and clear), but not child-directed speech. The speaker was trained in intonational phonology and the ToBI transcription system, and she was able to produce H* and L+H* pitch accents consistently across the different stimulus utterances. Each target word was produced naturally in an utterance with an H* accent, a L+H* accent, or as deaccented. The target word was then spliced out of its original utterance and into a carrier sentence. This ensured that the only difference between the different pitch type conditions was the intonation of the target word, and not the other parts of the sentence. All splices were made at zero-crossings.

Stimuli included two CVC monosyllabic target novel words and 4 CVC monosyllabic distractors (see Table 3.2 for a full list of stimuli by learning condition). All target words were phonologically distinct. All of the distractor words were animals commonly known by 24-month-olds. Previous knowledge of the words was confirmed by a vocabulary questionnaire completed by the caregiver. Novel words were associated with novel animals that were created for this experiment. There were four novel animals, two for each learning condition (one as a target and one as a distractor).

Table 3.2. *Audio stimuli for Experiment 2.*

Learning Condition	Novel Target Word	Distractors/Fillers
Noncontrastive	*wug* (IPA: /wʌg/)	sheep bear pig
Contrastive	*neem* (IPA: /nim/)	duck bear pig

3.2.3. Design and procedure

Participants were tested using a 2x3-mixed design with *contrastiveness* (within-subjects) and *pitch accent type* (between-subjects) as independent variables. *Contrastiveness* varied how the target word was presented during learning, either *contrastive* or *noncontrastive* in relation to a preceding discourse element/referent (see Figure 3.1 for an example of a noncontrastive learning condition). The *pitch type* variable had three levels: a novel target word could bear a *simple* monotonal (H*) pitch accent, a *complex* bitonal (L+H*) pitch accent, or a *deaccented* pattern.

In a *contrastive* learning condition, the target word was always presented in opposition to a referent in the preceding discourse element (e.g., *'Look! There is a bear and pig over there. Oh! There is a bear and a* $WUG_{contrastive}$ *over there.'*). In a *noncontrastive* learning condition, the target novel word was presented as presentationally new in relation to the previous discourse referents (e.g., *'Look! There is a bear and pig over there. Oh! There is a sheep and a* $wug_{noncontrastive}$ *over there.'*) (see

Figure 3.1). In the *noncontrastive* condition, the novel target was never presented in contrast to a previous referent.

An experimental condition consisted of a two Learning Phases and one Test Phase. The Learning Phase consisted of two 12-slide learning blocks (one *contrastive* and one *noncontrastive condition*, see Appendix A for a list of all learning blocks and test stimuli). A different novel word was presented in each of these learning blocks. Each

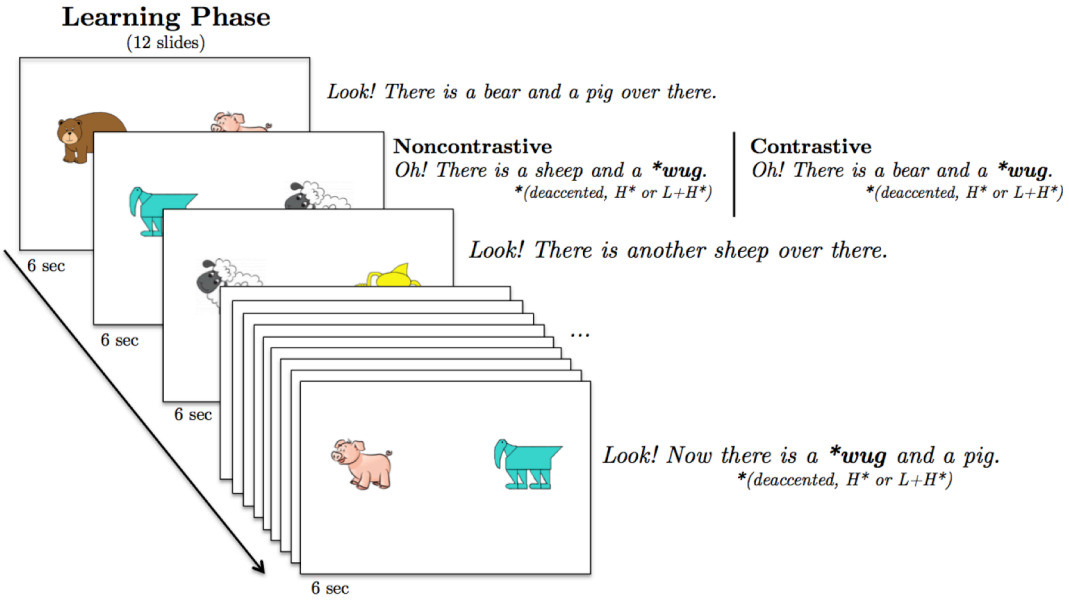


Figure 3.1. *Example noncontrastive Learning Phase from Experiment 2. The novel target word 'wug' can bear one of three pitch accent types depending on the condition.*

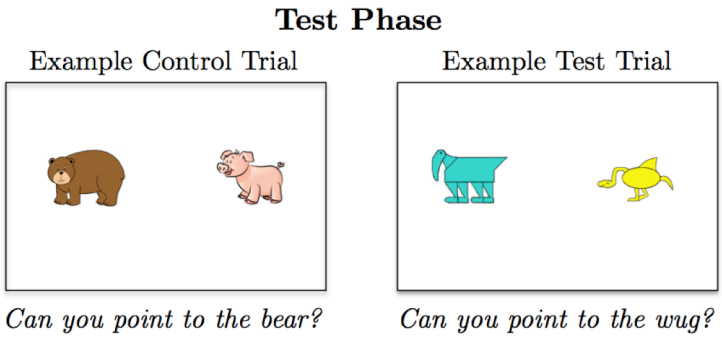


Figure 3.2. *Example control and test trials from the Test Phase of Experiment 2.*

block included 4 target slides (where the novel word-animal pairing is presented), 4 distractor slides (where a second unnamed novel animal is presented), and 4 filler slides of familiar animals. The Test Phase consisted of 8 trials: 4 test trials and 4 control trials (see Figure 3.2). During a control trial, two familiar animals were presented and the participant was asked to identify one of the referents. During a test trial, the target novel animal was presented with the distractor novel animal. Fixations to the target and the distractor were collected during the Test Phase to assess whether or not the child learned the novel word-animal pairings. Each test trial was six seconds in total duration. The onset of the critical word began at 3 seconds (3000ms) after the start of the trial (see Figure 3.3). The same eye-tracking paradigm and setup as Experiment 1 was used (i.e., the SMI iViewX™ RED was used to capture the eye fixations of the participant while they sat on their caregiver's lap).

Block order and novel animal assignments were counter-balanced across conditions. Test trials were randomized and the location of the target item (left or right) was counter-balanced within and across conditions.

In order to collect a vocabulary measurement for each child participant, the caregiver completed the age appropriate MacArthur Communicative Development Inventory (MCDI), Words and Sentences Short Form at the time of the laboratory visit (Fenson et al., 2007).

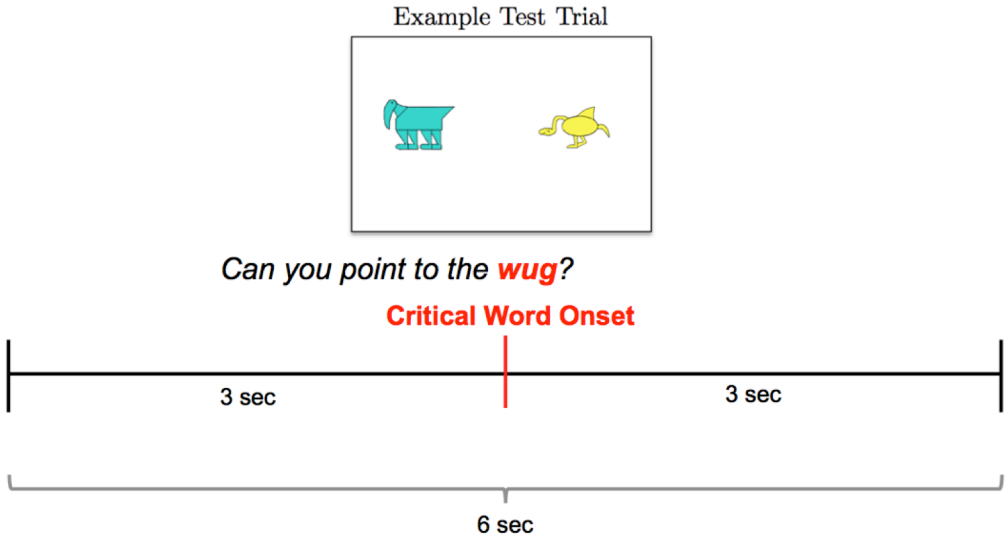


Figure 3.3. *Duration of an example test (and control) trial. Each trial lasted 6 seconds with the critical word onset beginning at the 3-second mark.*

3.2.4. Analysis

A time window of analysis was selected based on prior research as well as an evaluation of the control trials, which verified that subjects looked to the familiar target animals at test (statistics for the control trials are reported in §3.3 Results). Importantly, overall fixations to the target increased about 400ms after the onset of the target word, and decreased about 1500ms later. Given these observations, a 2000ms window of analysis (from 300-2300ms after onset of the critical word) for the control and the test trials was selected (see Figure 3.4). This window was sufficient to observe processing effects that may slow down or speed up gaze behavior and performance. These processing speeds are consistent with past literature showing that children process linguistic information at a slower rate than adults (beginning roughly 350ms after the onset of the critical word)

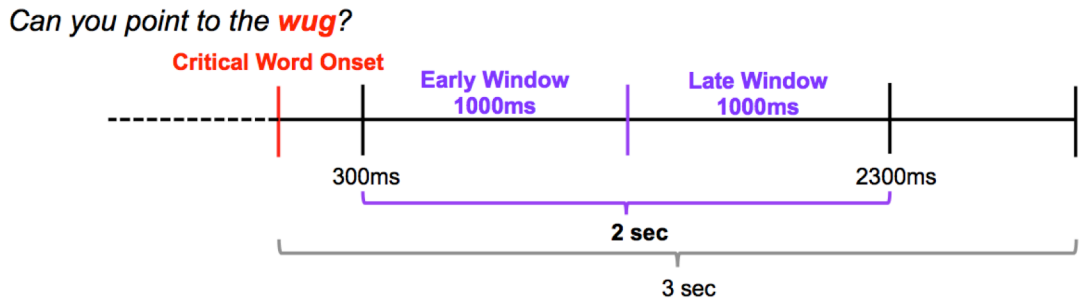


Figure 3.4. *Selected 2000ms time window for statistical analysis, ranging from 300ms to 2300ms after the onset of the critical word. Within the 2000ms range, there is an early window (300-1300ms) and a late window (1301-2300ms).*

(Matin, Shao, & Boff, 1993; Swingley & Fernald, 2002; Barr, 2008). See Figure 3.5 for a line graph displaying overall proportion of fixations to the target over the course of the 2000ms window of analysis for the control trials. In addition, the 2000ms window was split evenly into two 1000ms regions: 1) an *early window*, from 300-1300ms, and 2) a *late*

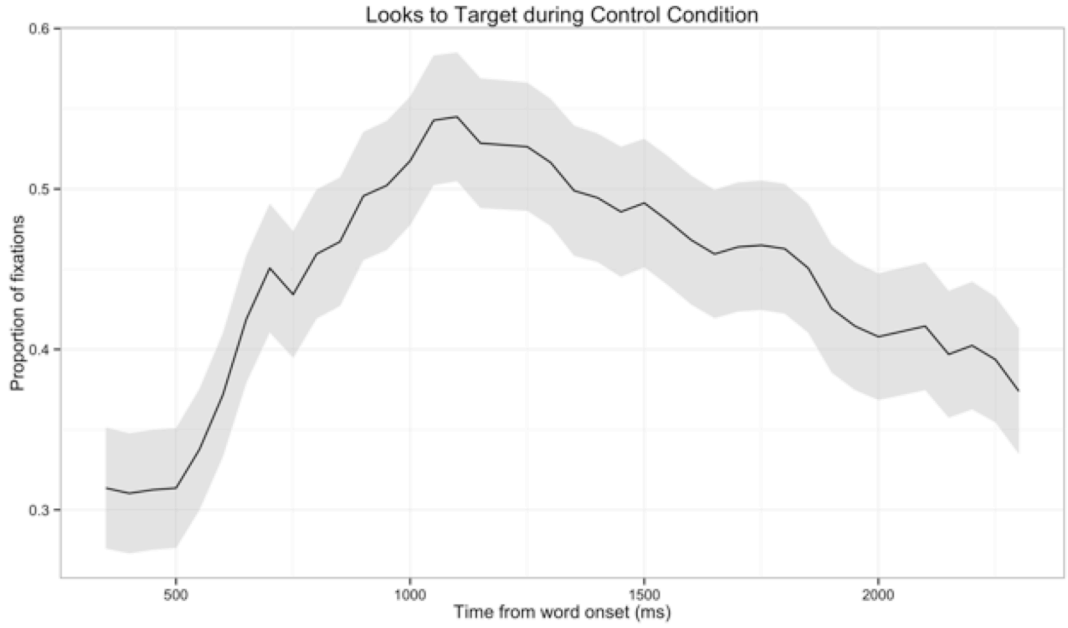


Figure 3.5. *Line graph of overall proportion of fixations to target over the 2000ms time window for the control trials (from 300-2300ms after the onset of the critical word). Shading shows standard error.*

window, from 1301-2300ms (Figure 3.4). This was included as an additional Time variable (*early* vs. *late*) in an effort to analyze how fixations to the target may change over the course of processing.

The dependent measure used for analysis was an advantage score, which demonstrates which referent (the *target* or the *distractor*) drew more looking after hearing the critical word. This score was calculated by subtracting the proportion of fixations to the distractor from the proportion of fixations to the target over the different windows of analyses.

3.2.5. Predictions

The first prediction is that because both the *simple H** and the *complex L+H** pitch accents are felicitous in a *contrastive* learning scenario, looking to the target should therefore increase in both of these contexts in comparison to the other conditions (Watson et al., 2008). Also for the *contrastive* contexts, increased looking to the target is predicted when the novel word is introduced with the more prominent L+H* pitch accent (vs. the *simple* monotonal H*) (Turnbull et al., 2014).

For *noncontrastive* contexts, the simple H* pitch accent is the most contextually appropriate (Watson et al., 2008). Moreover, Experiment 1 demonstrated that this pitch accent guides attention to a referent, even if it is *given* in the discourse. Therefore, the second prediction for Experiment 2 is increased learning of a novel word if it bears a *simple H** pitch accent during a *noncontrastive* learning condition (vs. *no accent*). Although the L+H* pitch accent is not contextually appropriate in a *noncontrastive*

context, the salience of this pitch accent may interact with the context and drive attention to the referent. Thus, the third prediction is that learning is expected to a lesser degree in the *noncontrastive-L+H** than the *contrastive-L+H** condition.

The fourth prediction is no learning is expected when the novel word is introduced with *no accent*, for either the *contrastive* or the *noncontrastive* learning conditions. This prediction is based on results from Experiment 1, showing that a deaccented word, if already *given* in the discourse, is not sufficient to guide attention to a referent. Overall, the *deaccented* conditions should exhibit less looking to the target at test than the other two pitch accent conditions.

3.3 Results

3.3.1 Control Trials

An analysis of the control trials over the 2000ms analysis window revealed a significant main effect of Time ($t = 2.48$, $p = .01$), with more looking to the target in the *late window* than the *early window.* This demonstrates an increase in proportion of fixations to the target over the 2000ms window of analysis. See Figure 3.5 for a line graph showing overall proportion of fixations to the control target for this 2000ms time window (versus the distractor or looking away).

3.3.2 Test Trials

I began analysis of the data using a conventional approach and performed a 2x3 repeated measures ANOVA with Contrastiveness (2-level, within-subjects: *contrastive*, *noncontrastive*) and Pitch Accent Type (3-level, between-subjects: *no accent*, *simple monotonal H**, *complex bitonal L+H**) as independent variables. The dependent variable was the proportion of fixations to the target during the 2000ms analysis window time-locked to begin 300ms after the onset of the target word (see Figure 3.3). Past literature has shown it takes a minimum of 180-200ms from the word onset to observe processing effects and gaze behavior driven by the linguistic signal (Matin, Shao, & Boff, 1993). For children (as well as adults viewing unfamiliar referents) this number has been said to be almost 100 to 200ms later (Haith, Wentworth, & Canfield, 1993; Canfield, Smith, Brezsnyak, & Snow, 1997; Swingley & Fernald, 2002; Barr, 2008). Thus, I employed a conservative time-lock of 300ms after the onset of the critical word to analyze the data.

The 2x3 repeated measures ANOVA did not reveal any main effects or interactions. Given these results, I examined more closely the time-course data over the 2000ms window of analysis. Figure 3.6 displays line graphs showing the proportion of fixations to the target for this time window. It is evident from these graphs that there are differences in the data over time (particularly when comparing the first 1000ms to the second 1000ms), but the proportional looking measure failed to capture these nuances. This is particularly apparent in the noncontrastive conditions, where there is no change in fixations until the later window.

An empirical logit regression was used to analyze the time-course data (Barr, 2008), which attempts to fit a linear model to the data. As with the ANOVA, the same 2000ms window was selected for analysis and Contrastiveness and Pitch Accent Type were the fixed factors. Data were aggregated over 50ms time bins (at a sampling rate of 120Hz this is equal to 6 frames of eye data) for each cell of the design and for each subject, ensuring that the data were not oversampled. Due to the observation that there were unique effects occurring early and late in the window of analysis, the 2000ms window was split in half for analysis: 1) the first 1000ms was labeled the *early window*;

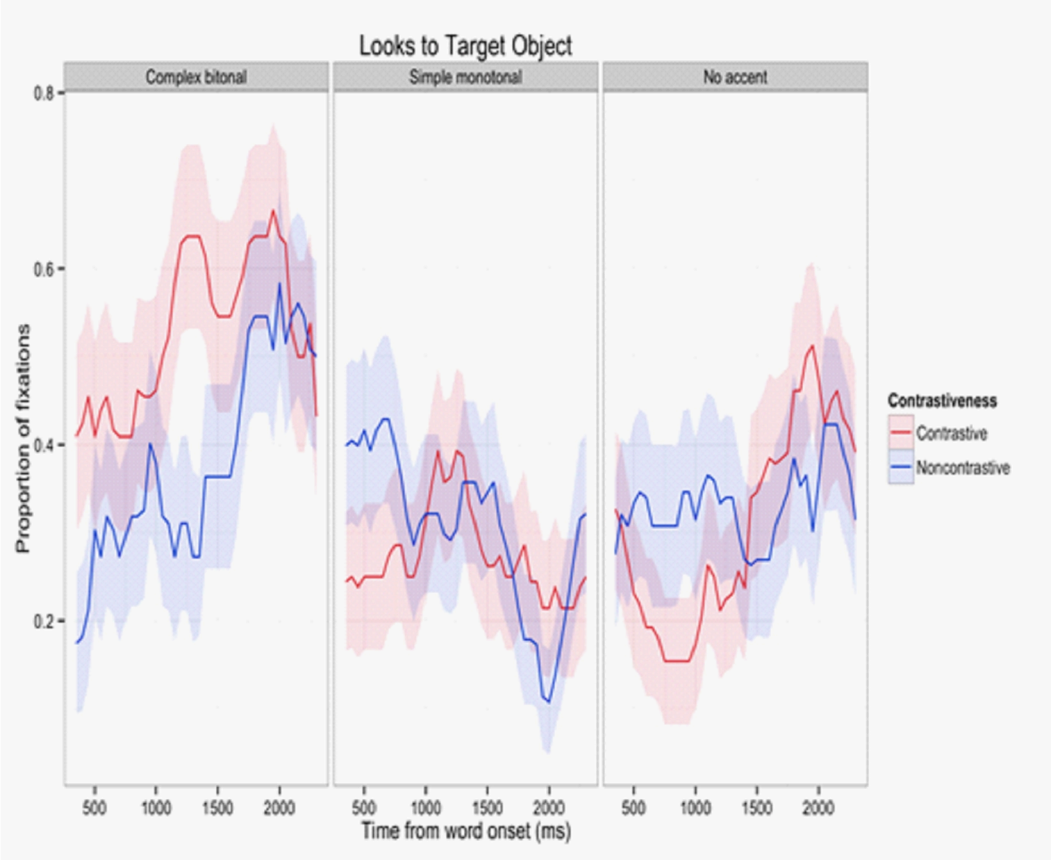


Figure 3.6. *Line graphs of proportion of fixations to target over the 2000ms time window (300-2300ms). There are separate graphs for each pitch accent type. Contrastive conditions are in red and noncontrastive are in blue. Shading shows standard error.*

and 2) the second 1000ms was labeled the *late window* (See Figure 3.7). This yielded a new factor of Window (*early* vs. *late*; similar to the Time variable used in the ANOVA analysis). The data were also transformed onto the log odds scale for the empirical logit regression. An omnibus model was first performed with Window included as a fixed factor. This was followed by separate analyses of the *early window* and the *late window* time periods.

Following Barr (2008), an empirical logit regression was conducted using the statistical software package R (R Development Core Team, 2007), with the lme4 package (Bates, 2007) to obtain parameter estimates and package languageR (Baayen, 2007) to obtain *p*-values. The 'lmer' function yields *t* statistics but the degrees of freedom are not estimated. To obtain *p*-values, a Markov Chain Monte Carlo (MCMC)

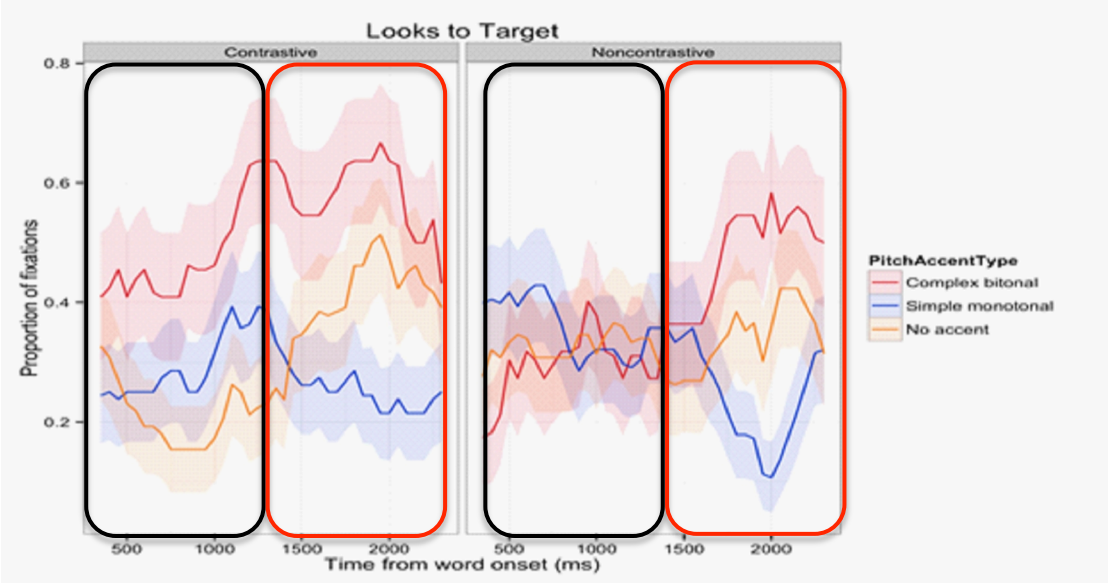


Figure 3.7. *Line graphs of proportion of fixations to target over the 2000ms time window, with separate graphs for contrastiveness. Pitch accent types are plotted in each graph. A black box indicates the early window; a red box indicates the late window. Shading shows standard error.*

method was used (Baayen, Davidson, & Bates, 2008).

No participants were eliminated on the basis of performance during the test trials. This was based on two factors. First, even when trials were eliminated due to less than 30% looking to the target or distractor during a trial, the results were not significantly altered (10% of trials met this criteria). Second, if this cutoff point had been used to eliminate trials from the analysis, it would have resulted in massive data loss and less power for the statistical analyses. Thus, all participants and trials were included in the final analysis.

An omnibus model was performed and included Pitch Accent Type, Contrastiveness, and Window as fixed effects. Since the Pitch Accent Type contained three levels, the *simple monotonal H** was selected as the reference condition for this analysis. There was no main effect of Contrastiveness for any of the omnibus models. The main effect of Window was significant ($t = 4.04$, $p < .001$). The *complex bitonal L+H** differed significantly from the *simple monotonal H** ($t = 2.66$, $p < .01$), with more looks to the target in the *L+H** condition. Two-way interactions of the pitch accent types (*complex bitonal* and *no accent*) by Window were significant (Complex by Window: $t = 8.66$, $p < .001$; No Accent by Window: $t = 8.98$, $p < .001$). Two 3-way interactions were also found to be significant (Complex by Window by Contrastiveness: $t = 2.40$, $p < .01$; No Accent by Window by Contrastiveness: $t = 4.69$, $p < .001$). These indicated more looking to the target for these two accents (vs. the *simple monotonal H**) in the *contrastiveness* conditions, and that this difference was mediated by early versus

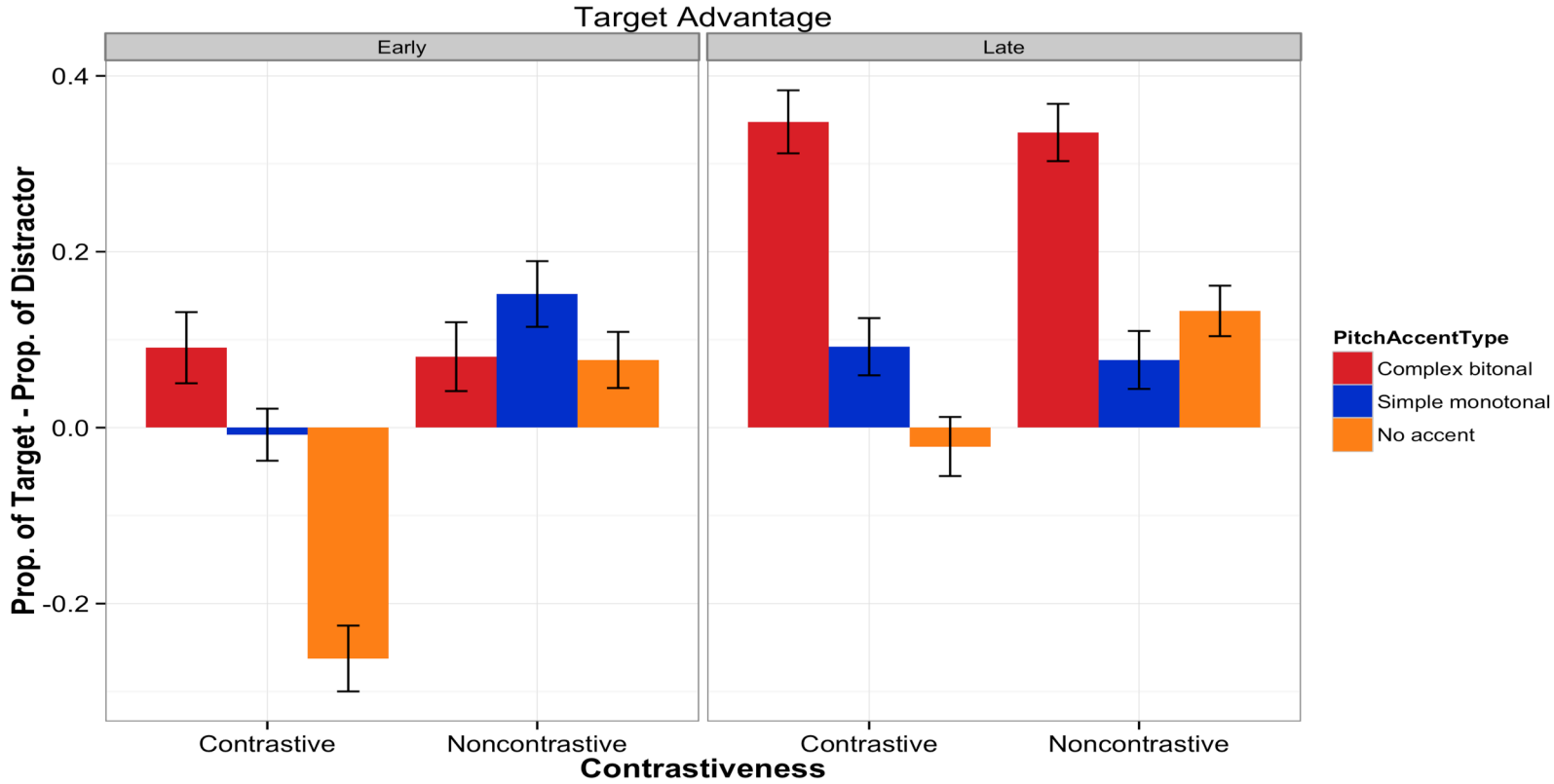


Figure 3.8. *Bar graphs of target advantage scores (proportion of fixations to the target minus proportion of fixations to the distractor). Analysis window begins 300ms after the onset of the target word and continues for 2000ms. Left panel (early) shows the first time window of 1000ms; right panel (late) shows the second time window of 1000ms. Each panel shows the between-subjects contrastive and noncontrastive conditions, with bars for each pitch accent type. Red bars show complex bitonal, blue bars show simple monotonal, and orange bars show no accent condition. Error bars show standard error.*

the late processing. (Figure 3.8 shows bar graphs of the proportion of fixations to the target for each condition by window of analysis (*early* vs. *late*)).

Taking the same approach but with the *complex bitonal L+H** as the reference condition, there was still a main effect of Window ($t = 8.06$, $p < .001$). There were also significant 2- and 3-way interactions mediated by Window and Contrastiveness (Simple by Window: $t = 8.63$, $p < .001$; Window by Contrastiveness: $t = 3.93$, $p < .01$; No Accent by Window by Contrastiveness: $t = 6.91$, $p < .001$), showing that contrastiveness has an effect in the *early window*, but not the *late window*.

Finally, a third omnibus version was run where *no accent* was used as the reference condition. There was only a significant main effect of Window ($t = 8.663$, $p < .001$), with neither Pitch Accent Type nor Contrastiveness reaching significance. The 2-way interactions of Simple by Window ($t = 8.98$, $p < .001$) and Contrastiveness by Window were significant ($t = 5.90$, $p < .001$). These data revealed more looking to the target in in the *simple* condition during the *early window*, and more looking in the *contrastive* conditions during the *late window*. The two 3-way interactions were also significant (Simple by Window by Contrastiveness: $t = 4.69$, $p < .001$; Complex by Window by Contrastiveness: $t = 6.91$, $p < .001$).

Next, the *early* and the *late windows* were analyzed separately. For the *early window*, regardless of which Pitch Accent Type condition was used as the reference condition, the only significant result was a 2-way interaction between Pitch Type (*complex L+H** vs. *no accent*) and Contrastiveness ($t = 2.54$, $p = .01$). The analysis

revealed that there was more looking during the *early window* in the *L+H*-contrastive* condition than the *L+H*-noncontrastive* one, and that there was more looking to the target in the *no accent-noncontrastive* condition than in the *no accent-contrastive* one. There were no significant main effects.

An analysis of the *late window* revealed two significant main effects. The *complex bitonal L+H** and the *no accent* pitch accents each differed significantly from the reference *simple monotonal H** condition (t =3.94, $p < .001$; t =2.18, $p < .03$), with more looking to the target in the *L+H** and the *no accent* conditions. There was also marginally more looking in the *complex L+H** condition than in the *no accent* condition ($t = 2.14$; $p = .06$). There were no significant interactions.

3.4 Discussion

The 2x3 repeated measures ANOVA that was first performed revealed no significant results. Since measures of overall proportional looking time failed to account for nuanced differences over time, an empirical logit regression was performed to reanalyze the data. This method proved fruitful by demonstrating how pitch accent type and contrastiveness interact to influence novel word learning.

The control trials and the test trials were analyzed separately. The control trials served two primary purposes. First, they ensured that the participants were completing the test phase successfully. Second, these trials provided a baseline of data to compare the learning conditions. The control condition was considered the standard for looking

patterns to a target stimulus, as well as whether or not the word was successfully associated with the object.

The results demonstrate that children learn novel word-to-object associations better when the word was produced with the *complex L+H** as compared to the *simple H** pitch accent. There is also a processing advantage for *L+H** over *H**, with children fixating faster to the target referent when the novel word was learned during the *complex L+H** condition. Additionally, a learning advantage was exhibited when the *L+H** accented word was introduced in a *contrastive* setting. This was true early and later in processing.

During the *early window* (from 300-1300ms after the critical word), the *complex L+H** showed a learning advantage when it was presented in a *contrastive* context. This supports the first prediction that learning is enhanced for the most salient or prominent conditions, which is the *contrastive-L+H** learning condition. The *complex L+H** accent also showed more fixations to the target than both the *simple H** and *no accent* types. Earlier in processing, the *simple H** and the *no accent* pitch accent types did not differ from one another for either the *contrastive* or the *noncontrastive* conditions.

Later in processing (from 1301-2300ms), words that had been learned bearing a *complex L+H** still showed an advantage, with more looks to the target than the *H** or *no accent* conditions. This falls in line with the prediction that L+H* is the most prominent pitch accent. Interestingly, no significant effect of contrastiveness was observed for any of the pitch accent type conditions by this later point in processing.

Perhaps one of the reasons there is an advantage for *L+H** in the *contrastive* condition is because this is the context where this pitch accent most frequently occurs in adult-directed speech. As predicted, the prominence of the L+H* pitch accent boosts learning in all conditions, regardless of context. Contrary to the predictions, the H* pitch accent did not show increased learning as compared to the *no accent* condition. Although there was a trend showing the H* pitch accent outperforming the *no accent* condition in *noncontrastive* settings (as predicted) early in processing, this trend did not reach statistical significance.

The *no accent* conditions yielded unexpected results. Early in processing, the *no accent* conditions showed no effect of learning. Later in processing though, there is an increase in fixations to the target when the deaccented target words were introduced in the contrastive learning condition. This may indicate the prominence of the *contrastive* scenario and its capacity to draw attention to a referent (versus the *noncontrastive* condition). An analysis of the line graphs (Figures 3.5 and 3.6) shows that learning does not seem to be occurring in these conditions. A more sensitive analysis may be necessary in order to better explain what is occurring in the *no accent* conditions.

Overall, the results suggest that the *complex L+H** pitch accent guides more looking to the target novel animal at test than the other pitch types. This was mediated by contrastiveness, suggesting that words introduced in contrast to a previous referent and that carry a L+H* pitch accent are learned better than words introduced with an H* or in a noncontrastive setting. Thus, the general pattern of results suggests learning

of the novel word-animal association is enhanced in the complex pitch accent conditions. Finally, the results showed processing differences across pitch accent conditions. Any effect of H* is seen earlier in processing, while the L+H* continues to gain more looking to the target over time.

3.5 Conclusion

The empirical logit regression was more effective in analyzing the data than the conventional 2x3 repeated measures ANOVA, insofar as the regression was able to capture certain nuanced behavior that the ANOVA could not. Still, this approach is not fully able to account for all of the patterns in the data. Due to the age of the population (2;0) and the tendency for their attention to vary considerably over time, an even more sensitive analysis would be appropriate. The empirical logit analysis attempts to fit a linear model to the data, but closer examination shows subtle changes in looking patterns over time. Higher order polynomial functions may better capture these changes in gaze behavior over time. Given the obvious changes and reversals in direction of the data, a more appropriate analysis would use growth curve modeling, which would look at patterns over a continuous time interval.

Nonetheless, the current analysis revealed several significant findings for how pitch accent type and contrastiveness interact to affect novel word learning. As predicted, the condition that was the best for learning the word-object association was the *contrastive* setting where the target word was introduced with a *complex L+H** pitch

accent. This condition showed effects early and late in processing in comparison to all of the other learning conditions. Later in processing, there was evidence that children learned the novel word when it was introduced in the noncontrastive context with a L+H* pitch accent. This is contrary to the idea that learning may be impaired in infelicitous contexts.

There are two possible explanations for this effect. First, learning may occur in the noncontrastive L+H* condition due to the salience and prominence of this particular pitch accent (Turnbull et al., 2014). Second, the L+H* is frequent in child-directed speech (Yu, Khan, & Sundara, 2014), and the high amount of exposure to this pitch accent may increase its acceptability. Thus, these results indicate that the prominence of the pitch accent may override the contextual appropriateness, since even a L+H* used in a noncontrastive setting shows increased effects of learning. The other two pitch accents have a less clear effect on learning within the current analysis, but there are indications in the data that the contrastive scenario is aiding learning even when the novel target is produced with an H* or as deaccented. Growth curve modeling will be able to better assess differences in looking patterns over time for the H* and no accent conditions.

Experiments 1 and 2 demonstrate the complex interaction of intonation and information structure on early word recognition and learning. There is a significant impact on attention and novel word learning abilities when the pitch accent of a referent and the discourse context are manipulated. Both of these experiments show that children are sensitive to these processes, even if not yet in a fully adult-like manner. Still, an

early perceptual sensitivity to changes in discourse and intonation proves beneficial for directing infant attention to the most prominent aspects of a communicative interaction. Child-directed speech may also be playing a critical role in processing at these young ages (1;6 and 2;0). If children learn primarily from child-directed speech, they cannot be expected to reflect patterns that are only present in adult-directed speech. Further research is necessary to explore how information structure categories are expressed in child-directed speech. Experiment 3 complements the first two experiments and analyzes how intonation patterns in child-produced speech compare to those in adult speech as the information structure varies.

Chapter 4

Experiment 3:

Speech Production

4.1. Introduction

The previous two chapters establish the importance of the interaction between information structure and intonation during early word recognition and learning. This chapter examines a different component of the developmental timeline and investigates how information structure and intonation are reflected in early child speech. Experiment 3 consists of two components. The first part verifies how adults use intonation to reflect informationally-structured categories. This component will serve as both a baseline for comparison with the toddler population, and an opportunity to replicate previous findings from the adult literature. The second part examines how young toddlers create these same distinctions as the adults in their own speech. For both adults and toddlers, I will analyze the phonological categories employed as well as the acoustic-phonetic underpinnings of those categories.

Experiment 1 demonstrated that for early receptive language, toddlers use prosodic cues in order to identify the most salient/prominent information in the discourse (cf. Fernald, 1985, for CDS; Arnold, 2008; Grassmann & Tomasello, 2010; Thorson & Morgan, 2014). Experiment 2 showed how this process of incorporating the information structure and intonation is critical in facilitating early word learning. The primary objective of this study is to analyze the interaction between intonation and information structure in toddlers' speech production. Intonation can be used in English to relay the information status of different components of an utterance (Jackendoff, 1972; Steedman, 1991; Pierrehumbert & Hirschberg, 1990; Vallduví & Engdahl, 1996). In

general, *new* information is typically accented, while *given* information is often deaccented in adult speech (Ladd, 1996). If attention and word learning are in fact mediated by intonation and specific pitch accents, how will children use these contours to reflect information structure in their own speech? Here, I ask how toddlers phonologically and phonetically signal *new*, *given*, and *contrastive* information during an ecologically valid speech task and then compare their productions to those of adults.

Like the previous two experiments, the two pitch accent types that are of interest for analysis are H* and L+H*. Previous research assigns these two pitch accents to different parts of the information structure. Steedman (2007) assigned the H* pitch accent to his notion of *theme*, which he identified as the part of the sentence corresponding to a question or topic that is presupposed by the speaker. He assigned the L+H* pitch accent to the counterpart *rheme*, or the part of the utterance that constitutes the speaker's novel contribution to that question or topic. Pierrehumbert and Hirschberg (1990) also made a distinction between these two types of pitch accents in the Autosegmental-Metrical (AM) framework, and identified the H* monotonal accent as representing presentationally *new* information and the L+H* bitonal accent as representing *contrastive* information. Although the AM framework and the ToBI transcription system originally identified these accents as being discrete and bimodal in their usage, recent research is mixed on their distributional patterns (Ito, Speer, & Beckman, 2004; Watson et al., 2008; Taylor, 2000).

Current research is split on the distribution of H* and L+H* pitch accents as they represent information status. Ito, Speer, and Beckman (2004) showed L+H* is more typically associated with *contrastive* meaning in discourse than the H* pitch accent. More recently, these two pitch accents have been claimed to have overlapping distributions in both perception and production, with H* understood as both presentationally *new* and *contrastive* and L+H* as primarily *contrastive* (Watson et al., 2008). Corpus studies on the phonetics of these two accents also reveal that in production these two pitch accents are difficult to distinguish (Calhoun, 2006; Hedberg & Sosa, 2007). Additionally, while automatic F_0 analyses show overlap of the two contours in their usage, the same F_0-only models lack a way to phonologically label these pitch accents (Taylor, 2000).

Given that the distributions and productions of these two pitch accents are ambiguous in adult speech, this study investigates how toddlers represent information structure categories in their own productions in a natural speech setting. In this type of context, will productions from adult controls confirm the results from recent work and show overlapping distribution patterns? Will toddlers show these same overlapping distributions of H* and L+H* as they relate to *new* and *contrastive* information?

Beyond this first possibility which reflects adult pitch accent distributions, two alternative possibilities are (1) that toddlers may either exhibit a more bimodal distribution for these two pitch accents with respect to meaning, or (2) that they may use a completely novel strategy (i.e., unique pitch accents) that differs from adults.

Previous work shows that young children often do not deaccent *given* information in the same way that adults do, and that they may exhibit distinct prominence patterns (Wieman, 1976; Behrens & Gut, 2005; Chen & Fikkert, 2007; Wonnacott & Watson, 2008). Specifically, it has been claimed that children may not deaccent at the same rate as adults, and that more of their words are assigned an accent. Thus, in addition to exploring how toddlers signal *new*, *given* and *contrastive* information distinctions in their own speech, a secondary consideration is to investigate the rate at which they deaccent *given* information.

A further objective of this study is to identify the primary acoustic-phonetic components that manifest in the different types of pitch accents employed. The primary acoustic correlates of intonation are fundamental frequency (~pitch), duration, amplitude/intensity (loudness), and segmental quality (Lehiste, 1970; Bolinger, 1989; Shattuck-Hufnagel & Turk, 1996). Past research on the relationship between information structure and prosody shows that focused elements (in relation to *given* ones) are more acoustically prominent. Debate has concerned which of the acoustic correlates of prosody is the most important for listener perception of focus, with different researchers suggesting pitch (Lieberman, 1960), duration (Beckman, 1986; Fry, 1955), intensity (Beckman, 1986), or vowel quality (Sluijter & van Heuven, 1996). Each of these studies found different acoustic correlates to be at the heart of focus and prominence. These differing findings led to work by Breen, Fedorenko, Wagner, and Gibson (2010) that

took a more direct approach to uncovering which components of the acoustic signal best predict information status.

Breen et al. (2010) identified several limitations of previous work including the questionable ecological validity of methodologies used (i.e., elicitation tactics, perceptual task judgments), the lack of acoustic evidence in addition to phonological labels (ToBI labels), the breadth of acoustic features studied (one feature vs. many), and the failure to take speaker variability into account. By contrast, their study tested which acoustic features out of a large set were the best predictors to focus type membership based on a more natural speech task where sentences were elicited during a question-answer setting. They used linear discriminant analysis (LDA) to show that duration plus silence (after the word), maximum F_0, mean F_0, and maximum intensity were the best predictors of information structure categories in adult speech. Their study focused on predicting focus type, location, and breadth (i.e., contrastiveness, sentence position, narrow vs. wide) in SVO sentences and found that these four correlates were consistently identified as the best predictors for these categories. To establish a starting point for the pitch accent analysis used in this study, I focus on the four acoustic correlates identified by Breen et al. (2010).

Overall, the aims of this study are three-fold. The first aim analyzes how adult controls intonationally and phonologically make information status distinctions during a spontaneous speech task (*new, given,* and *contrastive*), establishing a baseline of comparison for the toddler productions. Utterances will be analyzed by using the ToBI

transcription system (Beckman & Ayers, 1997). The second aim is to examine how these same information structure categories are signaled in toddlers' spontaneous speech. The motivating question is if toddlers show a one-to-one correspondence between pitch accent type and information status, or if they will they exhibit overlapping distributions like adults. The final aim is to analyze the acoustic features of fundamental frequency, duration, and amplitude for both the child and adult pitch accent productions, and investigate which correlates best predict pitch accent category membership.

4.2 Methods

4.2.1 Participants

Data were analyzed for 12 American English-acquiring toddlers (6 female). The age of participants ranged from 902 to 973 days, with a mean age of 933 days (2;6.19). Data from five additional child participants were discarded due to inattentiveness (3) or equipment malfunction (2). Data were also analyzed for 6 American English-speaking adults (3 female). The age of participants ranged from 20 to 51 years, with a mean age of 29 years. All participants were from Providence, RI, and immediately surrounding areas. Child participants received a t-shirt, book, or age-appropriate toy for their participation. Adult participants received either course credit or a small monetary gift for their participation.

Table 4.1. *Full list of stimuli used in Experiment 3.*

Stimuli	
bear	spoon
sun	balloon
moon	ball
lamb	tree
bunny	shoes
monkey	banana
bee	flower
mouse	cheese

4.2.2 Stimuli

All objects/words used in the design are commonly known at age 2 years and 6 months, and were selected from the MacArthur Communicative Development Index (MCDI), Words and Sentences, which evaluates receptive and productive abilities for children aged 18 to 30 months old (Fenson et al., 2007). Toddler comprehension and production of the words used in the experiment were verified with the caregiver before the experiment began. Additionally, an effort was made to select stimulus items that had names that consisted of primarily sonorant segments. This was done in order to ensure a more reliable and smooth F_0 track with fewer perturbations and tracking errors for pitch analysis (e.g., *ball*, *moon, balloon*). Table 4.1 lists the stimulus items used in the study.

4.2.3 Design

Speech was elicited during a spontaneous production task. The task was designed as an interactive game in order to collect speech that was as close to natural as possible.

Target stimuli (Table 4.1) were elicited during the task for three Information Structure categories. The three categories of interest were *new, given*, and *contrastive.* A target stimulus item was considered *new* if it had not been previously mentioned in the discourse, *given* if it had been recently mentioned in the discourse, or *contrastive* if it was in contrast to a previous referent in the context (or *corrective* if uttered in order to correct the experimenter). The discourse context is defined as the total interaction time between the participant and the experimenter over the entire length of the experimental session (the total duration of an individual session was between 10 and 20 minutes).

4.2.4 Procedure

During an interactive game, a set of target nouns were elicited in one of three ways: 1) *new,* uttered for the first time by the participant; 2) *given*, previously and recently uttered at least once by the participant; or 3) *contrastive*, uttered in direct opposition to a previously mentioned referent (also included here were *corrective* utterances, which were uttered to correct a previously mentioned referent). The game was designed as a matching game with two players, the participant and the experimenter. The materials used in the game included a large spaceship (3 by 1.5 feet), 16 laminated pieces with velcro on the back that each depicted one of the 16 target referents (~3 by 3 inches; e.g., a bear, a sheep, etc.), and a smaller version of the spaceship (8 by 5 inches) that showed the target referents already on the spaceship in designated locations (referred to as 'the map' or 'the seating chart'). See Figure 4.1 for one version of the spaceship 'map' with all sixteen characters and their objects on it.

The participant was first asked to select one of two small spaceship maps (version a and version b, each with a different configuration). Participants were recorded using a Blue Yeti freestanding microphone set to omnidirectional. This setting was selected so that all speech in the room could be recorded, even if the participant moved during the task. The microphone was suspended using an extendable arm and was positioned about 1.5 feet from the participant. All recordings were sampled at a rate of 44000 Hz in Bliss.[2] Recordings were later analyzed in Praat (Boersma & Weenink, 2014). See Figure 4.2 for an image of the experimental set up with the microphone.

Figure 4.1. *Image of spaceship 'map' with all target items in their locations.*

[2] Bliss is an in-house recording and acoustic analysis program designed by John Mertus at Brown University.

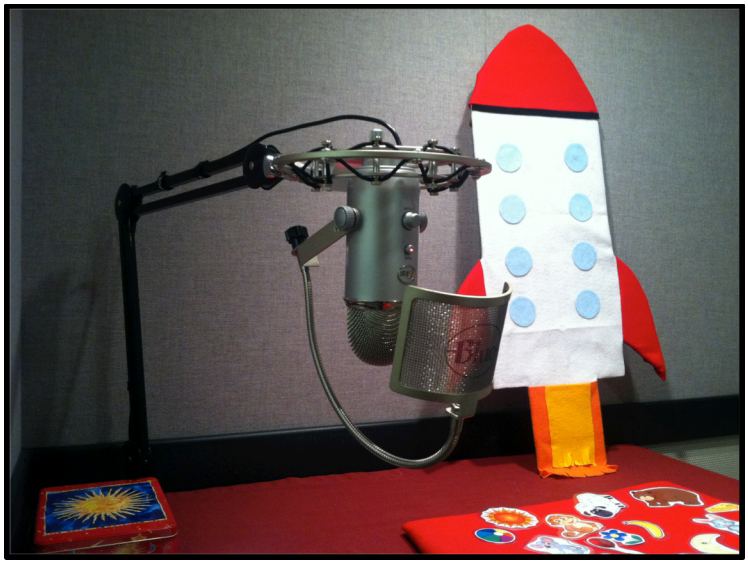


Figure 4.2. *Experimental set up for Experiment 3 showing microphone, spaceship, and target items.*

The directions given to the participants stated that their job was to help the other player (the experimenter) put the characters and their objects on the spaceship so that they could 'blast off to space'. The participant was told that the spaceship could not leave until all of the characters were on board with their belongings, and that each item needed to be labeled before it could be put on the ship. Participants were told to put the target items on the big spaceship in the same way as they appeared on the spaceship map that they had selected. After the participant helped place all of the items on the big spaceship, they would help in the 'cleanup game' and again label all of the items as they were taken off the spaceship by the experimenter.

In order to encourage saying the labels of the objects (either *new* or *given*), the experimenter asked questions about who might go where on the spaceship with an effort made to ask broad focus type questions. To elicit *contrastive* or *corrective* utterances the

experimenter would either make mistakes when placing items on the big spaceship (eliciting a *corrective* response from the participant), or ask questions about which character belonged with what object (eliciting a *contrastive* response). For the purposes of this experiment and for the analysis, *contrastive* and *corrective* utterances were collapsed into one information structure category. Two sample interactions between the experimenter and a (child) participant are listed in (4.1) and (4.2). Sample dialogue (4.1) demonstrates the participant uttering three *new* target words, and sample dialogue (4.2) demonstrates *contrastive* target word usage.

(4.1) Participant: *That's a [moon]*$_{\text{New}}$.

Experimenter: *Hum, what's next?*

Participant*: There's a [bear]*$_{\text{New}}$ *and [sheep]*$_{\text{New}}$.

(4.2) Experimenter: *Oh, was this the bee?*

Participant*: No, that's [sun]*$_{\text{Contrastive}}$.

4.2.4. Phonological and Phonetic Analyses

This study consisted of both phonological and phonetic analyses of the productions elicited during the task. Specifically, the target word productions were analyzed for each of the information structure categories labeled (*new, given, contrastive*). The target word was labeled following the ToBI transcription system guidelines for American English (Beckman & Ayers, 1997). For the phonological analysis, the pitch accent frequency and distribution for each information structure category were calculated.

For the phonetic analyses, a set of four acoustic variables were measured to test the acoustic underpinnings of the phonological categories. Using Praat (Boersma & Weenink, 2014), the values for each of these measurements were extracted from the stressed syllable portion of each target word analyzed. Table 4.2 lists all of the acoustic measurements collected for each target production. Each of the acoustic measurements was taken for the entire utterance in which the target word was produced, the target word itself, the stressed syllable of the target word, and the stressed vowel of the target word. All of the acoustic correlates were used for the discriminant function analysis (reported in §4.3.3), but only four of these correlates were used for the primary acoustic analysis (reported in §4.3.2). These four acoustic measurements were *duration (standardized), maximum F_0, mean F_0,* and *maximum*

Table 4.2. *List of acoustic measurements for Experiment 3. All measurements were taken at the utterance, target word, target stressed syllable, and target stressed vowel levels.*

Acoustic Label	Definition
Duration (raw)	Time from onset to offset of target
Duration (standardized)	Z-score of raw duration by participant
F_0 min	Minimum F_0 (Hz) during target
F_0 max	Maximum F_0 (Hz) during target
F_0 range	Difference between F_0 max and F_0 min (Hz)
F_0 mean	Average F_0 (Hz) over target segment
Intensity min	Minimum amplitude (dB) during target
Intensity max	Maximum amplitude (dB) during target
Intensity range	Difference between int. max and int. min (dB)
Intensity mean	Average amplitude (dB) over target segment

intensity.. The selection of these four measurements was based on research demonstrating that these are the four correlates that best predict information structure categories in adult speech (Breen et al., 2010). This set of measurements allowed for the comparison between the types of pitch accents used by children and adults.

In order to collect these measurements, all utterances containing a target word were segmented from the spontaneous speech task recording. A Praat text-grid was created for each utterance and contained seven tiers (Boersma & Weenink, 2014). The first four tiers were the Vowel Tier, Syllable Tier, Word Tier, and the Utterance Tier. Respectively, each of these tiers marked the beginning and end of the target noun vowels, target noun syllables, all words in the utterance, and the entire utterance itself.

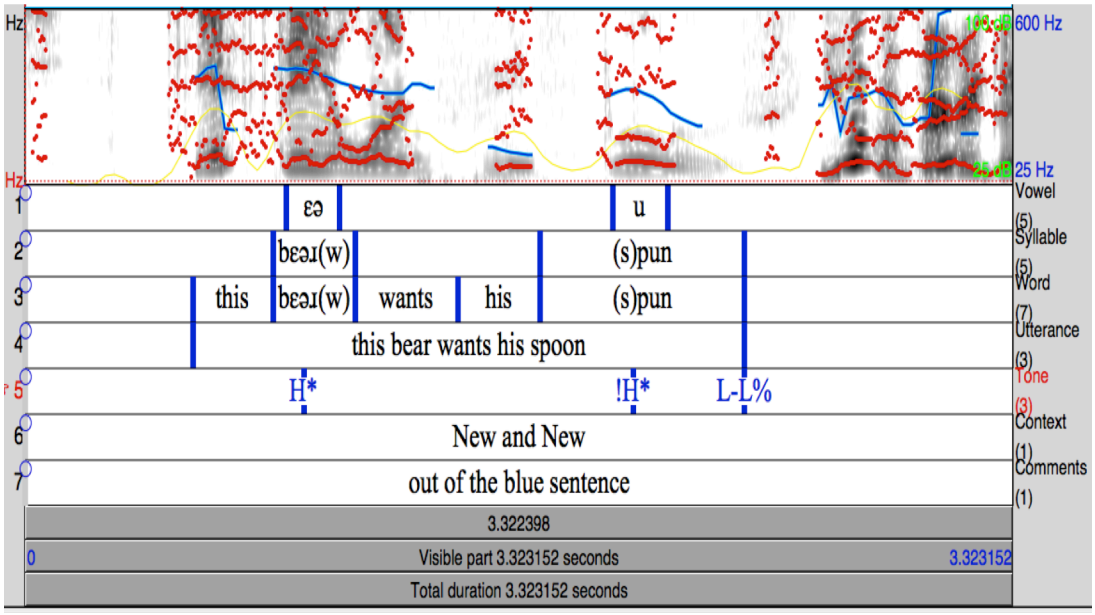


Figure 4.3. *Example Praat textgrid showing spectrogram and textgrid for the target utterance "this bear wants his spoon", target nouns underlined.*

The fifth tier was the Tone Tier and marked the pitch accents and boundary tones for the utterance. The sixth tier was the Context Tier and marked the information status of the target noun production in the utterance (*new*, *given* or *contrastive*). This portion was completed independently initially to avoid biasing the intonation analysis. Finally, the final Comments Tier marked the focus type of the utterance (*narrow* or *broad/out of the blue*) as well as any transcriber notes. Figure 4.3 shows an example spectrogram and Praat text-grid with all seven tiers for the target utterance "*this bear wants his spoon*" (target noun productions are underlined).

4.3 Results

4.3.1 Phonological Analyses

The frequency and distribution patterns of pitch accent types were calculated for the adult and the child participants. Figures 4.4 and 4.5 show the percentage of occurrence for each pitch accent during the task for the three information structure categories of *new, given,* and *contrastive*. Other categories that occurred in the data set were not included for analysis because they were outside the scope of the present study. For example, non-contrastive narrow focus productions were not included in this analysis. For the adults, a total of 232 targets were included in the frequency and distribution analysis. For the children, the analysis was based on 145 target productions.

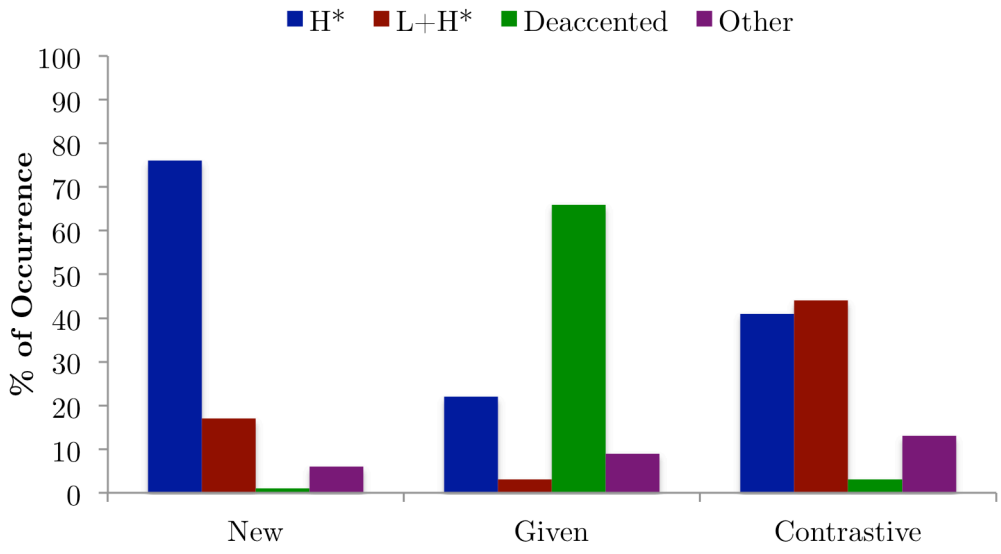


Figure 4.4. *Bar graph showing the adults' percentage of occurrence of each pitch accent type for the three information structure categories.*

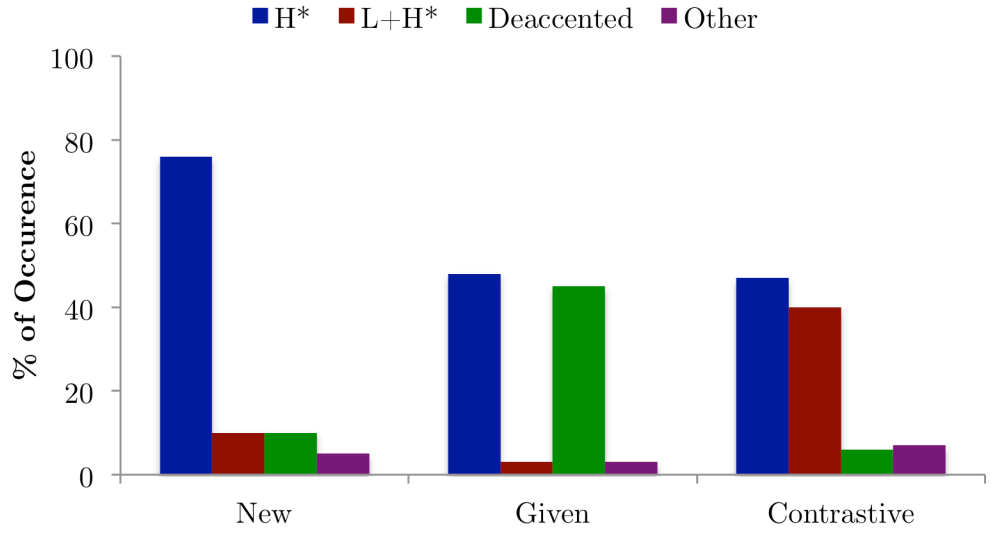


Figure 4.5. *Bar graph showing the children's percentage of occurrence of each pitch accent type for the three information structure categories.*

For the adult data, the H* pitch accent was the most frequent pitch accent used to express *new* information, representing 76% of the data (66 out of 87 total occurrences). The second most frequent type was the L+H* with 22% of the data (14 out of 87 occurrences) (Figure 4.4). For the *contrastive* information, distribution was almost evenly split between the H* and the L+H* pitch accents, which each represented 41% and 44% of the data, respectively (33 for H* and 35 for L+H* out of 80 total occurrences). Finally, the majority of the *given* material was deaccented (66%, or 43 out of 65 occurrences). In addition, 22% of the information marked as *given* was labeled with an H* pitch accent. Several other pitch accents were recorded for each of the information structure categories, but their frequency was never over 10% for any of the categories. The other most frequent types of pitch accents identified were the L* and the L*+H, represented by the 'other' pitch accent category label.

The child data showed distribution patterns similar to the adult data, particularly for *new* and *contrastive* information (see Figure 4.5 for child data). The H* pitch accent was used for 76% of the *new* information (48 out of 63 occurrences). The L+H* pitch accent and deaccentuation each made up 10% of the *new* information as well (each with 6 out of 63 occurrences). For the *contrastive* information, the H* and the L+H* pitch accents were roughly evenly distributed like the adults, representing 47% and 40% respectively (25 and 21 out of 53 occurrences). Unlike the adults, the children show a higher proportion of H* pitch accents for *given* information. For the *given* information, 48% of the targets carried an H* pitch accent and 45% were deaccented (14

and 13 out of 29 occurrences). The 'other' pitch accents types never made up more than 4% of the data for any of the categories. For the children, the only other type of pitch accent labeled was the L*.

4.3.2 Phonetic Analyses

The acoustic correlates of the H* and L+H* pitch accents as well as the deaccented target words were compared across the adult and child participants. As mentioned in §4.2.4, the four acoustic correlates selected for analysis were the *standardized duration*, *maximum F_0*, *mean F_0*, and *maximum intensity*. All values for these analyses were taken during the stressed syllable of the target word. Group results are presented in line graphs for the adults and the children by pitch accent type. The adult data were additionally split by gender. Adult males and females were analyzed separately for the acoustic results since acoustic realizations can vary substantially due to physical size and gender

Table 4.3. *Repeated measures ANOVA results for each acoustic variable, plus effect sizes (η^2).*

Group	Acoustic Measure	F-Value (DF)	*P* Value	*Effect Size (η^2)*
Males	Standardized Duration	$F(2,4) = 15.78$	$p = .013^{*}$	.89
	Mean F_0	$F(2,4) = 7.13$	$p = .048^{*}$	.78
	Maximum F_0	$F(2,4) = 14.09$	$p = .015^{*}$	.88
	Average Intensity	$F(2,4) = 19.51$	$p = .009^{**}$	.91
Females	Standardized Duration	$F(2,4) = 6.97$	$p = .050^{*}$	.78
	Mean F_0	$F(2,4) = 34.86$	$p = .003^{**}$	.95
	Maximum F_0	$F(2,4) = 32.95$	$p = .003^{**}$	.94
	Average Intensity	$F(2,4) = 7.51$	$p = .044^{*}$	.79
Children	Standardized Duration	$F(2,20) = 9.51$	$p = .001^{**}$	.48
	Mean F_0	$F(2,20) = 10.71$	$p = .001^{**}$	.52
	Maximum F_0	$F(2,20) = 6.09$	$p = .009^{**}$	.38
	Average Intensity	$F(2,20) = 24.51$	$p < .001^{**}$	.71

differences. Child data are presented as one group mean since gender does not have a significant effect at this early of an age (particularly on pitch).

First, repeated measure ANOVAs were conducted for each of the acoustic variables with Pitch Type as a within-subjects variable. Pitch Type consisted of three levels (*deaccented, H** and *L+H**). The average acoustic value at each level, for each participant, was entered in to the ANOVA. All of the ANOVAs were significant, showing that the Pitch Type was significantly different across participants. The results from each of the ANOVAs are presented in Table 4.3.

Following the significant repeated measures ANOVAs, pairwise comparisons were conducted to determine which groups differed significantly from each other. The results of paired t-tests using Bonferroni corrections between pitch accent groups are shown in Tables 4.4 through 4.7 for each acoustic measurement analyzed (Bonferroni adjusted alpha levels of .016 per test (.05/3)). As with the multivariate analyses, the average for each acoustic value for each participant was used for analysis. For the children, one participant did not have any deaccented or L+H* tokens, so these values are missing (this is reflected in the degrees of freedom). The results for the *standardized durations* of the deaccented, H*, and L+H* target words for males, females, and children are presented in Table 4.4 and Figure 4.6. For the adults, deaccented and H* targets were significantly shorter than the L+H* targets. The difference in *duration* between deaccented and H* targets did not reach significance for either the male or the female

adults. Children showed significant differences between the deaccented tokens and the two pitch accent types, with deaccented tokens having the shortest average duration.

Male adults did not show any significant differences between pitch types for *mean F_0* and *maximum F_0*. For female adults, *mean F_0* was significantly higher for H* target words than the deaccented targets and *maximum F_0* was higher for L+H* targets than deaccented ones (see Tables 4.5 and 4.6). Children showed a significant difference

Table 4.4. *Paired t-tests for standardized duration with Bonferroni corrected alpha levels.*

Group	Contrast	T-Value	*P* Value
Males	Deaccented vs. H*	$t(2) = 1.31$	$p = .321^{ns}$
	Deaccented vs. L+H*	$t(2) = 4.60$	$p = .044^{ns}$
	H* vs. L+H*	$t(2) = 3.99$	$p = .058^{ns}$
Females	Deaccented vs. H*	$t(4) = 0.70$	$p = .557^{ns}$
	Deaccented vs. L+H*	$t(4) = 2.67$	$p = .116^{ns}$
	H* vs. L+H*	$t(4) = 11.16$	$p = .008^{**}$
Children	Deaccented vs. H*	$t(10) = 4.01$	$p = .002^{**}$
	Deaccented vs. L+H*	$t(10) = 3.85$	$p = .003^{**}$
	H* vs. L+H*	$t(10) = 1.70$	$p = .120^{ns}$

Table 4.5. *Paired t-tests for mean F_0 with Bonferroni corrected alpha levels.*

Group	Contrast	T Value	*P* Value
Males	Deaccented vs. H*	$t(2) = 1.45$	$p = .283^{ns}$
	Deaccented vs. L+H*	$t(2) = 4.08$	$p = .055^{ns}$
	H* vs. L+H*	$t(2) = 2.04$	$p = .179^{ns}$
Females	Deaccented vs. H*	$t(2) = 30.34$	$p = .001^{**}$
	Deaccented vs. L+H*	$t(2) = 6.90$	$p = .020^{ns}$
	H* vs. L+H*	$t(2) = 3.17$	$p = .087^{ns}$
Children	Deaccented vs. H*	$t(10) = 1.88$	$p = .090^{ns}$
	Deaccented vs. L+H*	$t(10) = 3.88$	$p = .003^{**}$
	H* vs. L+H*	$t(10) = 3.30$	$p = .008^{**}$

Table 4.6. *Paired t-tests for maximum F_0 with Bonferroni corrected alpha levels.*

GROUP	CONTRAST	T VALUE	*P* VALUE
Males	Deaccented vs. H*	$t(2) = 0.21$	$p = .856^{ns}$
	Deaccented vs. L+H*	$t(2) = 5.83$	$p = .028^{ns}$
	H* vs. L+H*	$t(2) = 4.23$	$p = .052^{ns}$
Females	Deaccented vs. H*	$t(2) = 5.18$	$p = .035^{ns}$
	Deaccented vs. L+H*	$t(2) = 9.66$	$p = .011^{*}$
	H* vs. L+H*	$t(2) = 3.59$	$p = .070^{ns}$
Children	Deaccented vs. H*	$t(10) = 0.38$	$p = .710^{ns}$
	Deaccented vs. L+H*	$t(10) = 3.17$	$p = .010^{*}$
	H* vs. L+H*	$t(10) = 3.37$	$p = .007^{**}$

Table 4.7. *Paired t-tests for maximum intensity with Bonferroni corrected alpha levels.*

GROUP	CONTRAST	T VALUE	*P* VALUE
Males	Deaccented vs. H*	$t(2) = 11.50$	$p = .007^{**}$
	Deaccented vs. L+H*	$t(2) = 4.96$	$p = .038^{ns}$
	H* vs. L+H*	$t(2) = 1.83$	$p = .209^{ns}$
Females	Deaccented vs. H*	$t(2) = 8.23$	$p = .052^{ns}$
	Deaccented vs. L+H*	$t(2) = 2.85$	$p = .014^{*}$
	H* vs. L+H*	$t(2) = 0.85$	$p = .485^{ns}$
Children	Deaccented vs. H*	$t(10) = 2.63$	$p = .025^{ns}$
	Deaccented vs. L+H*	$t(10) = 6.24$	$p < .001^{**}$
	H* vs. L+H*	$t(10) = 5.39$	$p < .001^{**}$

between deaccented and H* targets and H* and L+H* target words for both *mean* and *maximum F_0*. See Figures 4.7 and 4.8 for line graphs of *mean* and *maximum F_0*.

Male adults showed higher *maximum intensity* values for H* targets in comparison to deaccented ones, but no difference between H* and L+H* or deaccented and L+H* target words (Table 4.7). Adult females only showed higher *maximum intensity* values for L+H* targets in comparison to deaccented ones, with no significant differences between deaccented and H* or H* and L+H* targets. Children showed

statistically significant differences between deaccented and L+H* as well as between H* and L+H* targets. They did not differ significantly between deaccented and H* tokens. Figure 4.9 shows the *maximum intensity* mean values of the different pitch accents for each group.

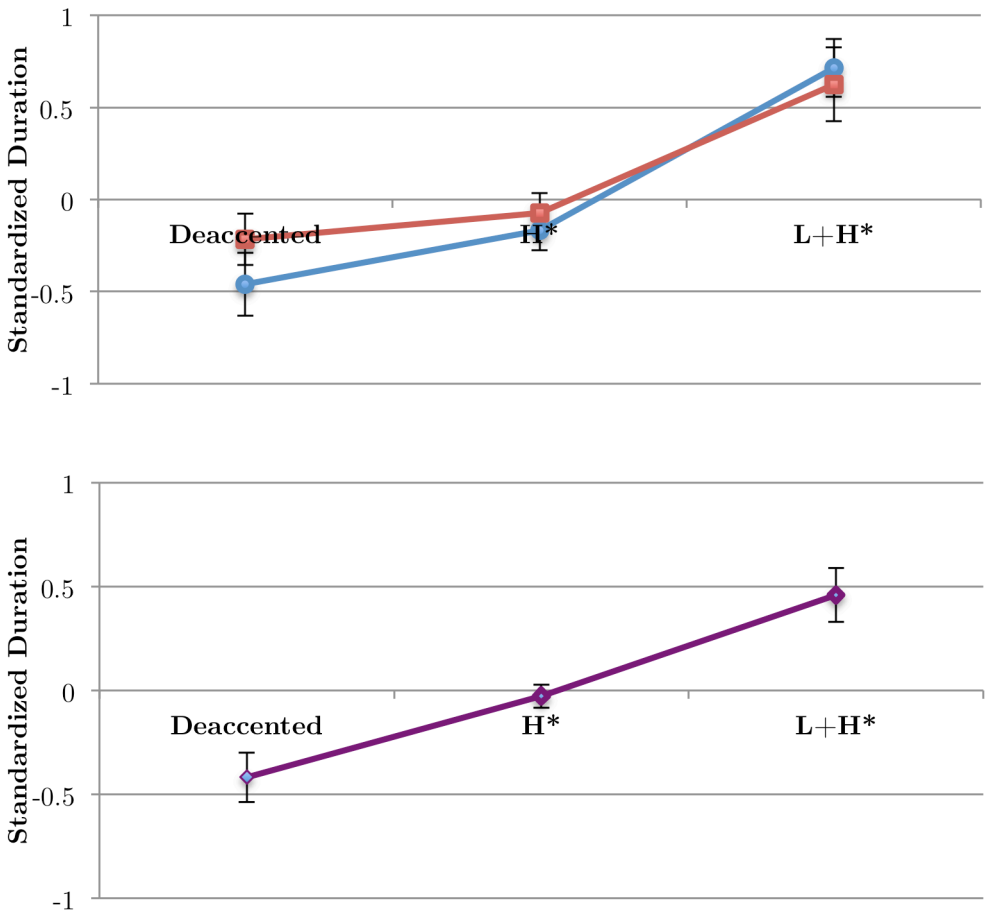


Figure 4.6. *Line graphs depicting standardized duration scores for deaccented, H*, and L+H* productions. Adults are on top and the children are on the bottom. In the adult graph, the red line shows the means for the men, and the blue line for the women. Error bars show standard error.*

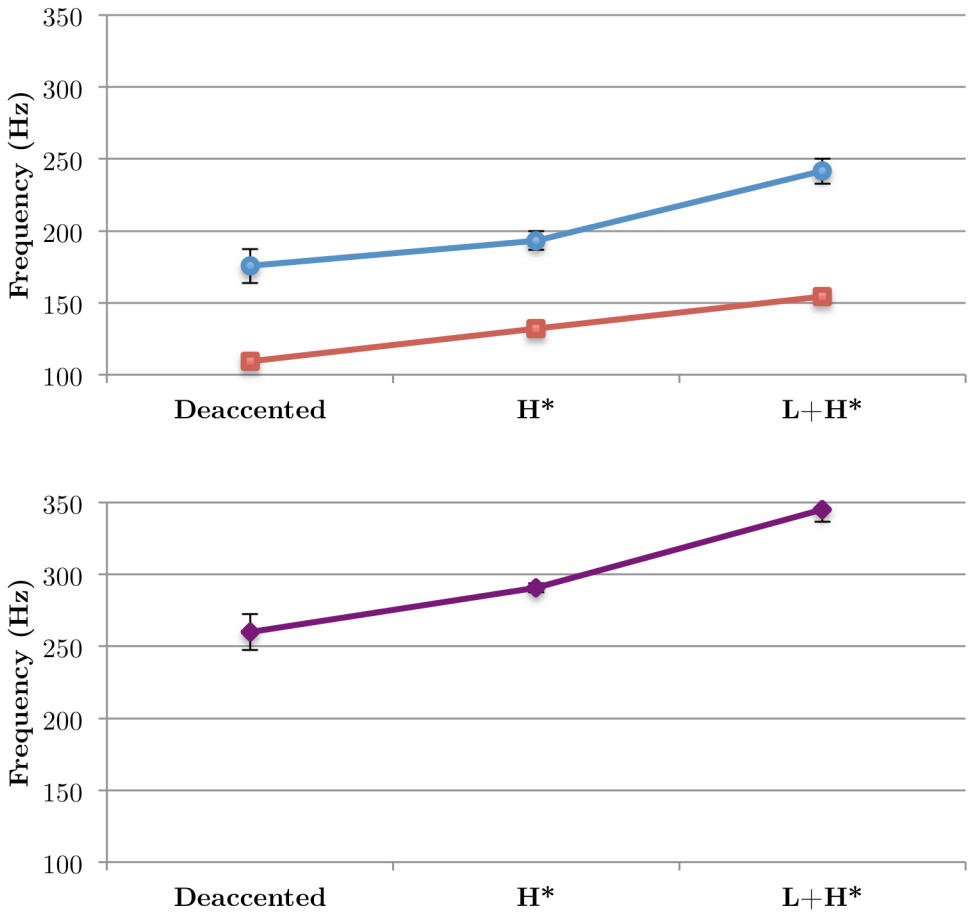


Figure 4.7. *Line graphs depicting mean F_0 for deaccented, H*, and L+H* productions. Adult means are on top and the children are on the bottom. In the adult graph, the red line shows the means for the men, and the blue line for the women. Error bars show standard error.*

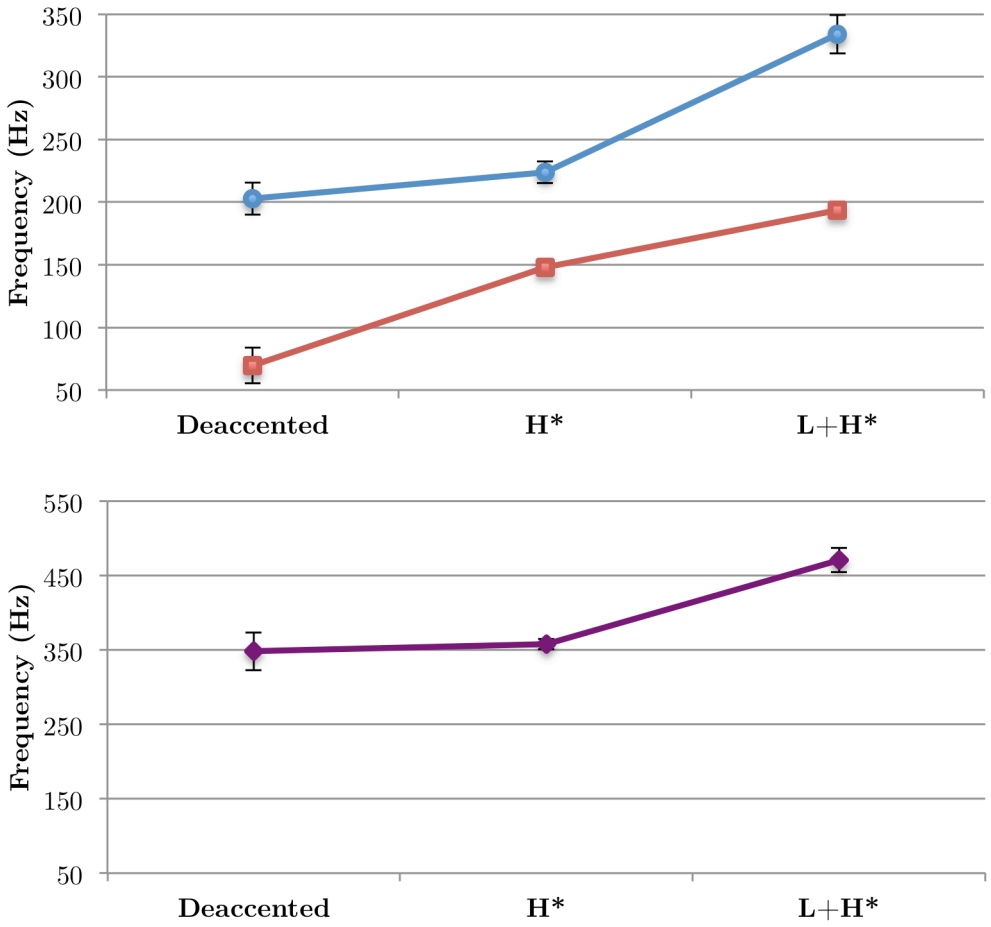


Figure 4.8. *Line graphs depicting maximum F_0 for deaccented, H*, and L+H* productions. Adult means are on top and the children are on the bottom. In the adult graph, the red line shows the means for the men, and the blue line for the women. Error bars show standard error.*

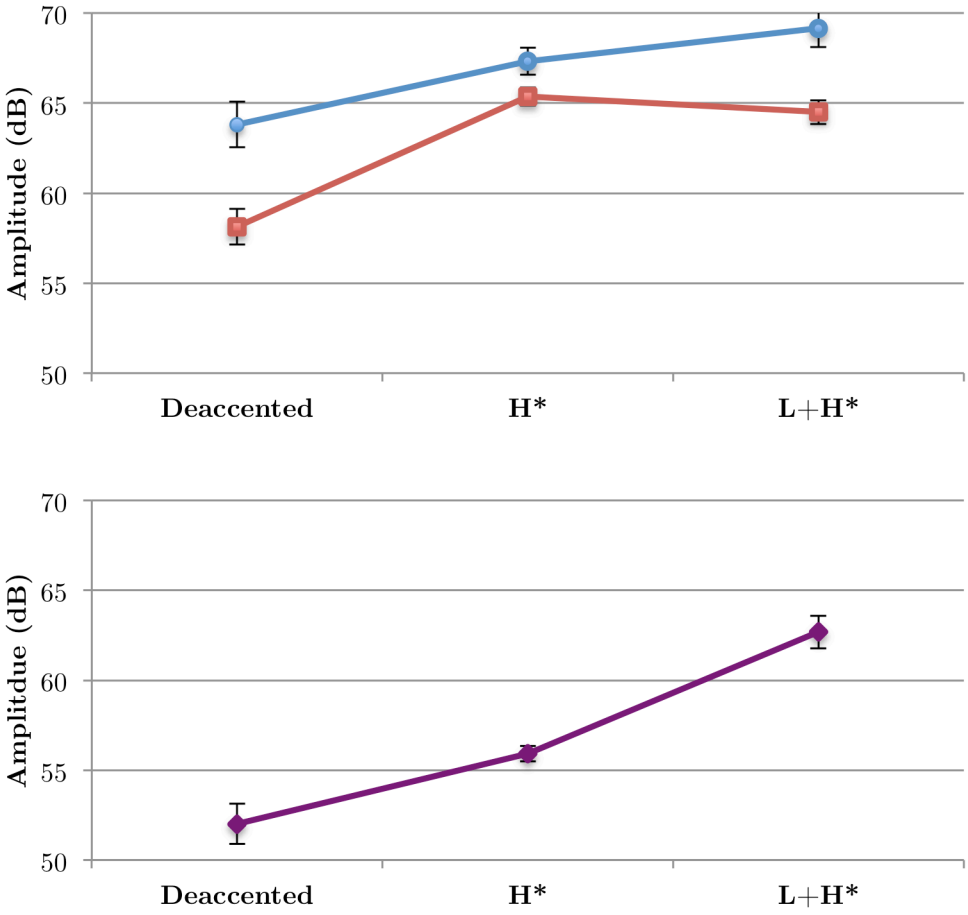


Figure 4.9. *Line graphs depicting maximum intensity values for deaccented, H*, and L+H* productions. Adult means are on top and the children are on the bottom. In the adult graph, the red line shows the means for the men, and the blue line for the women. Error bars show standard error.*

4.3.3 Discriminant Function Analyses

For this experiment, all participants were recorded during a spontaneous speech task where they uttered a set of target nouns with different information statuses. Target nouns were either *new, given*, or *contrastive* in the discourse context, and the pitch accents on these words were labeled according to the ToBI transcription system (Beckman & Ayers, 1997). The three pitch types that consisted of over 90% of the data for the three information categories were the deaccented, H*, and L+H* pitch accents. The *pitch accent groups* were analyzed for their primary acoustic correlates: fundamental frequency, duration, and intensity. One of the aims of the experiment was to investigate how well the acoustic correlates, the *predictors*, would accurately predict the pitch accent types used to label these different types of prosodic productions. A second aim was to discover which of these nine variables best predicted group membership, and whether they were similar to the acoustic measurements that have been identified in previous studies.

The nine predictor variables collected for each of the target pitch accents were: *standardized duration, mean F_0, maximum F_0, minimum F_0, F_0 range, mean intensity, maximum intensity, minimum intensity,* and *intensity range.* These measurements were taken over the course of the target stressed vowel, target stressed syllable, target word, and the target utterance. For this analysis, the values used are from the stressed syllable of the target word. Following the objective of this study, a stepwise discriminant function analysis (DFA) was performed using the previously mentioned nine acoustic

correlates of intonation as predictor variables of group placement (deaccented, H*, and L+H*). A stepwise analysis was chosen over a direct analysis to better identify which acoustic measurements were the most essential in determining group membership, a primary goal of the study. All analyses were conducted using SPSS (IBM Corp., 2013).

First, a check was performed to make sure all assumptions for running a DFA were met, particularly to check for multivariate normality and outliers. For the adults, out of the 264 cases there were no cases excluded and none of them were considered outliers. For the children, out of the 469 cases there were no cases excluded and none of them were outliers. Even though there were unequal group sample sizes, the overall large sample sizes of 264 and 469 ensured robustness of multivariate normality, and the DFA could be performed with reliable interpretations.

For the adult data, all scores were z-scored in order to reduce gender differences between the males and the females as well as to reduce speaker variability. The z-scored variables were then used in the discriminant function analysis. Additionally, the F_0 *range* predictor variable was positively skewed during a check for normality. This variable was transformed by taking the square root of each value. It was the *squared* F_0 *range* values that were used for the DFA. All of the child data passed checks of normality and no transformations were needed. The one variable that was standardized beforehand was duration in order to control for variation in vowel length and rate of speech. Finally, both sets of data failed Box's M test of equality of covariance matrices (adults: $M = 45.25$, $p <. 001$; children: $M = 113.46$, $p = .002$), which tests the

assumption that the vector of the dependent variables follow a normal distribution, and the variance-covariance matrices are equal across the cells formed by the between-subjects effects (SPSS, IBM Corp., 2013). Since population covariance matrices were determined to be unequal, separate- instead of within-group covariance matrices were substituted during classification.

Discriminant function analyses were performed separately for the adult data and child data sets. From the adult data, two discriminant functions (DF) were extracted, with a combined $\chi^2(8) = 106.676$, $\Lambda = .663$, $p < .001$. After the removal of the first DF, there was still a significant association between the groups and the predictors, yielding $\chi^2(3) = 30.151$, $\Lambda = .890$, $p < .001$. The amount of group-discriminating variance accounted for by the two DFs was 74% and 26%, respectively. The role of each function was as follows: the first separated deaccented from H* and L+H*, and the second separated H* from L+H*. The stepwise discriminant analysis yielded a final model with four predictor variables: *standardized duration, mean F_0, maximum F_0,* and *maximum intensity*. The loading matrix revealed which of the nine predictors most heavily loaded on each function.

For the first DF the strongest loading variable with value over .8 was *maximum F_0*, but overall, all of the other variables included in the final model contributed quite heavily with *mean F_0*, *standardized duration,* and *maximum intensity* having loadings of .749, .615, and .600, respectively. The only variable loading on the second DF was *maximum intensity* with a loading of .625. Variables that were not included in the final

stepwise model were not used in the analysis. Standardized canonical discriminant function coefficients are given in Table 4.8 for each predictor used in the final stepwise model.

The classification results for the adult model revealed that original group membership was correctly classified in a majority of the cases (63.5%). The best classification results were for the H* pitch accent, with 130 of the 149 cases (87.2%) accurately classified. The deaccented and L+H* pitch accents were not as well classified,

Table 4.8. *Standardized canonical discriminant function coefficients for each predictor variable that was included in the final adult model of the stepwise discriminant function analysis.*

	Discriminant Functions	
Predictor Variable	Function 1	Function 2
Standardized Duration	.381	-.263
Mean F_0	.141	1.002
Maximum F_0	.502	-1.130
Maximum Intensity	.407	.582

Table 4.9. *Classification results for adult final stepwise model.*

		Predicted Group Membership			
Original Group	Pitch Accent	deaccented	H*	L+H*	Total
Count	Deaccented	14	35	6	55
	H*	10	130	9	149
	L+H*	3	37	30	70
	Ungrouped cases	9	36	7	52
%	Deaccented	25.5	63.6	10.9	100
	H*	6.7	87.2	6.0	100
	L+H*	4.3	52.9	42.9	100
	Ungrouped cases	17.3	69.2	13.5	100

with only 25.5% of deaccented and 42.9% of L+H* accurately classified. There were a number of misses and false alarms for these two types, but overall the model performed with over 60% accuracy. See Table 4.9 for a full list of the classification results.

Prediction was performed by obtaining values for each case along the first and second discriminant function dimensions. Figure 4.10 shows a scatterplot of the three pitch accents as well as where each case is placed according to the two canonical discriminant functions. As visible from the plot, there is considerable overlap in the categories, especially between the deaccented and H* pitch categories.

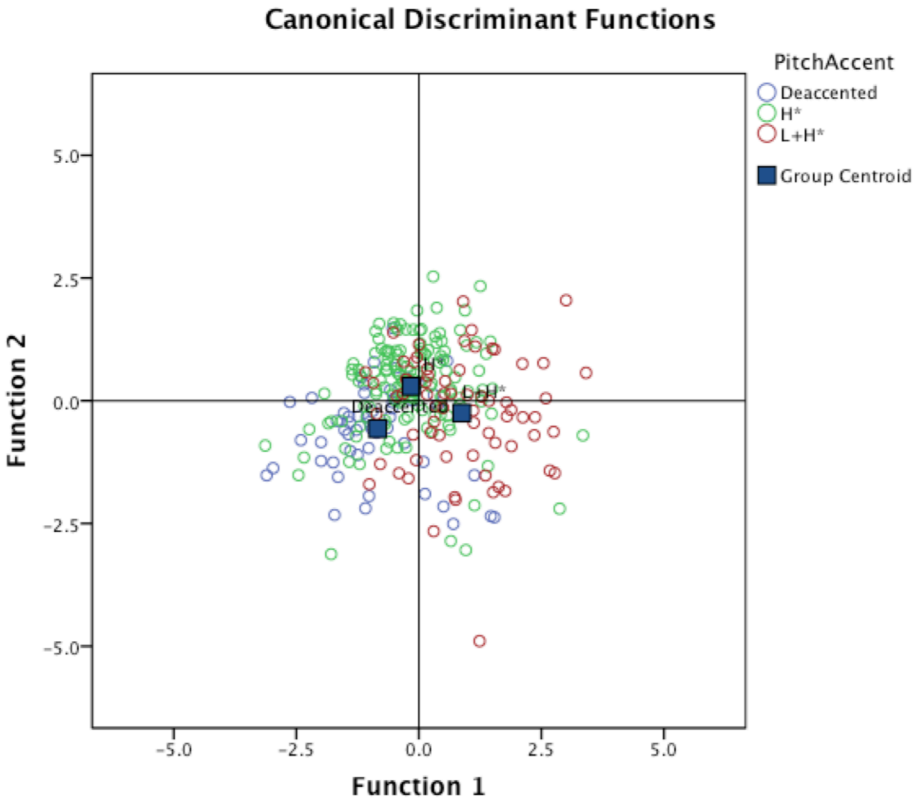


Figure 4.10. *Scatterplot of adult data along canonical discriminant function scores. Group centroids are marked with the large blue squares.*

For the child data, two discriminant functions (DF) were extracted, with a combined $\chi^2(12) = 112.340$, $\Lambda = .733$, $p < .001$. After the removal of the first DF, there was still a significant association between the groups and the predictors, yielding $\chi^2(5) = 20.715$, $\Lambda = .947$, $p = .001$. The amount of group-discriminating variance accounted for by the two DFs was 83% and 17%, respectively. The role of each function was as follows: the first separated deaccented from H* and L+H*, and the second separated H* from L+H*.

Table 4.10. *Standardized canonical discriminant function coefficients for each predictor variable that was included in the final child model of the stepwise discriminant function analysis.*

	Discriminant Functions	
Predictor Variable	Function 1	Function 2
Standardized Duration	.457	-.660
Mean F_0	.413	-.560
Maximum F_0	.230	.876
Minimum Intensity	-.363	.473
Mean Intensity	1.106	-2.531
Maximum Intensity	-.514	2.457

Table 4.11. *Classification results for child final stepwise model.*

		Predicted Group Membership			
Original Group	Pitch Accent	deaccented	H*	L+H*	Total
Count	Deaccented	3	44	1	48
	H*	5	279	11	295
	L+H*	0	44	23	67
	Ungrouped cases	0	51	8	59
%	Deaccented	6.3	91.7	2.1	100
	H*	1.7	94.6	3.7	100
	L+H*	0	65.7	34.3	100
	Ungrouped cases	0	86.4	13.6	100

The stepwise discriminant analysis yielded a final model with six predictor variables: *standardized duration, mean F_0, maximum F_0, minimum intensity, mean intensity,* and *maximum intensity*. The loading matrix revealed which of the nine predictors most heavily loaded on each function. For the first DF the strongest loading variables with values over .6 were *mean F_0, maximum F_0*, and *maximum intensity* (.713, .674, and .673, respectively). *Mean intensity* and *standardized duration* also had loadings on the first DF of .592 and .478, respectively. The variable with the largest absolute correlation with the second DF was *minimum intensity* with a loading of .230, but *mean F_0* also had a loading of .578. Variables that were not included in the final stepwise model were not used in the analysis (i.e., *intensity range* and *minimum F_0*). Standardized canonical discriminant function coefficients are given in Table 4.10 for each predictor used in the final stepwise model.

The classification results for the child model revealed that original group membership was accurately predicted a majority of the time (74.4%). The best classification results were for the H* pitch accent, with 130 of the 149 cases (94.6%) accurately classified. The deaccented and L+H* pitch accents were not as well classified with only 6.3% of deaccented and 34.3% of L+H* accurately classified. There were a number of misses and false alarms for these two types, but still the overall the model performed with over 70% accuracy. See Table 4.11 for a full list of the classification results. Figure 4.11 shows a scatterplot of the three pitch accents as well as where each case is placed according to the two canonical discriminant functions. Similar to the adult

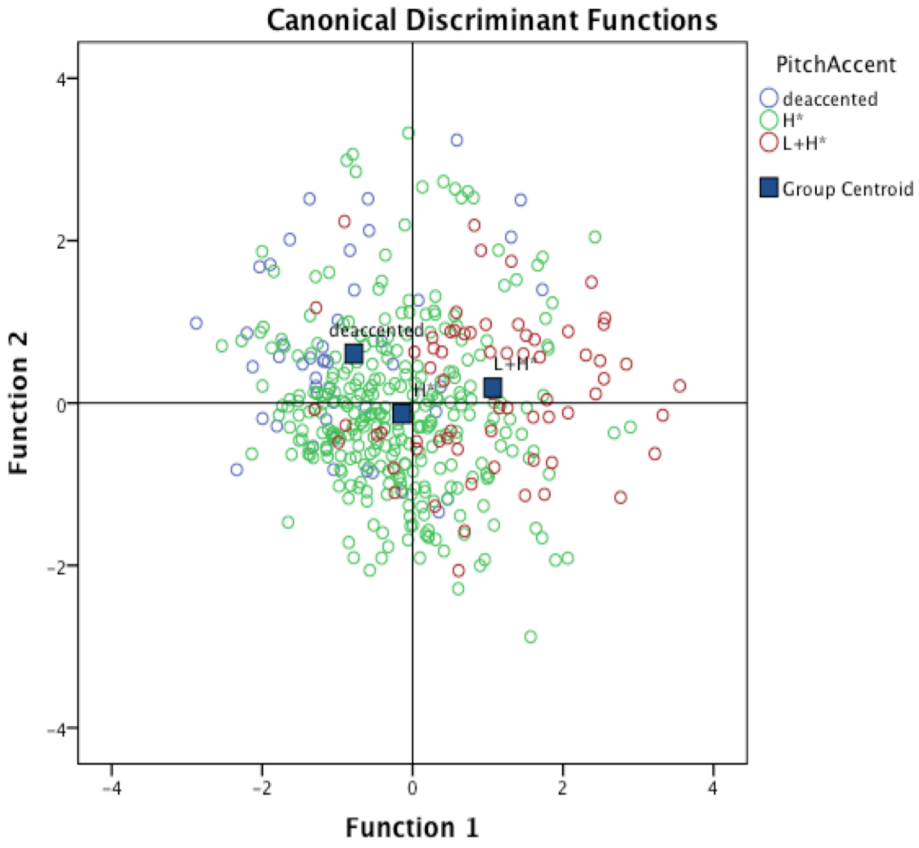


Figure 4.11. *Scatterplot of child data along canonical discriminant function scores. Group centroids are marked with the large blue squares.*

data and as visible from the plot, there is considerable overlap in the categories. For the child data, there is also more dispersion of values across the two discriminant functions.

Cross-validations were run on both models to check the stability of the classification procedure. The cross-validation randomly generated a sub-sample upon which follow-up DFAs were run to compare against the original DFAs described above. The sub-sample consisted of about half of the cases from the original run. There was a high degree in consistency, with the classification rates staying fairly high for group membership. The sub-sample produced accurate classification results above 60% for the adults and above 70% for the children. Thus, the cross-validation DFAs proved the

consistency and stability of the data. See Appendix B for the complete model outputs of the discriminant function analyses.

4.4 Discussion

This experiment explored how adults and young toddlers intonationally represent information structure categories. The information structure categories of interest for this study were *new, given,* and *contrastive*. During a naturalistic speech task, adults and children produced a set of target nouns that were labeled for both their information status and their pitch accent type according to the ToBI transcription system guidelines (Beckman & Ayers, 1997). There were three primary aims of this study, with the overall objective to phonologically and acoustic-phonetically investigate pitch accent productions as they relate to information status.

The first aim was to establish a comparative baseline for comparing child and adult data and analyze how adults phonologically represent the three information structure categories analyzed (i.e., *new*, *given*, and *contrastive*). Distribution and frequency analyses showed that the spontaneous speech produced by adults in this study support the conclusions of recent research by Watson and colleagues (2008). Adults produced all three information structure categories during the task and ToBI labels were applied to each production. Results demonstrate that adults deaccented *given* information in a majority of the target productions (66% of the *given* information was deaccented). In addition, they primarily used the H* pitch accent for *new* information,

and both the H* and L+H* pitch accents in nearly equal percentages to express *contrastive* information. These results replicate earlier findings where H* can be used for both *new* and *contrastive* information, and where L+H* is primarily used for *contrastive* information. Along with replicating earlier findings, this part of the study also produced the same type of data (i.e., from the same methodology and natural setting) for children and adults, and thereby allowed a clear comparison between the two.

Second, we investigated how two and half year old toddlers mark information status through the use of prosody in comparison to the adult population. Overall, the toddler productions exhibited intonational patterns akin to those of the adults. For *new* productions, the children produced nearly exclusively the H* pitch accent. For the *contrastive* productions, they produced H* and L+H* in nearly equal distributions. The point of divergence for the children from the adults was for the *given* cases. For the *given* target words, children produced roughly half as deaccented, like the adults, but the other half with an H* pitch accent.

There are two possible explanations for this pattern. First, as noted by other researchers, early child speech generally contains more stressed words. It has been reported that early two-word combinations often consist of two stressed elements (Behrens & Gut, 2005; Chen & Fikkert, 2007), and that children may lack the ability to deaccent due to the more complex articulatory control necessary in a two-word combination (Bolinger, 1982; Gussenhoven, 2002).

Alternatively, toddlers may be mimicking the patterns that they hear in their input. Bortfeld and Morgan (2010) showed that adults alternate stress patterns of repeated mentions of words when addressing their young children. The first, *new*, mention is pronounced with characteristic long duration and high fundamental frequency. The second, *given*, mention is pronounced with characteristic shorter duration and lower fundamental frequency. However, the third mention is pronounced with longer duration and higher fundamental frequency than the second. The fourth mention resembles the second, the fifth resembles the third, and so forth. Thus, overall, there is less acoustic reduction in child-directed speech than in adult-directed speech (i.e., less deaccentuation overall in child-directed speech). This is different from what is found in adult-directed speech, which reveals that attenuation of material typically continues after the first mention (Halliday, 1967; Chafe, 1976; Prince, 1981).

Bortfeld and Morgan (2010) proposed that this type of stress pattern is useful for calling attention to certain elements in a context, which aids in early word recognition and learning. If this pattern exists in the input for the toddlers in this study, it is possible that the pattern of both accented and deaccented productions is a reflection of the pattern that is most frequently heard by the toddler. Since the toddler participants used deaccentuation at a similar rate as they used accentuation for the *given* cases, it is unlikely that this is due to an articulatory inability. Rather, this appears to be a demonstration of young children's sensitivities to input statistics. Thus, the phonological analyses provide evidence that both adults and children mark *new, given*, and

contrastive information in a consistent manner. Importantly, the analyses validate the use of the H* and L+H* pitch accents, in addition to deaccentuation, to express these information structure categories.

There is still no completely clear answer in how the speech of adults towards infants reflects the intonationally constructed information categories. In certain respects, *given* and *new* information is presented to infants in a similar fashion as that to adults: by stressing the first occurrence and then with attenuation of the second repetition. Evidence from Bortfeld and Morgan (2010), though, demonstrates that there is more to this story when more than just the first two mentions of a word are analyzed. Their work suggests that infants construct detailed acoustic representations and then use these to aid in future word recognition, but they also show a general preference by infants for emphatic material in general. Infants are thus still tuned into the emphatic material, or the more prosodically expressive elements, that are often associated to the new and attention worthy portions of the discourse. Interestingly, the Bortfeld and Morgan (2010) study lends support to production work by Herbst (2007), showing that German-acquiring children stress the first and third mention of a referent, but not the second. More research is still needed to discover the exact types of intonational contours and pitch accents that adults use to express the given and new components in a discourse and how children use this information to formulate their own representations of the information structure.

Third, I examined how adults and toddlers compared in terms of the acoustic correlates of intonation, examining how *standardized duration, mean F_0, maximum F_0,* and *maximum intensity* differed or remained the same across pitch accent types for the different participant populations. All measurements differed significantly by pitch accent type grouping, with group means demonstrating a linear trend from deaccented, H* to L+H* productions. Overall, the *deaccented* tokens were the shortest in duration with lower fundamental frequency and intensity values than the accented tokens. The L+H* targets exhibited the longest durations, with higher fundamental frequency and intensity values. The H* pitch accent fell in between the two extremes, although not always significantly different along all four of these dimensions from the other two acoustic correlates. In general, patterns for toddlers most closely resembled those of female adults, possibly reflecting that they tended to mimic their (predominantly female) caretakers.

The second part of the final aim was to determine which acoustic correlates were most predictive of pitch accent group membership, and to investigate if this differed for adults versus toddlers. A stepwise discriminant function analysis was performed on the adult and the child data separately, with nine acoustic correlates initially entered into the model (see Table 4.2). Both analyses extracted two discriminant functions. The first discriminant function is referred to as Accentuation and separates deaccented tokens accented ones (deaccented vs. {H* and +H*}). The second discriminant function is

referred to as Pitch Complexity and separates the two pitch accents from each other (H* vs. L+H*).

The model that gave the best fit for the adult data consisted of *standardized duration*, *mean F_0, maximum F_0*, and *maximum intensity*. This confirms previous work that demonstrated that these four measurements were the best at differentiating various types of focus as it pertains to information status (Breen et al., 2010). The child model with the best fit consisted of six acoustic measurements. Besides the four listed in the adult model, it also included *minimum intensity* and *mean intensity*.

The adult and child models both show that the four acoustic measurements of *duration*, *mean fundamental frequency*, and *maximum fundamental frequency*, along with *maximum intensity*, are the best predictors of pitch accent type for both adults and toddlers (with toddlers also showing additional selectivity for intensity features). For the adults, the acoustic correlates that were the most important for the Accentuation function, and for discriminating the deaccented productions from the two accented ones, were *duration*, *mean*, *F_0* and *maximum F_0*. For the Pitch Complexity function, *maximum intensity* proved to be the most important in discriminating H* from L+H* productions in adult speech.

For the children, all six of the acoustic measurements used in their model were essential in order to discriminate deaccented from accented productions (the Accentuation discriminant function), but only *minimum intensity* and *maximum F_0* were predictive in separating H* from L+H* targets (the Pitch Complexity discriminant

function). While the two groups vary slightly in the exact mechanisms that differentiate pitch accent types, there is substantial overlap in features as well. One interpretation of these results is that children begin by employing more aspects of the acoustic signal than adults for making information structure distinctions, but that this narrows over time and eventually fewer acoustic correlates are needed to create these distinctions in the information structure.

Methodologically, Experiment 3 makes several specific contributions in the study of early prosodic development. First, it compared child and adult speech collected during the same naturalistic speech task. The majority of previous work is based on controlled speech elicitation tasks, or even rote repetition tasks for young children. By studying spontaneous speech, these results add to the growing body of research on information structure and intonation, and build a more compelling argument for the close inter-relation between these two domains.

Second, this experiment focused on comparing adult speech with that of two and a half year olds. Very few experiments have analyzed the relationship between information structure and intonation in early speech development, and fewer still have attempted to collect spontaneous speech data. Thus, the approach and procedure used in this experiment provide a starting point for future research on the intersection of intonation and information structure such as investigating pitch accent usage in a more natural setting and by using both phonological and phonetic analyses. Finally, the use of discriminant function analysis allowed for an unbiased approach with which to uncover

the acoustic correlates that are the most predictive of pitch accent type membership. This approach builds on recent work that uses a similar statistical methodology to investigate multiple acoustic measurements simultaneously.

4.5 Conclusion

This experiment compared the production abilities of toddlers with those of adults, conducting phonological and phonetic analyses on spontaneous speech data from 2.5-year-olds and adults completing the same task. Additionally, it extended the perceptual findings from experiments 1 and 2 into an investigation in the production domain. Analyses of adult speech replicated earlier findings, showing the use of H* and L+H* pitch accents for *new* and *contrastive* information, with L+H* used nearly exclusively for *contrastive* information. Also, children behaved in a manner very similar to the adults in their usage of pitch accents to represent *new* and *contrastive* information. The one divergence in patterns was for *given* information, where the children deaccented and used H* pitch accents, unlike adults. This pattern likely reflects sensitivities to the phonological statistics of child-directed input speech.

Children are able to employ several acoustic correlates of intonation in order to differentiate pitch accent type, and that they do so in a manner quite similar to adults. The four acoustic measurements explored in depth (*standardized duration, mean F_0, maximum F_0,* and *maximum intensity*) showed that children are able to make controlled distinctions between acoustic measures to express pitch accent differences. Finally, child

and adult pitch accent predictive models demonstrated that both groups employ the same set of acoustic measurements in order to maximally differentiate deaccentuation, H*, and L+H*. Children used a slightly more expanded set of acoustic cues, indicating a possible developmental trajectory of acoustic narrowing.

Chapter 5

Conclusion

5.1 Summary

This dissertation investigated the role of higher-level linguistic processes during first language acquisition. Through a set of three studies, it explored the early development of intonation and information structure from two perspectives: perception and production. Experiments 1 and 2 focused on perception and examined how intonation interacts with different types of information in discourse to guide attention and aid in early word learning. Experiment 3 focused on production and explored how young toddlers produced information structure distinctions in their own speech, both intonationally and acoustically, in comparison to adults. An acoustic component was included in Experiment 3 to provide a more complete intonational and prosodic analysis that included phonology and phonetics together.

Four questions were put forth in the Introduction in Chapter 1 and then addressed through these three studies presented in Chapters 2 through 4. The first question asked how intonation and information structure interact to guide attention. Experiment 1 explored this question by testing how the pitch movement type and information status of a referent affected 18-month-olds' attention. The pitch types tested included a complex bitonal movement (˜L+H*), a simple monotonal (˜H*), and a deaccented variety. The information status of the referent varied in newness; it was either *new* or *given* in the discourse context. An eye-tracking paradigm that analyzed proportional looking to the target showed that these two factors influenced reference resolution.

Control results from Experiment 1 showed that even in the absence of a linguistic label, children looked to the referentially new object. When labels were provided, attention was still guided to the referentially new item regardless of pitch accent type. When referents were *given*, children looked reliably to a target referent when they were introduced with either one of the two pitch movement types (*simple* or *complex*). This stands in contrast to previous work by Grassmann and Tomasello (2010) that claimed children only look to a referent if it is both referentially new and carries a stressed linguistic label.

The second fundamental question of this dissertation inquired into the role of the isolated acoustic correlate of fundamental frequency in guiding attention. The secondary aim of Experiment 1 addressed this question by examining how the primary acoustic correlate of intonation guided attention. The primary acoustic correlates to intonation are the fundamental frequency (F_0), duration, intensity, and segmental quality and reduction (Shattuck-Hufnagel & Turk, 1996). Research on early acoustic preferences has shown that F_0 is the preferred correlate to intonation for young infants (Fernald & Kuhl, 1985). This measurement is also typically considered to underlie the basic melody patterns found in the intonation (Ladd, 1996). The results of Experiment 1 reflect only differences in the F_0 contour, with duration and average intensity held constant. While Experiment 1 demonstrated robust results regarding F_0 without all of the other acoustic correlates in use, I hypothesized that the full range of cues will increase attention in the pitch accent conditions (H* and L+H*). This is especially expected for the L+H* accent,

which typically exhibits duration differences that enhance its prominence. Experiment 2 employed all of the acoustic cues when testing the effects of different pitch accents.

Experiment 2 focused on the third motivating question and asked if specific aspects of the intonation and information structure aid in early word learning. To explore this question, pitch accent type and contrastiveness were manipulated during a novel word learning task. During learning, two-year-olds were exposed to a novel word that was either deaccented or received a pitch accent (H* or L+H*). Additionally, the word was either presented in contrast to a previous referent or not (contrastive vs. noncontrastive setting). Results from this study confirmed predictions that the prominent L+H* pitch accent increased learning of a novel word, especially if it was used in a contrastive setting. The results from the H* and deaccented learning conditions proved more variable.

The effect of pitch accent type on novel word learning was most evident for the L+H* pitch accent. As predicted, there was no learning of the novel word for the deaccented conditions (particularly in the noncontrastive setting). Although learning was not observed for words carrying an H* pitch accent, there was a trend for increased learning when the H* productions were presented noncontrastively. This was particularly evident early in processing. Processing speed variations (early vs. late) were also shown for the other conditions as well, with L+H* showing an increase in fixations over time. More detailed analyses that include time as a continuous variable are necessary to better

account for the effects of pitch accent type and information status on novel word learning.

Experiment 3 examined the fourth and final question, asking if young toddlers make information structure distinctions in the same way as adults. Additionally, this experiment investigated whether or not toddler speech differed phonologically and/or acoustic-phonetically from the speech of adults. This experiment tested adults and two-and-half-year-olds in a natural environment and elicited speech during a spontaneous production task designed as an interactive game. Utterances containing target noun productions were segmented from the data afterward. The analysis looked at targets that were *new, given*, or *contrastive*. Each word was labeled following the ToBI transcription system (Beckman & Ayers, 1997).

For the adult participants, phonological results from Experiment 3 confirmed previous findings showing that *new* information is produced with an H* pitch accent, *contrastive* information with either an H* or a L+H*, and *given* information with deaccentuation. The child speech showed very similar patterns for *new* and *contrastive* information. For *given* referents, children used both deaccentuation and the H* pitch accent. This effect is likely to be a reflection of a pattern present in child-directed speech where referents are typically produced with alternating stress patterns (stressed, unstressed, stressed, etc.), and are deaccented to a lesser degree (Bortfeld & Morgan, 2010).

For the phonetic analysis of Experiment 3, a stepwise discriminant function analysis showed that the four acoustic measurements that best predicted pitch accent category membership were duration, maximum F_0, mean F_0, and maximum intensity. This confirmed recent findings by Breen et al. (2010) showing that these same four acoustic correlates were also predictive of information structure (i.e., focus type, breadth, and position). Analysis of the child speech data showed that these four acoustic measurements, along with minimum and mean intensity, best predicted pitch accent type membership. Neither discriminant function analysis in Experiment 3 classified the L+H* and the deaccented target nouns above 50% accuracy. This seems to be evidence in favor of the phonological or *indirect* status of pitch accent type. According to this type of analysis, acoustic measurements do not map directly onto phonological categories, which is consistent with the AM approach. Alternatively, this inability to classify by pitch complexity type (H* vs. L+H*) may be due to which acoustic features were analyzed. Future analyses will include acoustic measurements that better differentiate these two pitch accents, such as ones that better take into account contour shape and alignment.

All three experiments confirm that intonation and information structure both play a role in directing toddler attention and facilitating early word learning. Experiment 1 demonstrated that either newness or the presence of a pitch accent (i.e., F_0 variation only) guide 18-month-olds' attention during reference resolution. Experiment 2 showed that a complex pitch accent and contrastiveness aid 24-month-olds in learning a

novel word. Even with live interactions removed from the procedures, Experiments 1 and 2 revealed the interacting effects of prosody and information structure on attention as well as on successful word learning. Experiment 3 showed how toddlers (2;6) phonologically and acoustic-phonetically produced information structure differences in a manner comparable to the ones found in adult speech.

With prior experimental work employing complicated methodologies for the age ranges analyzed, this series of studies took extra care to ensure that all experimental methods were age-appropriate. This ensured a better and more accurate understanding of how young toddlers typically employ different aspects of the prosodic system. As discussed, most previous experiments lack a joint *phonological and acoustic analysis* of the data, and instead primarily rely on perceptual descriptions. To address this issue, I implemented the autosegmental-metrical approach for analysis, challenging current intonational development approaches and theories in American English by providing a more detailed and thorough evaluation of the data. In addition, I analyzed the data from an acoustic-phonetic perspective alongside the phonological one. A further advantage of this line of research is that by testing different age groups we can illuminate the interaction between intonation and information structure over different periods of development. The three experiments together support the fundamental goal of my dissertation: to explore the ways that intonation interacts with our ability to correctly interpret and produce informationally-structured categories in day-to-day discourse settings, and how this ability develops during early language acquisition.

5.2 Limitations and Extensions

All three of the experiments presented in this dissertation struggled with the issue of sample size. The two factors affecting the number of participants in the studies were recruitment and age. First, recruitment is difficult for any developmental study, but even more so in recent years as access to potential participants becomes increasingly limited. The in-house database used to recruit families with young children relies on calling landline phones, which in the past has proven to be an effective tool. Unfortunately, as the number of landline phones decreases, it has become more difficult to contact potential families for participation. Second, the age range analyzed was from 1.5- to 2.5-years-old, and there was a relatively high amount of attrition due to disinterest in the task and inattention. Steps were taken in all three experiments to reduce attrition as much as possible by using age-appropriate procedures and limiting the duration of the task.

A limitation and advantage of Experiment 1 is that it focused only on the isolated F_0 contour as a cue to intonation. Without an experiment showing the effect of the full range of acoustic cues on attention, it is currently unknown how children would behave when dealing with typical contour variability in this task. In order to strengthen Experiment 1, additional studies are necessary that would look at the combination of all of the acoustic cues together as well as each one individually (e.g., duration, intensity). Still, this experiment remains powerful in that it demonstrates a strong effect on attention and reference resolution when only one acoustic dimension is manipulated.

The primary limitation of Experiment 2 was due to task design. The procedure for this study included two learning phases followed by a single test phase. Consequently, participants needed to complete both learning blocks before the test phase began. Since inattention was a major limitation in all of the studies, many participants only completed half of the study and thus yielded no data on whether the word-object association had been learned or not. A possibility for future experiments is to include multiple test phases during the task. In this way, even if the study were not completed, there would still be partial data collected.

Finally, the main limitation of Experiment 3 was also its greatest strength: a natural speech setting. Spontaneous speech is difficult to analyze for many reasons including variations in the rate and loudness of speech, inconsistencies in naming (e.g., lamb vs. sheep), changes in the syntactic structure, sentence position of target words, etc. All of these also affect the realization of the acoustic signal, making analysis within and across subjects more difficult. Even so, it is essential to analyze ecologically valid speech samples alongside controlled laboratory experiments. The aim of this study was not to look at identical utterances across subjects, but to observe how participants employed intonation during a natural speech task. Future work would use both this method and more controlled designs in order to further explore the effects of information status on intonation in early speech development.

The overall success of this set of experiments outweighs the difficulties encountered. Experiments 1 and 2 clearly demonstrate how intonation interacts with

information structure to guide attention and aid early word learning. These experiments specifically isolated pitch accents and context and were thus able to explicitly analyze their effects on early perceptual processing and abilities. Experiment 3 strengthened phonological and acoustic models of child speech. It showed how children employ intonation to express information structure in a natural discourse setting, demonstrating a near adult-like level of acoustic control. Altogether, these experiments strengthen and nuance our understanding of how the mapping between intonation and information structure perceptually and productively develops during first language acquisition.

References

Akhtar, N., Carpenter, M., & Tomasello, M. (1996). The role of discourse novelty in early word learning. *Child Development, 67*, 635-645.

Allen, S., Skarabela, B., & Hughes, M. (2008). Using corpora to examine discourse effects in syntax. In H. Behrens (Ed.), *Corpora in Language Acquisition Research: Finding Structure in Data* (pp. 99-138). Amsterdam: John Benjamins.

Ariel, M. (1990). *Accessing Noun Phrase Antecedents*. London: Routledge.

Arnold, J. (2008). THE BACON not the bacon: how children and adults understand accented and unaccented noun phrases. *Cognition, 108*, 69-99.

Arvaniti, A., & Baltazani, M. (2000). Greek ToBI: a system for the annotation of Greek speech corpora. *Proceedings of the 2nd International Conference on Language Resources and Evaluation, Vol. II,* (pp. 555-562). Athens: European Language Resources Association.

Arvaniti, A., Ladd, D. R., & Mennen, I. (1998). Stability of tonal alignment: The case of Greek prenuclear accents. *Journal of Phonetics, 26*, 3-25.

Astruc, L., Payne, E., Post, B., Vanrell, M.M., & Prieto, P. (2013). Tonal targets in early child Catalan, Spanish, and English. *Language and Speech, 56*(2), 229-253.

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech, 47*, 31-56.

Baayen, R. H. (2007). languageR: Data sets and functions with "Analyzing Linguistic Data: A practical introduction to statistics". R package version 0.2.

Baayen, R.H., Davidson, D.J., & Bates, D.M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412.

Baldwin, D. A., & Markman, E. M. (1989). Establishing Word-Object Relations: A First Step. *Child Development, 60*, 381-398.

Balog, H. L., & Snow, D. (2007). The adaptation and application of relational and independent analyses for intonation production in young children. *Journal of Phonetics, 35*, 118-133.

Baltaxe, C. (1984). Use of contrastive stress in normal, aphasic, and autistic children. *Journal of Speech and Hearing Research, 27*, 97-105.

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language, 59,* 457-474.

Bates, D. (2007). *lme4: Linear mixed-effects models using S4 classes*. (R package version 0.99875-2).

Baumann, S., & Roth, A. (2014). Prominence and Coreference: On the Perceptual Relevance of F0 Movement, Duration and Intensity. *Proceedings of Speech Prosody 7* (pp. 227-231).

Beckman, M. E. (1986). *Stress and non-stress accent.* Dordrecht: Foris.

Beckman, M. E., & Ayers, G.M. (1997). Guidelines for ToBI labeling (Version 7.0). Columbus: Ohio State Uni-versify. Available from http://ling.ohio-state.edu/Phonetics/E-ToBI

Beckman, M. E., Hirschberg, J., & Shattuck-Hufnagel, S. (2004). The Original ToBI System and the Evolution of the ToBI Framework. In Sun-Ah Jun (Ed.), *Prosodic Typology: The phonology of intonation and phrasing* (Chapter 2). Oxford: Oxford University Press.

Beckman, M., & Pierrehumbert, J. B. (1986). Intonational structure in English and Japanese. *Phonology Yearbook, 3*, 255-310.

Behrens, H., & Gut, U. (2005). The relationship between prosodic and syntactic organization in early multiword speech. *Journal of Child Language, 32*, 1-34.

Birch, S., & Clifton, C. J. (1995). Focus, accent, and argument structure: effects on language comprehension. *Language and Speech, 38*(4), 365-391.

Boersma, P. & Weenink, D. (2014). *Praat: A system for doing phonetics by computer* [Software]. Available from http://www.praat.org

Bolinger, D. L. (1965). *Forms of English.* Cambridge, MA: Harvard University Press.

Bolinger, D. L. (1972). Accent is Predictable (If You're a Mind-Reader). *Language, 48*(3), 633-644.

Bolinger, D. L. (1982). Intonation and Its Parts. *Language, 58*(3), 505-533.

Bolinger, D. L. (1989). *Intonation and Its Uses: Melody in Grammar and Discourse.* Stanford: Stanford University Press.

Bortfeld, H., & Morgan, J. (2010). Is early word-form processing stress-full? How natural variability helps recognition. *Cognitive Psychology, 60*, 241-266.

Bortfeld, H., Morgan, J., Golinkoff, R. M., & Rathbun, K. (2005). Mommy and me: Familiar names help launch babies into speech-stream segmentation. *Psychological Science, 16*(4), 298-304.

Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes, 25*(7/8/9), 1044-1098.

Brooks, R., & Meltzoff, A. N. (2002). The importance of eyes: How infants interpret adult looking behavior. *Developmental Psychology, 38*, 958-966.

Büring, D. (2012). Focus and Intonation. In G. Russell & D. G. Fara (Eds.), *Routledge Companion to the Philosophy of Language* (pp. 103-115). London: Routledge.

Calhoun, S. (2006). *Intonation and Information Structure in English*. PhD dissertation. University of Edinburgh.

Canfield, R. L., Smith, E. G., Brezsnyak, M. P., & Snow, K. L. (1997). Information processing through the first year of life: A longitudinal study using the Visual Expectation Paradigm. *Monographs of the Society for Research in Child Development, 62*(2), 1-145.

Chafe, W. L. (1976). Givenness, constrastiveness, definiteness, subjects, topics and point of view. In C. N. Li (Ed.), *Subject and Topic* (pp. 27-55). New York: Academic Press.

Chafe, W. L. (1987). Cognitive constraints on information flow. In Tomlin (Ed.), *Coherence and Grounding in Discourse* (pp. 21-51). Amsterdam: John Benjamins.

Chen, A. (2010). Is there really an asymmetry in the acquisition of the focus-to-accentuation mapping? *Lingua, 120*, 1926-1939.

Chen, A. (2011). Tuning the information structure: intonational realization of topic and focus in child Dutch. *Journal of Child Language, 38*, 1055-1083.

Chen, A., & Fikkert, P. (2007). Intonation of early two-word utterances in Dutch. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the XVIth International Congress of Phonetic Sciences* (pp. 315-320).

Clark, H. H., & Haviland, S. E. (1977). Comprehension and the given-new contract. In R. O. Freedle (Ed.), *Discourse production and comprehension* (pp. 1-40). Hillsdale, NJ: Erlbaum.

Cruttenden, A. (1986). *Intonation*. University Press: Cambridge.

Cutler, A., & Swinney, D. A. (1987). Prosody and the development of comprehension. *Journal of Child Language, 14*, 145-167.

DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science, 208*, 1174-1176.

DeCasper, A. J., & Spence, M. J. (1986). Prenatal maternal speech influences newborn's perception of speech sounds. *Infant Behavior and Development, 9*, 133-150.

Diesendruck, G., Markson, L., Akhtar, N., & Reudor, A. (2004). Two-year-olds' sensitivity to speakers' intent: an alternative account of Samuelson and Smith. *Developmental Science, 7*(1), 33-41.

Dilley, L. C., Ladd, D.R., & Schepman, A. (2005). Alignment of L and H in bitonal pitch accents: Testing two hypotheses. *Journal of Phonetics, 33*(1), 115-119.

D'Odorico, L., & Carubbi, S. (2003). Prosodic characteristics of early multi-word utterances in Italian Children. *First Language, 23*(I), 97-116.

D'Odorico, L., & Fasolo, M. (2007). The prosody of early multi-word speech: word order and its intonational realization in the speech production of Italian children. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the XVIth International Congress of Phonetic Sciences* (pp. 321-325).

Estebas-Vilaplana, E., & Prieto, P. (2008). La notación prosódica en español. Una revisión del Sp_ToBI. *Estudios de Fonética Experimental, XVIII*, 263-283.

Fantz, R. L. (1964). Visual experience in infants: Decreased attention to familiar patterns relative to novel ones. *Science, 146,* 668-670.

Fenson, L., Marchman, V.A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (2007). *MacArthur-Bates Communicative Development Inventories: User's guide and technical manual (2nd ed)*. Baltimore, MD: Brookes.

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development, 8,* 181-195.

Fernald, A., & Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development, 10,* 279-293.

Fernald, A., Pinto, J. P, Swingley, D., Weinberg, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the second year. *Psychological Science, 9*(3), 228-231.

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology, 20*, 104-113.

Féry, C., & Kügler, F. (2008). Pitch accent scaling on given, new and focused constituents in German. *Journal of Phonetics, 36*, 680-703.

Fisher, C., & Tokura, H. (1995). The given/new contract in speech to infants. *Journal of Memory and Language, 34*, 287–310.

Fowler, C., & Housum, J. (1987). Talkers' signaling of 'new' and 'old' words in speech and listener's perception and use of that distinction. *Journal of Memory and Language, 26*, 489-504.

Frota, S., & Vigário, M. (2008). The intonation of one-word and first two-word utterances in European Portuguese. At *XI International Conference for the Study of Child Language Conference*.

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America, 27,* 765-768.

Golinkoff, R. M., Hirsh-Pasek, K., Bailey, L. M., & Wenger, N. R. (1992). Young children and adults use lexical principles to learn new nouns. *Developmental Psychology, 28*(1), 99-108.

Grassmann, S., & Tomasello, M. (2007). Two-year-olds use primary sentence accent to learn new words. *Journal of Child Language, 14*, 23-45.

Grassmann, S., & Tomasello, M. (2010). Prosodic stress on a word directs 24-month-olds' attention to a contextually new referent. *Journal of Pragmatics*, *42*(11), 3098-3105.

Greenfield, P. (1979). Informativeness, presupposition, and semantic choice in single-word utterances. In E. Ochs & B. B. Schieffelin (Eds.), *Developmental Pragmatics*. New York: Academic Press.

Greenfield, P. M., & Smith, J. (1976). *The structure of communication in language development.* New York: Academic Press.

Grice, P. (1975). Logic and Conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics, 3: Speech acts* (pp. 41-58). New York: Academic Press.

Gundel, J. K. (2003). Information Structure and referential givenness/newness. How much belongs in the grammar? *Journal of Cognitive Science, 4*, 177-199.

Gundel, J. K., & Fretheim, T. (1993). Topic and Focus. In G. Ward and L. Horn (Eds.), *Handbook of Pragmatics* (pp. 175-196). London: Blackwell.

Gundel, J. K., Hedberg, N., Zacharaski, R. (1993). Cognitive status and the form of the referring expressions in a discourse. *Language, 69*, 274-307.

Gussenhoven, C. (1983). Focus, Mode, and the Nucleus. *Journal of Linguistics, 19*, 377-417.

Gussenhoven, C. (2002). Intonation and Interpretation: Phonetics and Phonology. *Proceedings from the 1st International Conference on Speech Prosody* (pp. 47-57).

Gussenhoven, C. (2004). *The Phonology of Tone and Intonation.* Cambridge: Cambridge University Press.

Haith, M. M., Wentworth, N., & Canfield, R. L. (1993). The formation of expectations in early infancy. In C. Rovee-Collier & L. P. Lipsitt (Eds.), *Advances in infancy research* (pp. 251-297). Norwood, NJ: Ablex.

Halliday, M. A. K. (1967). Notes on transitivity and theme in English. Part II. *Journal of Linguistics, 3*, 199-244.

Hedberg, N., & Sosa, J. M. (2007). The Prosody of Topic and Focus in Spontaneous English Dialogue. In C. Lee, M. Gordon, and D. Büring (Eds.), *Topic and Focus: Cross-Linguistic Perspectives on Meaning and Intonation.* Dordrecht: Springer. 101-120.

Herbst, L. (2007). German 5-year-olds' Intonational Marking of Information Status. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the XVIth International Congress of Phonetic Sciences* (pp. 1557-1560).

Hornby, P. A. (1971). Surface structure and the topic-comment distinction: a developmental study. *Child Development, 42*, 1975-1988.

Hornby, P. A., & Hass, W. A. (1970). Use of contrastive stress by preschool children. *Journal of Speech and Hearing Research, 13*, 359-399.

IBM Corp. (2013). IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp.

Ito, K., Speer, S., & Beckman, M. E. (2004). Informational Status and Pitch Accent Distribution in Spontaneous Dialogues in English. *Proceedings of Speech Prosody 2004, an international conference* (pp. 289-282).

Jackendoff, R. (1972). *Semantic Interpretation in Generative Grammar*. Cambridge, MA: MIT Press.

Jun, S-A (Ed.). (2005). *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press.

Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology, 29*(1), 1-23.

Jusczyk, P.W., Houston, D., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology, 39*, 159-207.

Jusczyk, P. W., & Thompson, E. J. (1978). Perception of a phonetic contrast in multisyllabic utterances by two-month-old infants. *Perception & Psychophysics, 23*, 105–109.

Katz, J., & Selkirk, E. (2011). Contrastive focus vs. discourse-new: Evidence from phonetic prominence in English. *Language 87*(4), 771-816.

Kruijff-Korbayová, I., & Steedman, M. (2003). Discourse and Information Structure. *Journal of Logic, Language, and Information*, *12*(3), 249-259.

Kurumada, C., Brown, M., & Tanenhaus, M. (2012). Distributional learning in pragmatic interpretation of prosody. Poster presented at the *37th Annual Boston University Conference on Language Development (BUCLD).*

Ladd, D. R. (1996). *Intonational phonology.* Cambridge: Cambridge University Press.

Ladd, D. R., Faulkner, D., Faulkner, H., & Schepman, A. (1999). Constant "segmental anchoring" of F0 movements under changes in speech rate. *Journal of the Acoustical Society of America, 106*(3), 1543-54.

Ladd, D. R., Mennen, I., & Schepman, A. (2000). Phonological conditioning of peak alignment in rising pitch accents in Dutch. *Journal of the Acoustical Society of America, 107*(5), 2685-96.

Ladd, D. R., & Schepman, A. (2003). "Sagging transitions" between high accent peaks in English: Experimental evidence. *Journal of Phonetics, 31*, 81-112.

Lehiste, I. (1970). *Suprasegmentals.* Cambridge, MA: MIT Press.

Lieberman, P. (1960). Some acoustic correlates of word stress in American English. Cambridge, MA: MIT Press. *Journal of the Acoustical Society of America, 32,* 451-454.

Lieberman, P. (1967). *Intonation, Perception and Language.* Cambridge, MA: MIT Press.

MacWhinney, B., & Bates, E. (1978). Sentential devices for conveying givenness and newness: a cross-cultural developmental study. *Journal of Verbal Learning and Verbal Behavior, 17*, 539-558.

Mathesius, V. (1928). On linguistic characterology with illustrations from Modern English. *Actes du Premier Congrès International de Linguistes à La Haye,* 56-63.

Matin, E., Shao, K.C., & Boff, K.R. (1993). Saccadic overhead: Information processing time with and without saccade. *Perception & Psychophysics, 53*, 372-380.

Mehler, J., Bertoncini, J., Barrierre, M., & Jassik-Gershenfeld, D. (1978). Infant recognition of mother's voice. *Perception*, *7*, 491-497.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition,* 29, 143-178.

Mervis, C. B., & Bertrand, J. (1994). Acquisition of the novel name-nameless category (N3C) principle. *Child Development, 65*, 1646-1662.

Moore, C., Angelopoulos, M., & Bennett, P. (1999). Word learning in the context of referential and salience cues. *Developmental Psychology, 35*(1), 60-68.

Müller, A., Höhle, B., Schmitz, M., & Weissenborn, J. (2006). Focus-to-Stress Alignment in 4- to 5-year-old German-learning Children. In A. Belletti, E. Bennati, C. Chesi, E. Di Domenico, & I. Ferrari (Eds.), *Language Acquisition and Development. Proceedings of the GALA 2005 Conference* (pp. 393-407). Cambridge: Cambridge Scholars Press.

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance, 24*(3), 756-766.

Nazzi, T., Floccia, C., & Bertoncini, J. (1998). Discrimination of pitch contours by neonates. *Infant Behavior and Development, 21*(4), 779-784.

O'Neill, D., & Happé, F. (2000). Noticing and commenting on what's new: differences and similarities among 22-month-old typically developing children, children with Down Syndrome and children with autism. *Developmental Science, 3*, 457-458.

Oviatt, S. L. (1980). The emerging ability to comprehend language: an experimental approach. *Child Development, 51*, 97–106.

Palmer, H. E. (1922). *English intonation with systematic exercises*. Cambridge: Heffer.

Patel, R., & Brayton, J. (2009). Identifying Prosodic Contrasts in Utterances Produced by 4-, 7-, and 11-Year-Old Children. *Journal of Speech, Language, and Hearing Research, 52*, 206-222.

Peppé, S., & McCann, J. (2003). Assessing intonation and prosody in children with atypical language development: the PEPS-C test and the revised version. *Clinical Linguistics & Phonetics, 17*, 345-354.

Pierrehumbert, J. B. (1980). *The Phonology and phonetics of English intonation*. PhD dissertation. MIT.

Pierrehumbert, J. B., & Hirschberg, J. (1990). The Meaning of Intonational Contours in the Interpretation of Discourse. In P. Cohen, J. Morgan & M. Pollack (Eds.), *Intentions in Communication* (pp. 271-311). Cambridge, MA: MIT Press.

Prieto, P., Aguilar, L., Mascaró, I., Torres-Tamarit, F. J., Vanrell, M.M. (2009). L'etiquetatge prosòdic Cat_ToBI. *Estudios de Fonética Experimental XVIII*, 287-309.

Prieto, P., Estrella, A., Thorson, J., & Vanrell, M.M. (2012). Is prosodic development correlated with grammatical and lexical development? Evidence from emerging intonation in Catalan and Spanish. *Journal of Child Language, 39*(2), 221-257.

Prieto, P., & Vanrell, M.M. (2007). Early intonational development in Catalan. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the XVIth International Congress of Phonetic Sciences* (pp. 309-314).

Prince, E. F. (1981). Towards a taxonomy of given-new information. In P. Cole (Ed.), *Radical Pragmatics.* New York: Academic Press.

R Development Core Team. (2007). *R: A language and environment for statistical computing.* Vienna, Austria. (ISBN 3-900051-07-0).

Rose, S. A., Gottfried, A. W., Melloy-Carminar, P., & Bridger, W. H. (1982). Familiarity and novelty preferences in infant recognition memory: Implications for information processing. *Developmental Psychology, 18*, 704-713.

Samuelson, L. K., & Smith, L. B. (1998). Memory and attention make smart word learning: An alternative account of Akhtar, Carpenter and Tomasello. *Child Development, 1*, 94-104.

Seidl, A., & Johnson, E. K. (2006). Infant word segmentation revisited: edge alignment facilitates target extraction. *Developmental Science, 9*(6), 565-573.

Sgall, P. (I967). Functional sentence perspective in a generative description. *Prague studies in mathematical linguistics, 2*, 203-225. Prague: Academia.

Shattuck-Hufnagel, S., & Turk, A. E. (1996). A Prosody Tutorial for Investigators of Auditory Sentence Processing. *Journal of Psycholinguistic Research, 25*(2), 193-247.

Silverman, K., Beckman, M., Petrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., & Hirschberg, J. (1992). ToBI: A standard for labeling English prosody. *Proceedings of ICSLP, 92*(2), 867-870.

Sluijter, A.C.M., & van Heuven, V. J. (1996). Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America, 100*, 2471–2485.

Snow, D. (2006). Regression and reorganization of intonation between 6 and 23 months. *Child Development*, *77*, 281-296.

Song, H-j., & Fisher, C. (2005). Who's "she"? Discourse prominence influences preschoolers' comprehension of pronouns. *Journal of Memory and Language, 52*, 29-57.

Steedman, M. (1991). Structure and Intonation. *Language, 67*(2), 260-296.

Steedman, M. (2007). Information-Structural Semantics for English Intonation. In C. Lee, M. Gorgon & D. Büring (Eds.), *Topic & Focus: Cross-linguistic Perspectives on Meaning and Intonation*. Kluwer: Dordrecht.

Swingley, D., & Fernald, A. (2002). Recognition of words referring to present and absent objects by 24-month-olds. *Journal of Memory and Language, 46,* 39-56.

Taylor, P. (2000). Analysis and Synthesis of Intonation Using the Tilt Model. *Journal of the Acoustical Society of America, 107*(3), 1697-1714.

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-Directed Speech Facilitates Word Segmentation. *Infancy, 7*(1), 53-71.

Thorson, J. C., & Morgan, J. L. (2014). Directing Toddler Attention: Intonation and Information Structure. In *Supplement to the Proceedings of the 38th Annual Boston University Conference on Language Development.* Somerville: Cascadilla Press.

Tomasello, M., & Akhtar, N. (1995). Two-year-olds use pragmatic cues to differentiate reference to objects and actions. *Cognitive Development, 10*, 201-224.

Turnbull, R., Royer, A. J., Ito, K., & Speer, S. (2014). Prominence perception in and out of context. *Proceedings of Speech Prosody 7* (pp. 1164-1168).

Vallduví, E. (1992). *The Informational Component.* New York: Garland.

Vallduví, E., & Engdahl, E. (1996). The linguistic realization of information packaging. *Linguistics, 34*, 459-519.

Venditti, J. J., Hirschberg, J. (2003). Intonation and discourse processing. *Proceedings of the XVth International Congress of Phonetic Sciences*.

Watson, D. (2010). The Many Roads to Prominence: Understanding Emphasis in Conversation. *Psychology of Learning and Motivation, 52*, 163-183.

Watson, D., Gunlogson, C., & Tanenhaus, M. (2008). Interpreting pitch accents in on-line comprehension: H* vs L+H*. *Cognitive Science, 32*, 1232-1244.

Wagner, M., & Watson, D. G. (2010). Experimental and theoretical advances in prosody: a review. *Language and Cognitive Processes, 25*(7/8/9), 905-945.

Waxman, S. R., & Lidz, J. (2006). Early Word Learning. In D. Kuhn & R. Siegler (Eds.), *Handbook of Child Psychology, 6th Edition, Volume 2: Cognition, Perception, and Language* (pp. 299-335). Hoboken, New Jersey: Wiley.

Wells, B., & Peppé, S. (2001). Intonation within a psycholinguistic framework. In J. Stackhouse & B. Wells (Eds.), *Children's speech and literacy difficulties 2: identification and intervention*. London: Whurr Publishers.

Wells, B., & Peppé, S. (2003). Intonation abilities of children with speech and language impairment. *Journal of Speech, Language and Hearing Research, 46*(1), 5-20.

Wells, B., Peppé, S., & Goulandris, N. (2004). Intonation development from five to thirteen. *Journal of Child Language, 31*(4), 749-778.

Wiemen, L. (1976). Stress patterns of early child language. *Journal of Child Language*, *3*, 283-286.

Wonnacott, E., & Watson, D. (2008). Acoustic emphasis in four year olds. *Cognition, 107,* 1093-1101.

Woodward, A. L. (2003). Infants' developing understanding of the link between looker and object. *Developmental Science, 6*(3), 297-311.

Xu, Y. (1998). Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica*, *55*, 179-203.

Xu, Y., & Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics, 33*(2), 159-197.

Yu, K. M., Khan, S. D., Sundara, M. (2014). Intonational phonology in Bengali and English infant-directed speech. *Proceedings of Speech Prosody 7* (pp. 1130-1134).

APPENDIX A

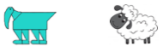
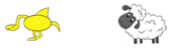
Experiment 2 Stimuli

Experiment 2 Stimuli

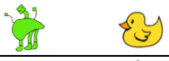
Full list of stimuli used the two learning conditions of Experiment 2: noncontrastive and contrastive. The audio stimuli are listed in the left column and the visual stimuli in the right. All images were counterbalanced for the side that they appeared on during the learning phase (left or right). Target novel animals were also counterbalanced across participants (alternating with the distractor animal in that condition).

For the spoken stimuli, if a word is marked in all caps then it was produced with an H* pitch accent, it the word is in lower case letters it was deaccented. The novel words of wug and neem (marked with underlining and an asterisk) varied by pitch accent type condition. In the deaccented learning condition the target words were all deaccented, in the simple monotonal H* learning condition they carried an H* pitch accent, and in the complex bitonal L+H* learning condition they carried a L+H* pitch accent.

Learning Phase – Noncontrastive

Spoken Stimuli	Visual Stimuli
1. Look! A BEAR and a PIG.	
2. Oh! There is a SHEEP and a wug.*	
3. Look! Now there is a SHEEP again.	
4. Look! There is a wug and a PIG.*	
5. Look! There is another PIG.	
6. Oh! Now there is a SHEEP and a BEAR over there.	
7. Oh! There is a PIG over there.	
8. Look! There is a wug and a SHEEP.*	
9. OH! There is a PIG and a BEAR.	
10. Look! There is another SHEEP over there.	
11. Oh, look! There is a sheep and a BEAR.	
12. Look! Now there is a PIG and a wug.*	

Learning Phase – Contrastive

Spoken Stimuli	Visual Stimuli
1. Look! A DUCK and a BEAR.	
2. Look! Now there is a neem and a bear.*	
3. Look! Now there is a DUCK over there.	
4. Oh! There is another DUCK over there.	
5. Oh, look! There is a duck and a BEAR.	
6. Look! Now there is a duck and a neem.*	
7. Oh! There is a BEAR over there.	
8. Oh! Now there is a bear and a DUCK over there.	
9. Look! There is a bear and a neem over there.*	
10. Look! There is another BEAR over there.	
11. Oh! There is a PIG and a DUCK.	
12. Look! There is a neem and a duck.*	

APPENDIX B

EXPERIMENT 3 FULL DFA MODELS

Experiment 3 Full DFA Models

Listed are the stepwise statistics for both the adult and child full model outputs for the discriminant function analyses.

Adult Model:

Stepwise Statistics

Variables Entered/Removed[a,b,c,d]

Step	Entered	Min. D Squared					
		Statistic	Between Groups	Exact F			
				Statistic	df1	df2	Sig.
1	Syllable_F0avg_Z	.421	H* and L+H*	20.006	1	261.000	1.153E-005
2	Syllable_F0max_Z	.667	Deaccented and H*	11.654	2	260.000	1.422E-005
3	Syllable_IntMax_Z	1.104	H* and L+H*	17.360	3	259.000	2.664E-010
4	Syllable_DurZscore	1.220	Deaccented and H*	10.582	4	258.000	5.914E-008

At each step, the variable that maximizes the Mahalanobis distance between the two closest groups is entered.[a,b,c,d]

a. Maximum number of steps is 18.

b. Maximum significance of F to enter is .05.

c. Minimum significance of F to remove is .10.

d. F level, tolerance, or VIN insufficient for further computation.

Variables in the Analysis

Step		Tolerance	Sig. of F to Remove	Min. D Squared	Between Groups
1	Syllable_F0avg_Z	1.000	.000		
2	Syllable_F0avg_Z	.344	.000	.110	Deaccented and H*
	Syllable_F0max_Z	.344	.000	.421	H* and L+H*
3	Syllable_F0avg_Z	.324	.005	.925	Deaccented and H*
	Syllable_F0max_Z	.342	.000	.426	H* and L+H*
	Syllable_IntMax_Z	.895	.000	.667	Deaccented and H*
4	Syllable_F0avg_Z	.316	.008	.927	Deaccented and H*
	Syllable_F0max_Z	.310	.000	.955	H* and L+H*
	Syllable_IntMax_Z	.891	.000	.674	Deaccented and H*
	Syllable_DurZscore	.870	.006	1.104	H* and L+H*

Summary of Canonical Discriminant Functions

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	.343[a]	73.6	73.6	.505
2	.123[a]	26.4	100.0	.331

a. First 2 canonical discriminant functions were used in the analysis.

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 2	.663	106.676	8	.000
2	.890	30.151	3	.000

Structure Matrix

	Function	
	1	2
Syllable_F0max_Z	.830*	-.278
Syllable_F0avg_Z	.749*	.221
Sqrt1_Syl_F0Range_Z[b]	.618*	-.558
Syllable_DurZscore	.615*	-.384
Syllable_IntRange_Z[b]	.370*	-.098
Syllable_IntAvg_Z[b]	.502	.636*
Syllable_IntMax_Z	.600	.625*
Syllable_F0min_Z[b]	.365	.513*
Syllable_IntMin_Z[b]	-.045	.417*

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions

Variables ordered by absolute size of correlation within function.

*. Largest absolute correlation between each variable and any discriminant function

b. This variable not used in the analysis.

Child Model:

Stepwise Statistics

Variables Entered/Removed[a,b,c,d]

Step	Entered	Min. D Squared					
		Statistic	Between Groups	Exact F			
				Statistic	df1	df2	Sig.
1	Syllable_DurZscore	.235	H* and L+H*	12.520	1	381.000	.000
2	Syllable_IntAvg	.380	deaccented and H*	6.476	2	380.000	.002
3	Syllable_IntMax	.539	deaccented and H*	6.118	3	379.000	.000
4	Syllable_IntMin	.745	deaccented and H*	6.322	4	378.000	6.214E-005
5	Syllable_F0avg	.838	deaccented and H*	5.673	5	377.000	4.648E-005
6	Syllable_F0max	.948	deaccented and H*	5.335	6	376.000	2.672E-005

At each step, the variable that maximizes the Mahalanobis distance between the two closest groups is entered.[a,b,c,d]

a. Maximum number of steps is 16.

b. Maximum significance of F to enter is .05.

c. Minimum significance of F to remove is .10.

d. F level, tolerance, or VIN insufficient for further computation.

Variables in the Analysis

Step		Tolerance	Sig. of F to Remove	Min. D Squared	Between Groups
1	Syllable_DurZscore	1.000	.000		
2	Syllable_DurZscore	.995	.000	.084	deaccented and H*
	Syllable_IntAvg	.995	.000	.235	H* and L+H*
3	Syllable_DurZscore	.750	.000	.090	deaccented and H*
	Syllable_IntAvg	.079	.023	.312	deaccented and H*
	Syllable_IntMax	.079	.009	.380	deaccented and H*
4	Syllable_DurZscore	.740	.000	.369	deaccented and H*
	Syllable_IntAvg	.057	.001	.330	deaccented and H*
	Syllable_IntMax	.065	.003	.440	deaccented and H*
	Syllable_IntMin	.542	.014	.539	deaccented and H*
5	Syllable_DurZscore	.736	.000	.438	deaccented and H*
	Syllable_IntAvg	.057	.001	.393	deaccented and H*
	Syllable_IntMax	.063	.003	.480	deaccented and H*
	Syllable_IntMin	.542	.014	.627	deaccented and H*
	Syllable_F0avg	.863	.000	.745	deaccented and H*

6	Syllable_DurZscore	.728	.000	.508	deaccented and H*
	Syllable_IntAvg	.055	.002	.584	deaccented and H*
	Syllable_IntMax	.061	.012	.667	deaccented and H*
	Syllable_IntMin	.537	.017	.768	deaccented and H*
	Syllable_F0avg	.443	.011	.746	deaccented and H*
	Syllable_F0max	.441	.013	.838	deaccented and H*

Summary of Canonical Discriminant Functions

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	.274[a]	83.0	83.0	.464
2	.056[a]	17.0	100.0	.231

a. First 2 canonical discriminant functions were used in the analysis.

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 2	.743	112.340	12	.000
2	.947	20.715	5	.001

Structure Matrix

	Function	
	1	2
Syllable_F0avg	.713*	.192
Syllable_F0max	.674*	.578
Syllable_IntMax	.673*	.243
Syllable_IntRange[b]	.604*	-.016
Syllable_IntAvg	.592*	.118
Syllable_DurZscore	.478*	-.294
Syllable_IntMin	-.065	.230*
Syllable_F0min[b]	.138	-.184*

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions

Variables ordered by absolute size of correlation within function.

*. Largest absolute correlation between each variable and any discriminant function

b. This variable not used in the analysis.