TRENDS IN
LANGUAGE ACQUISITION RESEARCH 23

The Development of Prosody in First Language Acquisition

Edited by Pilar Prieto and Núria Esteve-Gibert

JOHN BENJAMINS PUBLISHING COMPANY

The Development of Prosody in First Language Acquisition

Trends in Language Acquisition Research

ISSN 1569-0644

TiLAR publishes monographs, edited volumes and text books on theoretical and methodological issues in the field of child language research. The focus of the series is on original research on all aspects of the scientific study of language behavior in children, linking different areas of research including linguistics, psychology & cognitive science.

For an overview of all books published in this series, please see
http://benjamins.com/catalog/tilar

Volume 23

The Development of Prosody in First Language Acquisition
Edited by Pilar Prieto and Núria Esteve-Gibert

The Development of Prosody in First Language Acquisition

Edited by

Pilar Prieto

ICREA-Universitat Pompeu Fabra

Núria Esteve-Gibert

Aix-Marseille Université, CNRS, LPL / Universitat de Barcelona

John Benjamins Publishing Company

Amsterdam / Philadelphia

The paper used in this publication meets the minimum requirements of the American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

DOI 10.1075/tilar.23

Cataloging-in-Publication Data available from Library of Congress:
LCCN 2017058997 (PRINT) / 2018000598 (E-BOOK)

ISBN 978 90 272 0059 4 (HB)
ISBN 978 90 272 6421 3 (E-BOOK)

John Benjamins Publishing Company · https://benjamins.com

Table of contents

CHAPTER 1

Introduction

An overview of research on prosodic development

Pilar Prieto[1,2] & Núria Esteve-Gibert[3,4]
[1]Institució Catalana de Recerca i Estudis Avançats (ICREA) / [2]Universitat Pompeu Fabra, Department of Translation and Language Sciences / [3]Aix Marseille Université, CNRS, LPL UMR 7309 / [4]Universitat de Barcelona

Prosody in spoken language, realized through patterns of timing, melody, and intensity, is used across languages to convey a wide range of language functions, which are crucial for both structuring information in speech and encoding an important set of semantico-pragmatic meanings. First and foremost, prosody constitutes the 'organizational structure of speech' (Beckman, 1996). We use it to separate our speech into chunks of information, thus helping our interlocutor to parse our discourse into meaningful syntactic units but also sending signals about when to take turns in our conversation. Secondly, prosody plays a key pragmatic role in conversation because it can convey a broad panoply of communicative meanings, ranging from the type of speech act (assertion, question, request, etc.), information status (given vs. new information, broad focus vs. narrow focus, contrast), belief status (or epistemic position of the speaker with respect to the information exchange), politeness, and affective states, to indexical functions such as gender, age, and the sociolectal and dialectal status of the speaker (Gussenhoven, 2004; Ladd, 2008; Nespor & Vogel, 2007; see Prieto, 2015, for a review). Finally, in many languages of the world, prosody can also encode phonological contrasts at the lexical level through stress or tonal marking.

These various organizational and semantico-pragmatic functions are manifested by means of prosodic phrasal grouping (via phrasal intonation markers), intonational prominence, and intonational modulations. While phrasal intonation markers signal the ends and beginnings of turns, sentences and words, prominence markers such as pitch accents allow listeners to pinpoint the important information in an utterance (for example, contrastive focus in English, in a sentence like *It was MARY, who painted the wall* is produced with pitch accent prominence on this element). For their part, intonational modulations typically play a vital role in the

DOI 10.1075/tilar.23.01pri

pragmatic interpretation of speech (for example, the falling-rising pitch modulations at the end of interrogative sentences can convey an incredulity meaning). As we will see, given the key nature of these three prosodic elements, at least one them will be central to every chapter in this book.

Given these tight links between prosody and syntax, pragmatics, and phonology, the study of how prosody is first acquired by young children constitutes an important and exciting field of research, particularly because it can shed considerable light on the emergence throughout childhood of linguistic, communicative, and cognitive skills in general. In fact, the development of the ability to detect and decode prosody starts in very early infancy, even before birth, and continues through childhood until early adolescence. Research on early infancy has highlighted the essential role of prosody in early word segmentation and word-semantic mapping, as well as its facilitating or 'bootstrapping' role in the decoding of syntactic and pragmatic information (Höhle, 2009, among others). In addition, the special prosodic patterns often used by caregivers when addressing infants (at least in Western cultures) – what is known as infant- or child-directed speech – have been found to facilitate infants' language learning (Thiessen, Hill, & Saffran, 2005, among others).

Given the multiple functions of prosody in communication as well as its links to so many other factors, it should be evident that achieving a truly comprehensive picture of this process must require input not only from linguistics but also fields like developmental psychology, neuroscience, speech therapy, education, and computation. Indeed, recent decades have begun to see an increasing trend for collaboration between such disciplines for this purpose, resulting in significant advances. The present book is intended to reflect these advances by providing a comprehensive, interdisciplinary, state-of-the-art overview of the multiple strains of research exploring the field of prosodic development. Each of the sixteen chapters that comprise the central part of this volume will provide a particular disciplinary perspective, with one exploring how prosody is linked with other communicative resources like gestures, for example, another looking at how it is acquired by bilinguals and special populations, and so on. In the following section we will summarize the various broad topics covered in this book, noting after each summary the specific focus of the chapters that fall within that topic.

Early sensitivity to prosody

A large number of experimental studies have shown that, from a very early age, infants are acutely sensitive to the prosodic patterns present in human speech. From around two months after birth, infants have already developed some basic knowledge about the way their mother tongue sounds, being able, for example, to discriminate among languages which differ in their rhythm and intonation patterns. Together with these rhythm discrimination abilities, infants are also able to distinguish between words with different patterns of lexical stress, or distinguish function words (e.g., *it*, *this*) from content words (e.g., *baby*, *happy*) on the basis of their acoustic characteristics. This early capacity to perceive prosodic cues in language lays the foundations for later language learning in domains such as word segmentation, word mapping, and syntactic development (see Gervain & Mehler, 2010, for a review). For example, 6-month-olds can map an auditory word onto a visual referent only when the words are aligned with phrasal prosodic constituents, a universal structural property of natural languages (see Jusczyk, 1999; Shukla, White, & Aslin, 2011).

Four chapters in this book deal with the effects of infants' perception of prosody as a bootstrapping mechanism for these milestone abilities in language development. In **Chapter 2**, "Early perception of phrasal prosody and its role in syntactic and lexical acquisition", de Carvalho, Dautriche, Millotte, and Christophe summarize evidence for newborns' early sensitivity to prosodic structure and how they use it for word segmentation in continuous speech between 6 and 12 months of age. By 18 months, toddlers can readily exploit the prosodic structure of an utterance to access some of its syntactic constituents. Together with function words, this allows them to build a partial syntactic representation which can, in some cases, help them to assign a syntactic category (e.g., noun vs. verb) to a novel word and thus to distinguish objects and actions from speech. **Chapter 3**, "Early sensitivity and acquisition of prosodic patterns at the lexical level", by Bhatara, Boll-Avetisyan, Höhle, and Nazzi, focuses on infants' early preferences for particular stress patterns. The authors show that 9-month-old infants can discriminate between different patterns of lexical stress, lexical pitch contours and lexical tones if these different patterns are relevant and contrastive in their ambient language. Between 6 and 9 months of age, infants already prefer the lexical stress pattern of their ambient language, and they seem to consider duration, pitch, and intensity as a whole to build their preferences. A section on bilingual infants is also included, showing that their developmental patterns are similar to those of their monolingual peers.

Infants' early sensitivity to prosody for word learning is the focus of two chapters. **Chapter 4**, "The role of prosody in early word learning: Behavioral evidence", by Thorson, provides behavioral evidence of how infants and toddlers use their early sensitivity to prosody to segment words from the speech stream, identify discourse referents, and map a word onto its meaning. The author argues for an integrative approach to word learning, in which prosodic cues work together with social and cognitive strategies to help infants learn words. **Chapter 5**, "The role of prosody in early speech segmentation and word-referent mapping: Electrophysiological evidence," by Teixidó, François, Bosch, and Männel, approaches the same issue but from a neural processing point of view. After providing a brief introduction to the technique of event-related potentials (ERP) in language acquisition, the authors review electrophysiological studies showing the key role of prosodic cues in young infants' abilities to segment words and map words to meaning. The authors include one section on what ERP studies have shown about the role of infant-directed speech and also, interestingly, the special melodic properties of sung speech in infants' early word segmentation processes.

Infants' early sensitivity to prosodic cues as seen in early discrimination and preference patterns is thus crucial for early language learning. Young infants have been shown to be sensitive to a varied set of phrasal and lexical prosodic cues (or lexical pitch accents and lexical tones, if they are important in their ambient language), which they can then use to access lexical and syntactic information in their ambient language. In this fashion, prosodic cues serve to bootstrap language acquisition, since it is precisely the ability to perceive the prosodic patterns of their target language(s) that seems to trigger the infants' learning of other linguistic domains.

Learning to produce prosody

The early sensitivity to prosody seen in infants contrasts sharply with their very limited production capabilities. It is only between 6 and 10 months that children start to produce what is called 'canonical babbling', an early type of syllabic babbling which resembles consonant and vowel combinations in the ambient language. During the babbling stage, infants are also in the process of developing their motor skills, and interesting parallel development can be seen between babies' repetitions of sounds and their arm movements. The potential connection between these two phenomena is discussed in **Chapter 6**, "Set in time: temporal coordination of prosody and gesture in the development of spoken language production", by Rusiewicz and Esteve-Gibert. This chapter shows the strong link between prosodic and visual/gestural prominences in both adult speech and infant development. The authors review the literature that describes the synchronization between speech and motor activities of young infants before one year of age, and discuss how babbling and meaningless arm movements begin to give way to more temporally and linguistically controlled vocal and motor behaviors. For example, almost as soon as they begin to produce words, children start to properly align the prominent parts of their gestures with the prosodically prominent parts of speech.

Interestingly, the authors include a final section where they address the implications of this research for children with speech disorders.

Evidence of the lag between the development of perception and production abilities also comes from studies of rhythm development in children. There is consensus in the literature that, although language-specific perceptual sensitivity to rhythm emerges very early in infancy, the ability to produce language-specific rhythmic patterning emerges around the time children have achieved a four-word vocabulary (around age one) and very gradually develops during the second year of life. For example, a longitudinal examination of rhythmic production in young infants from the four-word point to the 25-word point (i.e., between 12 and 17 months) in typologically diverse languages found that the duration patterns of syllables differed in ways that were consistent with the language-specific input (De Paolis, Vihman, & Kunnari, 2008). In **Chapter 7**, "Speech rhythm in development: What is the child acquiring?", a comprehensive review of perception and production studies related to the production of prosodic rhythm, Payne and Post discuss how the varying typology of rhythm across languages affects acquisition by children at different ages, as well as the effect of acquiring multiple linguistic systems as seen in bilinguals. They highlight the multiple set of language-specific factors that need to be taken into account to explain the different developmental paths reported for speech rhythm crosslinguistically.

Another area which shows the delay between perception and production is **intonation**. The early ability that infants show with respect to distinguishing pitch contrasts (for example, they can discriminate between rising and falling intonation patterns between 4 and 9 months of age and several different tonal shapes in the case of tonal languages; Frota, Butler, & Vigário, 2014; Yeung, Chen, & Werker, 2013) sets the scene for the production of stress contrasts and basic pitch contrasts around the end of their first year. In **Chapter 8**, "Early development of intonation: perception and production", Frota and Butler summarize findings related to the development of language-specific intonational phonology from both perception and production points of view. On the perception side, they describe evidence showing when children begin to be sensitive to the inventory of pitch accents and boundary tones (together with the tonal alignment and tonal scaling patterns) that are present in their ambient language. Language-specific effects emerge very early on here, suggesting that pitch perception is not simply guided by domain-general abilities but is subject to underlying cross-linguistic differences. On the production side, crosslinguistic evidence carried out within the Autosegmental Metrical model shows that around the end of the first year infants begin to produce language-specific intonational patterns which develop quickly as they learn to map pragmatic meanings onto prosodic features during their second year.

Infants' first words usually appear side by side with babbling, these first words being sequences of sounds which approximate the phonetic forms of adult words. A great deal of research has concentrated on analyzing the **prosodic forms** of the **early words and phrases** produced by children, and determining whether we can predict these early speech forms by the prosodic characteristics of the ambient language as well as the frequency with which these structures appear in the input. This book contains three chapters that complement each other by covering different aspects of early word and phrase production. Nevertheless, all of them highlight the fact that **prosodic phonology** is now the preferred framework for describing the development of not only early words but also early phrases and even higher levels of prosodic structure like phonological and intonational phrases. In terms of prosodic constituents, the prosodic structure of an utterance is hierarchical, in that all sentences contain intonational phrases, intonational phrases contain phonological phrases, and phonological phrases contain prosodic words which in turn contain syllables (Nespor & Vogel, 1986). The three chapters review what we know so far about how infants acquire the prosodic structure of their ambient language, offering complementary views of the nature of early words and phrases, how they vary cross-linguistically and the mechanisms that can explain their forms and order of emergence.

Chapter 9, "Prosodic phonology in acquisition: A focus on children's early word productions", by Kehoe, presents an overview of the body of research carried out on early word production within the framework of prosodic phonology. It begins with an introduction to the background linguistic theory and then moves to a discussion of major early accounts of prosodic development, nonprosodic factors, and more recent findings from a wide range of languages. In order to account for children's early speech patterns, while earlier approaches focused almost exclusively on the constraints imposed by prosodic structure, more recent approaches have also taken into account segmental factors, as well as frequency, articulatory and perceptual bias effects. Kehoe argues that it is essential to integrate all these factors in order to support a cohesive framework of cross-linguistic prosodic development.

In **Chapter 10**, "The development of prosodic structure: A usage-based approach", Vihman proposes that the notion that children have an innate knowledge of universal principles of prosodic structure is of little help in accounting for the structures found in the early word forms that they produce. Vihman proposes an alternative view of the development of word prosodic structures, which is supported by her quantitative analysis of early child word patterns in four typologically varied languages. Her claim is that the origins of children's initial word representations can be traced back to three types of learning mechanisms that increasingly interact over the course of the first years of life: (a) experience of salient/prominent as well as frequent elements of the input speech stream; (b) constraints imposed by the neurophysiology of vocal production; and (c) memory processes that relate the acoustic patterns children hear in their target language to existing knowledge based on production.

Chapter 11, by Demuth, is entitled "Understanding the development of prosodic words: The role of the lexicon". Using evidence from cross-linguistic studies, Demuth argues that the prosodic shape and structure of early child productions is heavily influenced by the characteristics of the lexicon of the ambient language. The Chapter also includes an analysis of higher levels of prosodic structure and highlights the prosodic restrictions on the development of grammatical morphemes, articles, and determiners in children's early productions. The chapter ends by discussing the implications of these findings for the development of higher levels of prosodic competence, as well as the need to develop a model of speech planning and production.

Moving to meaning and interaction: Prosody and pragmatic development

In contrast with the wealth of research exploring the early perception of prosodic cues by children, and how they use these cues to support their acquisition of lexical (including both word segmentation and word-meaning mapping) and syntactic skills, research on the relationship between prosody and the acquisition of sociopragmatic meanings is still sparse. As noted above, an important function of prosody in language is that of marking pragmatic information ranging from emotional meanings to epistemic meanings, something which is crucial for communicative success in social interactions. Infant's first vocalizations can be uttered with prosodic patterns which serve many functions, such as to express comfort or discontent, to make a request or statement, or to express surprise (Esteve-Gibert & Prieto, 2013). Prosody is also an important cue to mark the informational status of an element in the discourse (new vs. given, all new vs. partially new, or contrastive) and the timing of social exchanges, especially in signalling turn-taking in conversation. In essence, the acquisition of the pragmatic uses of prosody, also called prosody-pragmatics mapping, has not been investigated in detail, especially in the earliest stages of language development.

During the second half of their first year of life, infants become able to engage in joint attentional frames where infant and caregiver both attend to the same external object or event and are mutually aware that they are doing so. In these contexts, infants begin to communicate intentionally using vocalizations, pointing gestures or a combination of these two modes. In **Chapter 12**, "Early development of the prosody-meaning interface", Esteve-Gibert and Prieto review the evidence on how infants aged between 9 and 12 months begin to map the prosodic patterns of others' speech with their intended speech acts. Adults use different prosodic and gesture patterns accompanying actions to signal their speech act motivation, and one-year-old children can understand the main intentionality of an adult's speech act (e.g., a requestive speech act means they are asking for an object, an expressive speech act means they want to share their interest in an object, and a declarative speech act means they are offering information about some feature of an object). Likewise, infants start to signal their intended meaning by controlling the prosodic patterns of their own speech.

Newborns have a strong interest in social stimuli (such as faces, eye contact, etc.), a 'social sense' (Kovács, Téglás, & Endress, 2010) that sets the stage for the development of communicative and interaction skills. Towards the sixth or eighth week of life, babies often produce their first smiles, begin to interact face-to-face with adults for longer periods and start to exchange emotions with others. Research has shown that vocal cues are key for young infants to recognize the emotions of other people in the first half year of life. It has been demonstrated, for example, that 5-month-olds (but not 3.5-month-olds) can detect, distinguish among, and match the facial and vocal affective displays of other infants (Vaillant-Molina, Bahrick, & Flom, 2013).

Three chapters in the volume deal with the development of prosody for the expression and comprehension of affect, reviewing different ages and levels of affect information. Though their primary focus is on intentionality, in Chapter 12, Esteve-Gibert and Prieto also review studies of early infancy and show that, before the emergence of lexical and grammatical skills, infants can 'read' the emotional states of other people by processing the prosodic cues in a person's speech, and that they learn to control prosody in their own production in order to express their own emotional states. **Chapter 13**, "Gradual development of focus prosody and affect prosody comprehension: A proposal for a holistic approach", by Ito, and **Chapter 14**, "Children's development of internal state prosody", by Armstrong and Hübscher, show that although pre-lexical infants can use and process prosody as a tool to mark affect, they are still not adult-like in this respect and this ability continues to develop in later stages of language acquisition. Both chapters review the literature on preschool and school-age children and stress the fact that the study of children's use of prosody to comprehend affective states cannot be isolated from the study of how infants use prosody to comprehend other pragmatic meanings that involve complex cognitive skills. In Chapter 13, Ito relates children's comprehension of affect meaning with their ability to comprehend other uses of prosodic prominence like focus marking. In Chapter 14, Armstrong and Hübscher relate emotional uses of prosody to the expression of other internal states, like belief states.

As we have noted, prosody – and specifically prosodic prominence – is used in many languages of the world to encode information status, that is, whether some piece of information is old/new in the discourse or important/unimportant, and even whether it should be assigned a contrastive interpretation (see Speer & Ito, 2009, for a review). When we assign focus prosody to a given word within a sentence (e.g., *Some BOYS came* vs. *SOME boys came*), we are guiding the hearer towards the central focus of the utterance's message. Thus, prosodic prominence is a strong perceptual highlighter for the location of critical information in speech. Two chapters in this volume look at how children understand and then use prosody to convey information status. In Chapter 13, Ito reviews current research on how infants develop an ability to comprehend prosodic focus and highlights the fact that prosodic prominence is used to mark not only information status across languages but also attitudinal and affective information. She calls for an integrative approach that considers the interaction between affect and focus prosody as a future direction for research on prosodic development within and across individuals. In **Chapter 15**, "Get the focus right: Acquisition of prosodic focus-marking in production", Chen reviews the sharp cross-linguistic differences in the ways that children use prosody (and morphosyntax) to mark focus, and then traces the different paths followed by children as their ability to produce focus develops. She presents results of studies that analyzed the production of broad and narrow focus as well as contrastive focus by children 2 to 11 years of age who were learning to speak typologically distinct first languages, primarily Dutch, Seoul Korean, Swedish, and Mandarin, but also including English, German, and Finnish. The author proposes a cross-linguistic theory for the acquisition of prosodic focus according to which the specific means of marking focus in a particular language and the lexical relevance of pitch will have a predictable impact on the age at which children are able to produce adult-like focus prosody and the order in which different prosodic means are acquired.

Prosody (and in particular intonation pitch contours) can be a strong marker of speaker belief and epistemic states across languages. Speakers use prosody (often in combination with visual cues like facial cues or body gestures) to express their (un)certainty, (dis)belief, or (non)surprise about the propositional content of an utterance. Little work has been carried out on how children develop the ability to comprehend these epistemic meanings in others' speech, and even less on how they learn to use prosody to express such epistemic meanings themselves. In Chapter 14, Armstrong and Hübscher review the evidence available so far in this regard, which shows how children's mastery of prosody for the expression of belief states interacts with their use of lexical (e.g., modal verbs), morphosyntactic (e.g., adverbs in specific constructions), or non-verbal (e.g. facial cues) cues, their acquisition of these skills being tightly linked to their overall development of their understanding of other's minds as measured by Theory of Mind tasks.

Prosody in bilingualism and in specific populations

Most of the world population speaks more than one language (Grosjean, 2010), so most children in fact have to learn more than one prosodic system. Several chapters of the volume include specific sections on the developmental patterns of bilingual children. In general, bilingual infants seem to be able to extract, generalize, and produce the specific prosodic properties of their ambient languages at a similar age to monolingual infants, although some cues like rhythmic metrics may take them a bit longer to acquire, and their dominant language will play a larger role than their less dominant language in the acquisition process. **Chapter 16**, "Bilingual children's prosodic development", by Lleó, looks more closely at a bilingual population. The author presents data on how the different domains of prosodic structure (syllables, metrical feet, stress, intonation and rhythm) develop in bilingual children, revealing that although the two systems are acquired independently, there are cross-linguistic interactions between the systems.

Research describing the development of prosodic skills in children with communication disorders has also progressed. In **Chapter 17**, "Prosodic development in atypical populations", Peppé provides an overview of how conditions such as autism spectrum disorders, hearing impairment, Down syndrome, Williams syndrome, and disfluency disorders can affect prosodic abilities in terms of both perception and expression. She concludes the chapter with a discussion of recent intervention programs based on prosody.

Directions for future research

The chapters included in this book bring us a step forward in our understanding of prosodic development in children. One inescapable conclusion to be drawn from all this work is that prosodic development is closely intertwined with many other systems of language as well with as the acquisition of social cognition skills and pragmatic knowledge. Mastering the comprehension and use of prosody is therefore vitally important for an infant's development of later linguistic abilities and general communication skills. Given the complexity of this process as well as its importance, it is not surprising that a full understanding of prosodic development in children constitutes an ongoing challenge for not only linguists but also developmental psychologists, neuroscientists and speech therapists.

Though the multidisciplinary research reflected in this volume has clearly yielded a significant body of essential information regarding the acquisition of prosody, we still lack a comprehensive and widely accepted theory of prosodic development. We need a theory that can not only encompass and explain perception and production patterns–which have traditionally been studied separately–but also take into account the complex relationships between prosodic abilities and other linguistic, communicative, and cognitive skills. Given the range of fields involved in such an endeavor, this overarching goal calls for a significantly high level of interdisciplinary awareness. There are also methodological challenges ahead, including the need to find more ecologically valid research methods that combine experimental and computational methods in future studies with children. For example, for both perception and comprehension, behavioral data should be complemented by ERP and fMRI studies for a fuller picture of how an infant's brain processes prosodic features. On the other hand, recent technological developments have greatly facilitated data collection, leading to the creation of freely accessible, large-scale audio and video corpora from different languages, such as the PhonBank initiative (Rose & MacWhinney, 2014), which constitute a potential goldmine of information on the development of prosodic production. Similarly, acoustic/phonetic tools such as Praat (Boersma & Weenink, 2017) have had a profound impact on our ability to measure and analyze prosodic data. This combination of high quality recorded corpora and a tool to automatically code acoustic cues has proved invaluable to research and must be further exploited, for it has huge potential to yield important results.

The chapters in this volume have outlined a developmental path along which a set of prosodic abilities are acquired at different stages in a child's development. For example, a child's initial abilities related to rhythm discrimination give way to later prosodic skills related to word segmentation and syntactic grouping, and later still, the ability to comprehend epistemic information in the speech of others. However, there remains a need to investigate whether certain prosodic abilities at one stage can act as precursors and even predictors of later linguistic abilities. While the role of prosody as a syntactic bootstrapper has been explored (see de Carvalho, Dautriche, Millotte, & Christophe, this volume), the role of prosody as a pragmatic bootstrapper is still neglected (yet see a recent proposal by Hübscher, Esteve-Gibert, Igualada, & Prieto, 2017). Interestingly, researchers interested in gesture development have found evidence that the use of gesture by children precedes and predicts changes in language acquisition and learning (see Goldin-Meadow & Alibali, 2013, for a review). Given the strong semantic co-dependency between co-speech gestures and pragmatic prosody in languages, we would expect the learning of prosody to run in tandem with language learning.

Though sometimes neglected, prosody is a robust cue for the conveyance of essential pragmatic information in communication exchanges. An important line of investigation that needs to be strengthened is research into the relationship between prosodic development and pragmatic development. Many questions remain unanswered about the role of prosody in first language acquisition, including the mapping between prosody and meaning, and the relationship between the development of prosody and a child's social cognition skills and pragmatic knowledge (see Stephens & Matthews, 2014). Though some of the chapters in this book have addressed the question of how prosody develops in relation to the understanding and expression of emotional information, as well as speech act, focus, and belief state information, much work remains to be carried out in this area. One of the more important challenges ahead is to unravel how children acquire pragmatic meanings through prosody using an integrative approach that considers the interaction between the multidimensional meanings encoded by prosody (see Ito, this volume), as well as to relate what we know about the development of prosody to other linguistic and sociocognitive abilities in children.

Another line of research which warrants further attention is the area of cross-linguistic comparisons, as well as the interaction of prosody with other types of grammatical domains. Data from these two directions could work together to clarify which patterns of prosodic development are universal and which are language-specific. As several authors in this volume point out, we still need to have a full picture of prosodic development across languages (e.g., Demuth, this volume; Frota & Butler, this volume; Chen, this volume). Only by means of cross-linguistic findings we will be able to discover the forces that, despite individual differences and language-specific effects, guide infants in their path to language acquisition.

As a final remark, the research reviewed in this book has important practical and educational implications which merit further exploration. Given that behavioral studies show that prosodic skills can be precursors and predictors of later language abilities as well as predictors of communication disorders, prosodic cues could potentially be put to use as tools to not only facilitate language learning on the one hand but also diagnose and treat communication disorders on the other. For example, a lack of sensitivity to pitch or durational contrasts, or emotional marking in speech, might be a precocious warning of later language delays. This suggests the need to establish standard profiles of prosodic development (and therefore universally and easily applicable prosodic assessment tools), which can be matched to profiles of pragmatic and sociocognitive development to help us identify children who show signs of non-normative development. By the same token, there is a need to assess current intervention paradigms which include prosody (see Peppé, this volume) and see how much farther this approach can take us. In general, we need to foster closer collaboration with researchers who are interested in applying this research to new educational and speech therapy practices and strategies.

All in all, we hope that this volume will be of interest to both experienced and novice researchers and will help delineate a road map for a more comprehensive theory of prosodic development.

Acknowledgements

All chapters were reviewed by at least one external reviewer (sometimes two) and one internal reviewer. Our external reviewers were Lluïsa Astruc, Laura De Ruiter, Marion Dohen, Marisa Filipe, Caroline Floccia, Ulrike Gut, Alfonso Igualada, Brian MacWhinney, Danielle Matthews, Mits Ota, Ferran Pons, Yvan Rose, Elena Sakkalou, Shari Speer, Marc Swerts, Maria del Mar Vanrell, and Laurence White, and our internal reviewers were Meghan Armstrong, Aoju Chen, Katherine Demuth, Kiwako Ito, Margaret Kehoe, Sue Peppé, Brechtje Post, and Heather Leavy Rusiewicz. We are profoundly indebted to all reviewers for the time they devoted to providing feedback that has made this a vastly better volume. We would also like to thank Paolo Roseano, who provided key assistance with the final formatting of the materials and the creation of the index.

This work was supported by a grant awarded by the Spanish Ministry of Science and Innovation (FFI2015-66533-P, "Intonational and gestural meaning in language") and a grant awarded by the Generalitat de Catalunya (2014 SGR 00925) to the Prosodic Studies Group. The second author also acknowledges a grant awarded by the Labex BLRI (ANR-11-LABX-0036) under the project title Investments of the Future A*MIDEX (ANR-11-IDEX-0001-02), a postdoctoral grant awarded by the Spanish Ministry of Economy, Industry, and Competitivity (FJCI-2015-26845), and a financial contribution from the Laboratoire Parole et Langage, Aix-Marseille Université and CNRS, France, to the publication of this book.

References

Beckman, M. E. (1996). The parsing of prosody. *Language and Cognitive Processes*, 11, 17–67. doi: 10.1080/016909696387213

Boersma, Paul & Weenink, David (2017). Praat: doing phonetics by computer [Computer program]. Version 6.0.36, retrieved 11 November 2017 from http://www.praat.org/

De Paolis, R. A., Vihman, M. M., & Kunnari, S. (2008). Prosody in production at the onset of word use: A cross-linguistic study. *Journal of Phonetics*, 36, 406–422. doi: 10.1016/j.wocn.2008.01.003

Esteve-Gibert, N. & Prieto, P. (2013) Prosody signals the emergence of intentional communication in the first year of life: Evidence from Catalan-babbling infants. *Journal of Child Language*, 40(5), 919–944.

Frota, S., Butler, J., & Vigário, M. (2014). Infants' perception of intonation: Is it a statement or a question? *Infancy*, 19(2), 194–213. doi: 10.1111/infa.12037

Gervain, J., & Mehler, J. (2010). Speech perception and language acquisition in the first year of life. *Annual Review of Psychology*, 61, 191–218. doi: 10.1146/annurev.psych.093008.100408

Goldin-Meadow, S., & Alibali. M. W. (2013). Gesture's role in speaking, learning, and creating language. *Annual Review of Psychology*, 64, 257–283. doi: 10.1146/annurev-psych-113011-143802

Grosjean, F. (2010). *Bilingual: Life and reality*. Cambridge: Harvard University Press.

Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511616983

Höhle, B. (2009). Bootstrapping mechanisms in first language acquisition. *Linguistics*, 47: 359–382. doi: 10.1515/LING.2009.013

Hübscher, I., Esteve-Gibert, N., Igualada, A., & Prieto, P. (2017). Intonation and gesture as bootstrapping devices in speaker uncertainty. *First Language*, 37 (1): 24–41.

Jusczyk, P. W. (1999). How infants begin to extract words from speech. *Trends in Cognitive Sciences*, 3(9), 323–328. doi: 10.1016/S1364-6613(99)01363-7

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science*, 330(6012), 1830–1834. doi: 10.1126/science.1190792

Ladd, D. R. (2008). *Intonational phonology* (2nd ed.). Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511808814

Nespor, M., & Vogel, I. (2007). *Prosodic phonology* (2nd ed.). Dordrecht: Walter de Gruyter. doi: 10.1515/9783110977790

Prieto, P. (2015). Intonational meaning. *Wiley Interdisciplinary Reviews: Cognitive Science*, 6, 371–381.

Rose, Y., & MacWhinney, B. (2014). The PhonBank project: Data and software-assisted methods for the study of phonology and phonological development. In J. Durand, U. Gut, & G. Kristoffersen (Eds.), *The Oxford handbook of corpus phonology* (pp. 380–401). Oxford: Oxford University Press.

Shukla, M., White, K. S., & Aslin, R. N. (2011). Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-month-old infants. *Proceedings of the National Academy of Sciences*, 108(15), 6038–6043. doi: 10.1073/pnas.1017617108

Speer, S. R. & Ito, K. (2009). Prosody in first language acquisition: Acquiring intonation as a tool to organize information in a conversation. *Language and Linguistic Compass*, 3, 90–110. doi: 10.1111/j.1749-818X.2008.00103.x

Stephens, G. & Matthews, D. (2014). The communicative infant from 0–18 months: The social-cognitive foundations of pragmatic development. In D. Matthews (Ed.), *Pragmatic development in first language acquisition* (pp. 13–36). Amsterdam: John Benjamins.

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, 7(1), 53–71. doi: 10.1207/s15327078in0701_5

Vaillant-Molina, M., Bahrick, L. E., & Flom, R. (2013). Young infants match facial and vocal emotional expressions of other infants. *Infancy*, 18(1), E97–E111. doi: 10.1111/infa.12017

Yeung, H. H., Chen, K. H., & Werker, J. F. (2013). When does native language input affect phonetic perception? The precocious case of lexical tone. *Journal of Memory and Language*, 68(2), 123–139. doi: 10.1016/j.jml.2012.09.004

PART 1

Early sensitivity to prosody

CHAPTER 2

Early perception of phrasal prosody and its role in syntactic and lexical acquisition

Alex de Carvalho[1], Isabelle Dautriche[2], Séverine Millotte[3] & Anne Christophe[1]
[1]Laboratoire de Sciences Cognitives et Psycholinguistique (ENS, EHESS, CNRS), Département d'Études Cognitives (École Normale Supérieure – PSL Research University) / [2]Centre for Language Evolution, School of Philosophy, Psychology and Language Sciences, University of Edimburgh / [3]Laboratoire d'Étude de l'Apprentissage et du Développement, CNRS UMR 5022, UB, Université Bourgogne Franche-Comté

This chapter will review empirical findings on the perception of phrasal prosody in very young infants, and how it develops in first language acquisition. The ability to process phrasal prosody impacts learning of important aspects of language, specifically word segmentation and syntactic parsing. We will see that infants are able to perceive crucial aspects of phrasal prosody before the end of their first year of life, and that a few months later they are able to exploit the prosodic structure of an utterance to constrain its syntactic analysis, and therefore, to infer the meaning of unknown words.

Introduction

Infants acquiring language face the challenging task of having to learn by extracting information directly from the speech stream. Given the complexity of this spoken input, it is impressive that around the age of two or three years, toddlers already show a deep knowledge about aspects of the phonology, semantics and syntax of their native language. How were they able to build this knowledge during the first steps of language acquisition? In this chapter we will review empirical findings on the perception and use of a potential cue that is directly accessible from the speech stream: phrasal prosody. Phrasal prosody reflects the organization of sentences into prosodic constituents above the word level, and it is conveyed by variations in pitch, duration, and energy, of speech sounds over the course of an utterance.

DOI 10.1075/tilar.23.02car

In all the languages of the world, the words and syllables composing the speech stream are not produced in a monotonous way. Rather, speech is characterized by rhythmic and intonational variations over the course of an utterance: its prosody. The prosody of a language can be described at different levels (i.e., syllables, words, phrases, and whole utterances) and impacts linguistic interpretations in various ways. For example, prosodic information can be used to convey lexical meaning at the word level through variations in stress pattern or lexical tone (e.g. 'beBE' and 'BEbe' are two different words in Spanish, differing only in their stress patterns). Prosody is also used to mark whether a sentence is declarative or interrogative (e.g. Zhou, Crain, & Zhan, 2012) and it conveys useful discourse information such as information structure (e.g. focus, new vs. old information; Hirschberg & Pierrehumbert, 1986; Pierrehumbert & Hirschberg, 1990). Prosody can also be used to change the interpretation of an utterance (e.g. irony, disbelief, etc.), and it can even reflect the emotional state of the speaker (Armstrong, Andreu, Esteve-Gibert, & Prieto, 2016; Jun, 2005, 2014; Ladd, 2008). As for phrasal prosody, the prosodic structure of an utterance can be described hierarchically, in the sense that utterances contain one or more intonational phrases, intonational phrases contain one or more phonological phrases, and phonological phrases contain one or more prosodic words, which in turn contain syllables (e.g. Nespor & Vogel, 1986). In this chapter, we will focus on the role of phrasal prosody on speech segmentation and parsing, and its impact on syntactic and lexical acquisition.

Our interest in this topic stems from the fact that the prosodic boundaries at the phrasal level (i.e., phonological and intonational phrases) always coincide with the boundaries between syntactic constituents in a sentence (Nespor & Vogel, 1986). Thus, the prosody of an utterance can reflect aspects of the syntactic structure of a sentence and provide important cues for parsing (e.g., Morgan & Demuth, 1996; Nespor & Vogel, 1986). For example, a sentence like: "The little boys run really fast", tends to be produced as: [the little boys] [run really fast] (brackets indicate phonological phrases), where the three words "the little boys" are grouped into one single prosodic unit, corresponding to the noun phrase in this sentence, and "run really fast" are grouped into a second prosodic unit, corresponding to the verb phrase. Although it is not the case that all syntactic constituent boundaries are marked by a prosodic boundary, whenever a prosodic boundary is perceived in the speech stream, it always coincides with a syntactic constituent boundary (Nespor & Vogel, 1986). As a result, salient prosodically-conditioned acoustic information such as phrase-final lengthening, pitch variations, and pauses, may allow listeners to identify prosodic boundaries which in turn might be useful to segment the continuous speech stream into relevant units, such as words and syntactic constituents. In other words, phrasal prosody might be particularly important for young listeners, who still do not know much about their native language, because it could represent an important tool to parse sentences into meaningful constituents and to identify some of the syntactic constituents of a sentence even before infants have acquired an extensive vocabulary (Christophe, Millotte, Bernal, & Lidz, 2008; Morgan, 1986; Morgan & Demuth, 1996). In this chapter we will review empirical findings supporting this hypothesis.

The first section will show that infants are sensitive to prosodic information from birth onwards and that this cue can help them to learn important aspects of their native language. The second section will present empirical findings showing that infants can rely on phrasal prosody to segment the speech stream into words and constrain lexical access. Finally, in the third section we will present recent findings suggesting that infants can exploit the relationship between the prosodic and syntactic structures of sentences to constrain their syntactic analysis – which in turn would allow them to constrain the acquisition of word meanings.

Early perception of prosodic cues

Considerable work has shown that infants have extensive experience with prosody from their first days of life (Christophe, Mehler, & Sebastián-Gallés, 2001; Decasper & Spence, 1986; Mehler et al., 1988; Nazzi, Bertoncini, & Mehler, 1998; Shi, Werker, & Morgan, 1999). An important study conducted by Mehler et al. (1988) showed that four-day-old infants are already sensitive to prosodic information when listening to sentences. Testing French newborns with French and Russian, and American infants with English and Italian, the authors observed that infants showed a preference for listening to their native language over a foreign one. Given that a few days after birth, infants have received very little postnatal language input, it is possible that prenatal exposure plays a role in this early preference for native speech (e.g., Decasper & Spence, 1986). Indeed, some prosodic characteristics from the speech stream, such as rhythm, stress and intonation, pass through the skin and uterus to the fetus (Decasper & Spence, 1986). To test whether newborns' preference arises from an early knowledge of the prosodic characteristics of their native language, Mehler et al. (1988) tested another group of French and American infants using low-pass filtered speech samples, where the prosodic information remained intact while phonetic information was stripped away. Their results showed that French newborns preferred to listen to low-pass filtered French speech over filtered Russian speech and American infants showed the same pattern for filtered English vs. filtered Italian speech. These results suggest that from birth onwards, infants are already sensitive to prosodic information, which helps them to distinguish their native language from a foreign language.

In a subsequent study, Nazzi et al. (1998), used low-pass filtered sentences in foreign languages to test infants' sensitivity to prosody to distinguish between non-native languages. They showed that French newborns were also able to discriminate foreign languages based on their prosodic patterns. For example, French newborns were able to discriminate stress-timed English from mora-timed Japanese, but failed to discriminate between English and Dutch because these languages share the same rhythmic properties (both stress-timed languages). Taken together, these results show not only that infants can use prosodic cues to discriminate their native language from a foreign one but also that they can use prosody to categorize languages based on their rhythmic and intonational properties.

Subsequent studies provided further evidence about the role played by prosody during the first steps of language acquisition, showing in particular that infants use it to segment the continuous speech stream into chunks. One of the earliest infant studies on continuous speech perception showed that 7-month-old infants are able to rely on prosodic information to recognize clauses in the speech stream: Hirsh-Pasek, Kemler Nelson, Jusczyk, Cassidy, Druss, and Kennedy (1987) showed that 7-to-10-month-olds are sensitive to the coherence of intonational phrases in the speech stream. Inserting a one-second pause either at clause boundaries or at within-clause locations in the speech stream, the authors observed that infants prefer to listen to speech containing pauses at clause boundaries (i.e., "*Cinderella lived in a great big house*-PAUSE-*but it was sort of dark*-PAUSE-…") than within clauses (i.e., "*Cinderella lived in a great big house, but it was*-PAUSE-*sort of dark because she had*-PAUSE-…"; see also Männel & Friederici, 2009, for similar results at 5 months with EEG and Teixidó, François, Bosch, and Männel, this volume, for a review of the role of prosody in early speech segmentation). Subsequent studies showed that infants are sensitive to smaller prosodic units, phonological phrases, that may correspond to noun phrases or parts of verb phrases in a sentence, from about the age of 6 months (e.g. Gerken, Jusczyk, & Mandel, 1994; Johnson & Seidl, 2008; Soderstrom, Seidl, Kemler-Nelson, & Jusczyk, 2003). Moreover, infants show better memory for units from the speech stream that correspond to whole prosodic units than for chunks of speech that span prosodic boundaries (Mandel, Jusczyk, & Nelson, 1994; Nazzi, Iakimova, Bertoncini, Frédonie, & Alcantara, 2006). This shows that infants are not only sensitive to prosodic grouping information, but that they recognize the prosodic well-formedness of speech chunks, and find it harder to process chunks of speech that are not well-formed prosodically.

It is important to note that some of the cues that mark prosodic units are the same in all the world's languages, especially for larger prosodic units such as intonational phrases. These tend to be followed by a silent pause, systematically exhibit final lengthening, and often show marked pitch excursions (e.g., a decline, or a rise for questions; note that these boundary cues are also found in other domains such as musical phrases). However, other cues may vary between languages, so that infants need to learn some language-specific properties before they can exploit those cues efficiently, which may explain why sensitivity to smaller prosodic units such as phonological phrases arises later during development. For instance, the pitch contours typical of smaller prosodic units such as phonological phrases vary between languages, which means that children need experience with their native language before they can exploit these cues as reliable boundary markers in their language. For instance, experimental work suggests that while younger infants (even across different languages) tend to rely on strong and universal prosodic markers, such as pauses, older infants can make use of more subtle prosodic cues such as pitch contours and lengthening even in the absence of pauses (Johnson & Seidl, 2008; Seidl, 2007; Wellmann, Holzgrefe, Truckenbrodt, Wartenburger, & Höhle, 2012).

Prosody has also been proposed to help infants discover the word order of their native language. In fact, many languages differ with respect to the position of the verb and its object in a sentence, or more generally, the respective position of heads and complements. Either complements tend to follow their heads, or they tend to precede them. This organization, in turn, can impact the order of function words (highly frequent functional morphemes, such as articles, auxiliaries, etc…) with respect to content words (much less frequent lexical items, such as nouns, verbs, adjectives, etc…) in a sentence. For example, in VO (Verb-Object) languages such as English and French, function words typically appear before content words and at the beginning of phrases (e.g., *Le bateau* 'The boat'; *de Paris* 'from Paris' – head-initial languages), while in OV languages like Turkish and Japanese, function words tend to appear after content words and at the end of phrases (e.g., *Tokyo kara* 'Tokyo from' – head-final languages; Dryer, 1992). In addition, the head-direction of a language also determines which element will be more prosodically prominent within phonological phrases: The first one in head-final languages (typically marked with higher pitch), and the last one in head-initial languages (typically marked with lengthening; Nespor, Shukla, van der Vijver, Avesani, Schrandolf, & Donati, 2008). Thus, infants could use both the prosodic cues and the relative position of frequent and infrequent elements to infer the basic word order of their native language (Bernard & Gervain, 2012; Christophe, Nespor, Guasti, & Van Ooyen, 2003; Gervain, Mehler, Horie, Mazuka, & Nespor, 2008; Gervain & Werker, 2013; Höhle, Weissenborn, Schmitz, & Ischebeck, 2001). For instance, Christophe et al. (2003) showed that 2-month-old infants were able to distinguish between two languages that have very similar phonology, but differ in their head-direction: French (head-initial) versus Turkish (head-final) suggesting that this kind of prosodic information might be used by young listeners to obtain information about the word order in their native language. Additionally, Gervain and colleagues (2008) showed that 8-month-olds are already sensitive to the typical position in which frequent and infrequent elements appear in their native language in order to infer the position of function and content words (e.g., Italian: VO, frequent-infrequent; or Japanese: OV, infrequent-frequent). Thus, in an artificial grammar experiment, when exposed to an unsegmented string of syllables in which some syllables were highly frequent (i.e., playing the role of function words) and others infrequent (i.e. playing the role of content words), infants segment this continuous signal such that the position of the frequent elements respects the typical order of function and content words in their native language (Italian infants prefer to have the frequent elements in initial position, and Japanese infants in final position).

Interestingly however, in the case of bilingual infants acquiring both VO and OV languages, frequency alone does not provide enough information about word order since both frequent-final and frequent-initial phrases occur in their input. Researchers have proposed that prosodic information could cue word order in this case. For instance, in a recent experiment, Gervain and Werker (2013) showed that bilingual 7-month-olds acquiring simultaneously an OV and a VO language exploit prosodic information to determine the relative order of frequent and infrequent elements in an unsegmented string of syllables. When familiarized with strings of syllables consistent with an OV prosodic pattern (with high-low-high-low pitch alternations), infants preferred to listen to chunks of syllables with the frequent elements at the end; in contrast, when familiarized with strings of syllables consistent with a VO prosodic pattern (short-long-short-long), they showed the reverse preference. These studies suggest that prosody, together with frequency information, impacts word order acquisition. Taken together, the studies reviewed in this section show that prosody is indeed an important source of information, directly accessible from the speech stream, which may promote early language learning in a variety of ways.

Using phrasal prosody to segment the speech stream into words

Discovering the words composing their language is one of the challenging tasks faced by language learners. During their development, infants have to extract the word-forms from the continuous speech stream, and associate the extracted word-forms with a possible meaning. However, given that in fluent speech there are no pauses between two consecutive words, how can infants segment the speech stream into words? Several studies have shown that infants can use a variety of cues to discover word boundaries; although none of them is sufficient on its own, together they may allow infants to discover many words in their input (see Thorson, this volume). Word segmentation cues that have been proposed and studied include phonotactic constraints (e.g., Jusczyk, 1997), words heard in isolation (e.g., Bortfeld, Morgan, Golinkoff, & Rathbun, 2005; Lew-Williams, Pelucchi, & Saffran, 2011), the statistical structure of the input (e.g., transitional probabilities; e.g., Saffran, Aslin, & Newport, 1996), word-level prosodic information such as typical word stress pattern (e.g., Cutler & Butterfield, 1992; Echols, 1993; Echols & Newport, 1992; Jusczyk, Houston, & Newsome, 1999; see Bhatara, Boll-Avetisyan, Höhle, & Nazzi, this volume, for a detailed review), as well as phrasal prosody (e.g., Gout, Christophe, & Morgan, 2004; Johnson, 2008; Shukla, Nespor, & Mehler, 2007; Shukla, White, & Aslin, 2011).

Turning our attention toward the role of phrasal prosody on segmentation into words, several studies have shown that infants are able to perceive prosodic phrase boundaries and exploit them to find the boundaries between words from 6 months onwards (Gout et al. , 2004; Johnson, 2008; Millotte et al., 2010; Shukla et al., 2011). Since prosodic units such as phonological phrases are constructed by grouping words together, whenever a prosodic boundary is perceived, it has to correspond to a word boundary (Nespor & Vogel, 1986). Thus, sensitivity to phrasal prosody might provide cues to speech segmentation and, therefore, constrain lexical search. Of course, since most prosodic units contain more than one word, many word boundaries will not be marked by phrasal prosody – word segmentation within phonological phrases will have to rely on some other cues mentioned above. As an illustration of the impact of prosodic boundaries on word segmentation (1), the sentences below both contain the two syllables *pay* and *per*, however, only the first one contains the word *paper*:

(1) a. [The college] [with the biggest *paper* forms] [is best]
    b. [The butler] [with the highest *pay*] [*per*forms the most]

In the second sentence, the prosodic boundary between the syllable 'pay' and the syllable 'per' should block lexical access to the word *paper*: indeed, it has been shown with adult listeners that prosodic boundaries constrain lexical access (Christophe, Peperkamp, Pallier, Block, & Mehler, 2004; Endress & Hauser, 2010; Warner, Otake, & Arai, 2010). To test infants' ability to use phrasal prosody to segment the speech stream into words, Gout et al. (2004) used the above sentences (1) in a conditioned head-turn procedure. In a first session, they trained American 10-month-olds to turn their heads toward a puppet whenever they heard the word *paper* (for instance). Then, during a test phase infants were exposed to full sentences, such as (a) and (b). Their results showed that infants trained to respond to the word *paper* turned their head more often when listening to *paper*-sentences (a) than to *pay#per*-sentences (b). In contrast, infants trained to respond to the target word *pay* turned equally often for both types of sentences (since the syllable *pay* was present in both sentences, the target word *pay* might have been noticed in both sentences). This results show that 10-month-old American infants can use phrasal prosody to segment the speech stream into words and, therefore, to constrain their lexical access. Further studies showed similar results with French 16-month-olds (Millotte et al., 2010) and other experiments in English confirmed these results, showing that 12-month-olds can use phrasal prosody to constrain lexical access within strings of nonsense syllables differing in their prosodic structure (Johnson, 2008). Moreover, Shukla et al. (2011) showed that 6-month-olds were able to better associate a visual referent to a novel word aligned with a prosodic phrase boundary, than to a novel word that straddled a prosodic boundary. This suggests that infants are sensitive to the fact that words are aligned with prosodic phrase boundaries, and exploit this information to facilitate word learning. Taken together these results highlight the importance of phrasal prosody for segmenting the speech stream into words and constraining lexical access and lexical acquisition.

The role of phrasal prosody for syntactic parsing in children

The empirical studies presented above show that phrasal prosody has an important role in infant speech perception, allowing them to discriminate between languages, infer aspects of the word order of their language, and segment the speech stream into words and clauses. Since the prosodic structure of an utterance partially reflects the syntactic structure of a sentence (Nespor & Vogel, 1986), the *prosodic bootstrapping* hypothesis proposes that phrasal prosody could also help infants to discover the syntactic structure of sentences (e.g., Morgan, 1986; Morgan & Demuth, 1996). As mentioned above, the relationship between prosodic structure and syntactic structure is such that prosodic boundaries are aligned with syntactic constituent boundaries. Thus, prosodic information such as phrase-final lengthening, pitch variation and pauses can help listeners to identify prosodic boundaries, and therefore to find some syntactic constituent boundaries. Taking it into account, prosodic information might facilitate on-line sentence processing in adults, and might provide a way for infants to identify some of the syntactic constituents of an utterance even before they have acquired an extensive vocabulary.

Supporting this hypothesis, many studies have shown that adults integrate phrasal prosody online to recover the syntactic structure of sentences (Kjelgaard & Speer, 1999; Michelas & D'Imperio, 2015; Millotte, René, Wales, & Christophe, 2008; Millotte, Wales, & Christophe, 2007; Snedeker & Trueswell, 2003; Weber, Grice, & Crocker, 2006). Given the extensive literature we reviewed in this chapter showing that infants have an early access to phrasal prosody, and are able to exploit it for lexical segmentation, one would naturally expect that infants might also be able to use phrasal prosody, as adults do, to constrain syntactic analysis. However several studies investigating whether preschoolers can exploit phrasal prosody to constrain their syntactic analysis have found that children have difficulties using prosody for syntactic ambiguity resolution in English (Snedeker & Trueswell, 2001, 2003; Vogel & Raimy, 2002) and in Korean (Choi & Mazuka, 2003). Most of these studies in English used sentences with a prepositional phrase attachment ambiguity, such as "Can you touch the frog with the feather?", in which the prepositional phrase "with the feather" can be interpreted either as an instrument of the verb "touch" or as a modifier of the noun "frog". In such sentences, the default prosodic structure is the same for the two possible interpretations (i.e. [Can you touch] [the frog] [with the feather] – Snedeker & Yuan, 2008), but speakers who are aware of the ambiguity can intentionally disambiguate by exaggerating one of the prosodic breaks, in order to favor one interpretation over the other, i.e. "[Can you touch the frog] [with the feather]?" for the instrument interpretation, vs. "[Can you touch] [the frog with the feather]?" for the modifier interpretation (Snedeker & Trueswell, 2003). Snedeker and Trueswell (2001) found that children failed to use this kind of prosodic information when interpreting this kind of sentences. In a subsequent experiment, which controlled for children's perseveration biases, Snedeker and Yuan (2008) observed that children succeeded in this task, when they were presented with only one kind of sentences: either modifier-only, or instrument-only. However, when they were presented with both instrument and modifier sentences across the experiment, as in the previous study, children failed to use prosody to constrain their syntactic interpretations (Snedeker & Yuan, 2008). Children's difficulty using prosody in this kind of sentences might be due to the fact that the disambiguating prosodic breaks they needed to use are not part of the normal prosodic structure of these sentences, but these cues are only produced when the speaker is consciously trying to disambiguate (Snedeker & Trueswell, 2003). Thus, children may have had difficulties using this kind of optional prosodic information.

To avoid this problem of optional prosodic disambiguation, de Carvalho and colleagues exploited locally ambiguous sentences featuring noun/verb homophones, in which the default prosodic structure differed between conditions (de Carvalho, Dautriche, & Christophe, 2016, for French; and de Carvalho, Lidz, Tieu, Bleam, & Christophe, 2016, for English). For example, the word "watch" is a verb in the sentence: [Mommies] [**watch** TV every night], but it is a noun in the sentence: [Mommy's **watch**] [ticks very noisily]. Here, brackets indicate prosodic units, which reflect the syntactic structure of each sentence. Crucially, in both cases, there is a prosodic break (marked by phrase-final lengthening and pitch change) between the subject Noun Phrase and the Verb Phrase: this break falls after the critical word when it is used as a noun, and before it when it is used as a verb.

American and French 4-year-olds presented with the beginning of these ambiguous sentences (e.g., "Mommies watch …") interpreted the ambiguous target word as a noun or as a verb depending on the prosodic structure in which it was embedded, as shown by the words they used to complete the sentences in a sentence completion task (de Carvalho, Dautriche, & Christophe, 2016; de Carvalho, Lidz, Tieu, Bleam, & Christophe, 2016). These results show that preschoolers can use prosodic boundaries to infer the presence of a syntactic constituent boundary, and in turn use that information to figure out the syntactic category of an ambiguous word. The discrepancy between these recent findings and the previous literature showing children's failure to use prosody to constrain syntactic analysis, rests on the disambiguating prosodic information used. In the recent studies, the prosodic boundary between the noun phrase and the verb phrase is part of the normal prosodic structure of sentences, and it is present even in non-ambiguous sentences (e.g. [the little frog] [eats a lot of food]).

The important question then is when do children become able to use phrasal prosody to constrain their syntactic analysis? Given that infants were shown to be sensitive to prosodic boundaries very early during development (e.g. Gerken et al., 1994; Gout et al., 2004; Johnson, 2008; Soderstrom et al., 2003), it is possible that not only preschoolers but even younger children would be able to exploit phrasal prosody to constrain syntactic analysis. De Carvalho, Dautriche, Lin, and Christophe (2017) tested two-year-old French toddlers with the sentences featuring noun/verb homophones described above (e.g., 'ferme' in French can be used as a noun in: [*la petite **ferme***] [*est très jolie*] – [the small **farm**] [is very nice] or as a verb in: [*la petite*] [***ferme** la boîte à poupée*] – [the little girl] [**closes** the toy box]). To test toddlers' ability to use prosody to disambiguate these sentences, they were presented with two images displayed side-by-side on a TV screen: one associated with the noun interpretation of the target word (e.g., a farm) and the other one with the verb interpretation (e.g., a little girl closing something). At the same time, toddlers listened to the sentence beginnings pronounced in a noun-prosody or a verb-prosody condition. Crucially, the end of the sentences was replaced with babble noise, such that only prosodic information could be used to disambiguate. The results, depicted in Figure 1, show that toddlers in the noun prosody condition switched their gaze toward the noun image around the end of the ambiguous word, while toddlers in the verb prosody condition looked more toward the verb image. Taking into account saccade preparation time (e.g., + / – 300ms: Allopenna, Magnuson, & Tanenhaus, 1998), this suggests that toddlers computed the syntactic category of the critical word online, before the word offset.

These results show that two-year-olds, upon hearing the first words of a sentence, exploit its prosodic structure to group words into constituents, and exploit this constituent structure in their on-line syntactic analysis of spoken sentences.

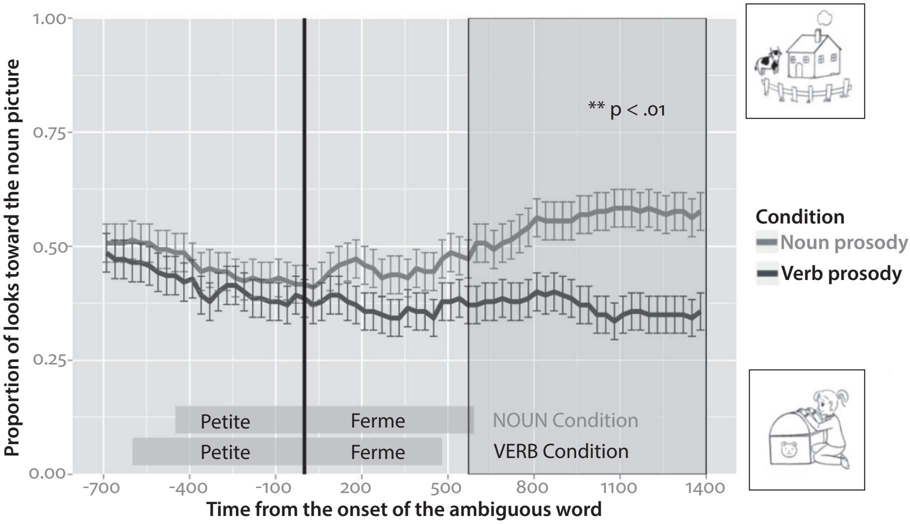


Figure 1. Proportion of looks toward the noun image, time-locked to the onset of the ambiguous word (vertical black line), for the noun prosody condition (light gray curve) and the verb prosody condition (dark gray curve). Error bars represent the standard error of the mean. A nonparametric cluster-based permutation test (Maris & Oostenveld, 2007) revealed a significant difference between the noun prosody and the verb prosody conditions, starting 600 ms after the onset of the ambiguous target word (gray time-window). Figure adapted from de Carvalho et al. (2017).

The ability to exploit information from the speech to access the syntactic structure of sentences can be extremely useful for infants during language acquisition. As proposed by Gleitman (1990), having access to the syntactic structure of sentences can help infants to discover the meaning of novel words (the *syntactic bootstrapping* hypothesis). For example, in a very recent study, He and Lidz (2017) showed that 18-month-olds listening to a sentence like "*Look, it's a doke*!" were able to infer that the novel word '*doke*' referred to an object (i.e., a penguin), but when they were listening to sentences like "*Look! It's pratching*!", they inferred that '*pratching*' referred to an action (i.e., a spinning action). In this case, the critical word, '*doke*' or '*pratch*', was preceded and/or followed by disambiguating function morphemes (*a, it's… -ing*). However, in everyday life, not all content words are immediately preceded by function words, as shown in the earlier examples with noun/verb homophones (e.g., *bears* can be either a noun or a verb, in: "The giant bears…"). In such cases, a representation in terms of syntactic constituent could be very useful. For example, *bears* is a noun in: [The giant bears]$_{NP}$ [are very hungry], and a verb in: [The giant]$_{NP}$[bears$_{v}$ a heavy load]$_{VP}$). But can infants exploit prosodic boundaries together with function words to constrain their syntactic analysis of sentences containing novel words? If so, can they use this information to determine the syntactic nature of these novel words and therefore constrain their meaning?

To investigate this question, de Carvalho, He, Lidz, and Christophe (2015) used two novel words in French (e.g., *bamoule* and *doripe*) to construct minimal pairs of sentences that differed only in their prosodic structures. For instance, the novel word *bamoule* was used as a noun in the sentence: [*Regarde la petite bamoule*]! – [*Look at the little bamoule*]!, where all the words in this sentence were grouped together into a single prosodic unit, and the novel word *doripe* was used as a verb in the sentence: [*Regarde*]! [*la petite*] [*doripe*]! – [*Look*]! [*The little one*] [*is doriping*]!, where all the words in the sentence were spread into three different prosodic units (square brackets indicate prosodic phrase boundaries). Crucially, note that when *doripe* was used as a verb there was a prosodic boundary preceding it (i.e., the boundary between the noun phrase and the verb phrase) but, when *bamoule* was used as a noun, it was embedded in a single prosodic unit together with the other words of the sentence, corresponding to the verb 'look' and the following noun phrase.

In a Habituation-Switch paradigm (e.g., Werker, Cohen, Lloyd, & Casasola, 1998), 18-month-old monolingual French-learning infants were first habituated[1] to two video stimuli showing a penguin doing two different intransitive actions (e.g., spinning, cartwheeling), one in each video. During the presentation of one of the videos (e.g., a penguin spinning), infants heard sentences in the Noun-prosody condition (e.g., [*Regarde la petite* ***bamoule***!] – Look at the little **bamoule**!, where *bamoule* was a noun, naming an object, here the penguin), and during the presentation of the other video (e.g., a penguin cartwheeling), they heard sentences in the Verb-prosody condition (e.g., [*Regarde*],[*la petite*][***doripe***!] – Look, the little$_{\text{(one)}}$ is **doriping**, where *doripe* was a verb, naming an action, here cartwheeling). Then, to test whether infants were able to correctly interpret these sentences, during the test phase, the audio tracks of the two videos were switched. Half of the participants were exposed to the Noun-Switch-condition, that is, they heard the noun sentence with 'bamoule' while seeing the cartwheeling video, and half were exposed to the Verb-Switch-condition, that is, they heard the verb sentence with "doripe" while seeing the spinning video. If infants have learned, like adults would, that "bamoule" meant "penguin" and "doripe" meant "cartwheeling", they should be surprised in the Verb-Switch condition (look longer to the

1. The habituation criterion was reached when an infant's average looking time during any block of 3 consecutive trials dropped to less than 65% of the average looking time for the most-attended block (i.e. the 3-trial block that had the longest total looking time to the video stimuli).

video), because they are listening to the "doripe" sentence while the penguin is "spinning" instead of "cartwheeling". However in the Noun-Switch condition infants should not be surprised when listening to the sentence with "bamoule", since there is still a penguin present in the video, even if it is now doing a different action.

The results of this experiment, presented in Figure 2, showed that infants' looking time toward the videos was significantly longer in the Verb-Switch condition than in the Noun-Switch condition, during the test phase. These results fit with the idea that the action change was inconsistent with their interpretation of the novel verb, while it didn't matter for their interpretation of the novel noun (a penguin was present in both videos). Thus, we observe that, from 18 months of age, infants can exploit prosodic structure to group words into constituents, and calculate the syntactic structure of sentences, which can then help them to infer the probable meaning of novel words: mapping nouns to objects and verbs to actions. Note that since both noun and verb sentences in this experiment were composed of exactly the same words in the same order (*regarde-la-petite-bamoule/doripe*), a simple analysis in terms of which words is preceding *bamoule* or *doripe* was not enough to determine the syntactic category of the novel words, since they are the same in both conditions. Remarkably, in this situation, toddlers were then able to integrate information coming from phrasal

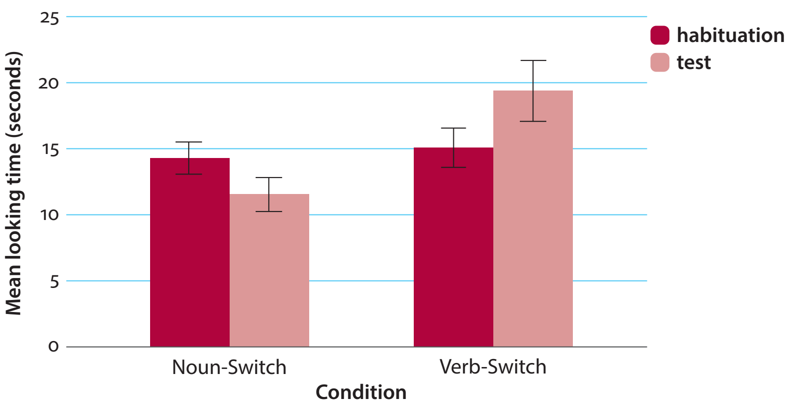


Figure 2. Mean looking time in seconds toward the videos during the last two trials of the habituation phase (in dark red) and during the two trials of the test phase (in light red) for participants assigned to the Noun-Switch Condition (on the left; $N = 24$) and to the Verb-Switch Condition (on the right; $N = 24$). Error bars represent the standard error of the mean. Each trial has a maximal duration of 50 seconds. There was a significant interaction between Condition and Phase (habituation/test). Figure adapted from de Carvalho et al. (2015).

prosody (that delimits syntactic constituents), together with the function words (to label these syntactic constituents), in order to access the syntactic category of unknown words. In other words, putting together phrasal prosody and function words allows infants to construct a first-pass syntactic structure, which may be extremely important to constrain the acquisition of word meanings (a *syntactic skeleton*, Christophe et al., 2008).

Other recent studies using an artificial language (Hawthorne & Gerken, 2014; Hawthorne, Mazuka, & Gerken, 2015) or jabberwocky sentences (Massicotte-Laforge & Shi, 2015) have confirmed the idea that phrasal prosody may help infants organize sentences into syntactic constituents even when they do not know most of the words. For example, Hawthorne and Gerken (2014) tested the syntactic organization of sentences composed exclusively of non-words. For instance, when familiarized with non-word sentences exhibiting the prosodic pattern of two syntactic constituents, such as "[*bup div kagî*] [*feb zaf vot*]", 19-month-olds then preferred to listen to sentences like [*feb zaf vot*] [*bup div kagî*] – where the order of the two constituents is changed – than to sentences like [*zaf vot bup*] [*div kagî feb*] in which words are moved around irrespective of the constituent structure (whereas in natural languages, only words that are grouped in the same constituent can 'move' together – Hawthorne & Gerken, 2014). Hawthorne et al. (2015) extended these findings to strings of nonwords with non-native prosody. Testing English- and Japanese- acquiring 19-month-olds with sentences from an artificial language with a non-native prosodic contour (i.e., English infants with Japanese prosody and Japanese infants with English prosody), the authors demonstrated that both groups were able to use phrasal prosody to parse the speech into cohesive and re-orderable syntactic constituent-like units. This finding suggests that the cues that marked syntactic constituents boundaries in these experiments (e.g., pauses, pitch variation and phrase-final lenghthining cues) are important prosodic cues that toddlers can exploit even in non-native prosody. More generally, since phrasal prosody is found in all languages (e.g., Shattuck-Hufnagel & Turk, 1996) one would expect infants to be able to use phrasal prosody to identify word and syntactic boundaries in any language of the world.

Conclusions

This chapter provided an extensive review of empirical findings on the perception of phrasal prosody during the first steps of language acquisition. We showed that infants are sensitive to prosodic information from birth. Before their first birthday they can use phrasal prosody to segment continuous speech into words and they can also exploit prosodic information together with other cues (i.e., the position of function words relative to content words) to infer the typical word order in their language. A few months later, they readily exploit the prosodic structure of an utterance to access some of its syntactic constituents. Together with function words, this allows young children to build a partial syntactic representation, which can in some cases help them to assign a syntactic category to a novel word (e.g. noun vs. verb). Taken together, these results demonstrate the important role played by phrasal prosody during the first steps of language acquisition.

Acknowledgements

The authors thank Anne-Caroline Fiévet, Michel Dutat and Viréack Ul for technical support setting up some of the experiments reported in this chapter and all the babies and families who participated in these studies.

Funding

This work was supported by a PhD fellowship from the École normale supérieure to Alex de Carvalho, and grants from the Région IIe-de-France, Fondation de France, LabEx IEC (ANR-10-LABX-0087), IdEx PSL (ANR-10-IDEX-0001- 02), as well as the ANR 'Apprentissages' (ANR-13-APPR-0012).

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, 439(38), 419–439. doi: 10.1006/jmla.1997.2558

Armstrong, M. E., Andreu, L., Esteve-Gibert, N., & Prieto, P. (2016). Children's processing of morphosyntactic and prosodic cues in overriding context-based hypotheses: An eye tracking study. *Probus*, 28(1), 57–90.

Bernard, C., & Gervain, J. (2012). Prosodic cues to word order: What level of representation? *Frontiers in Psychology*, 3(October), 451.

Bortfeld, H., Morgan, J. L., Golinkoff, R. M., & Rathbun, K. (2005). Mommy and me: Familiar names help launch babies into speech-stream segmentation. *Psychological Science*, 16(4), 298–304. doi: 10.1111/j.0956-7976.2005.01531.x

Choi, Y., & Mazuka, R. (2003). Young children's use of prosody in sentence parsing. *Journal of Psycholinguistic Research*, 32(2), 197–217. doi: 10.1023/A:1022400424874

Christophe, A., Mehler, J., & Sebastián-Gallés, N. (2001). Perception of prosodic boundary correlates by newborn infants. *Infancy*, 2(3), 385–394. doi: 10.1207/S15327078IN0203_6

Christophe, A., Millotte, S., Bernal, S., & Lidz, J. (2008). Bootstrapping lexical and syntactic acquisition. *Language and Speech*, 51(1–2), 61–75. doi: 10.1177/00238309080510010501
Christophe, A., Nespor, M., Guasti, M. T., & Van Ooyen, B. (2003). Prosodic structure and syntactic acquisition: The case of the head-direction parameter. *Developmental Science*, 6(2), 211–220. doi: 10.1111/1467-7687.00273
Christophe, A., Peperkamp, S., Pallier, C., Block, E., & Mehler, J. (2004). Phonological phrase boundaries constrain lexical access I. Adult data. *Journal of Memory and Language*, 51(4), 523–547. doi: 10.1016/j.jml.2004.07.001
Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language*, 31(2), 218–236. doi: 10.1016/0749-596X(92)90012-M
de Carvalho, A., Dautriche, I., Lin, I., & Christophe, A. (2017). Phrasal prosody constrains syntactic analysis in toddlers. *Cognition*, 163, 67–79.
de Carvalho, A., Dautriche, I., & Christophe, A. (2016). Preschoolers use phrasal prosody online to constrain syntactic analysis. *Developmental Science*, 19(2), 235–250.
de Carvalho, A., He, X. A., Lidz, J., & Christophe, A. (2015). *18-month-olds use the relationship between prosodic and syntactic structures to constrain the meaning of novel words*. Paper presented at the 40th Boston University Conference on Language Development, Boston, USA.
de Carvalho, A., Lidz, J., Tieu, L., Bleam, T., & Christophe, A. (2016). English-speaking preschoolers can use phrasal prosody for syntactic parsing. *The Journal of the Acoustical Society of America*, 139(6), EL216–EL222.
Decasper, A., & Spence, M. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior and Development*, 9, 133–150. doi: 10.1016/0163-6383(86)90025-1
Dryer, M. (1992). The greenbergian word order correlations. *Language*, 68(1), 81–138. doi: 10.1353/lan.1992.0028
Echols, C. H. (1993). A perceptually-based model of children's earliest productions. *Cognition*, 46, 245–296. doi: 10.1016/0010-0277(93)90012-K
Echols, C. H., & Newport, E. (1992). The role of stress and position in determining first words. *Language Acquisition*, 2, 189–220. doi: 10.1207/s15327817la0203_1
Endress, A. D., & Hauser, M. D. (2010). Word segmentation with universal prosodic cues. *Cognitive Psychology*, 61(2), 177–199. doi: 10.1016/j.cogpsych.2010.05.001
Gerken, L., Jusczyk, P. W., & Mandel, D. R. (1994). When prosody fails to cue syntactic structure 9-month-olds' sensitivity to phonological versus syntactic phrases. *Cognition*, 51, 237–265. doi: 10.1016/0010-0277(94)90055-8
Gervain, J., Nespor, M., Mazuka, R., Horie, R., & Mehler, J. (2008). Bootstrapping word order in prelexical infants: A Japanese-Italian cross-linguistic study. *Cognitive Psychology*, 57(1), 56–74. doi: 10.1016/j.cogpsych.2007.12.001
Gervain, J., & Werker, J. F. (2013). Prosody cues word order in 7-month-old bilingual infants. *Nature Communications*, 4, 1490. doi: 10.1038/ncomms2430
Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, 1, 3–55. doi: 10.1207/s15327817la0101_2
Gout, A., Christophe, A., & Morgan, J. L. (2004). Phonological phrase boundaries constrain lexical access II. Infant data. *Journal of Memory and Language*, 51(4), 548–567. doi: 10.1016/j.jml.2004.07.002

Hawthorne, K., & Gerken, L. (2014). From pauses to clauses: Prosody facilitates learning of syntactic constituency. *Cognition*, 133(2), 420–428. doi: 10.1016/j.cognition.2014.07.013

Hawthorne, K., Mazuka, R., & Gerken, L. (2015). The acoustic salience of prosody trumps infants' acquired knowledge of language-specific prosodic patterns. *Journal of Memory and Language*, 82, 105–117. doi: 10.1016/j.jml.2015.03.005

He, A. X., & Lidz, J. (2017). Verb learning in 14- and 18-month-old English-learning infants. *Language Learning and Development*, 5441(March).

Hirschberg, J., & Pierrehumbert, J. (1986). The intonational structuring of discourse. *Proceedings of the 24th Meeting of the Association for Computational Linguistics* – New York, Vol. 1, 136–144.

Hirsh-Pasek, K., Kemler Nelson, D. G., Jusczyk, P. W., Cassidy, K. W., Druss, B., & Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition*, 26, 269–286. doi: 10.1016/S0010-0277(87)80002-1

Höhle, B., Weissenborn, J., Schmitz, M., & Ischebeck, A. (2001). Discovering word order regularities: The role of prosodic information for early parameter setting. In J. Weissenborn, & B. Höhle (Eds.), *Approaches to bootstrapping: Phonological, lexical, syntactic and neurophysiological aspects of early language acquisition* (pp. 246–263). Amsterdam: John Benjamins. doi: 10.1075/lald.23.15hoh

Johnson, E. K. (2008). Infants use prosodically conditioned acoustic-phonetic cues to extract words from speech. *The Journal of the Acoustical Society of America*, 123(6), EL144–EL148. doi: 10.1121/1.2908407

Johnson, E. K., & Seidl, A. (2008). Clause segmentation by 6-month-old infants: A crosslinguistic perspective. *Infancy*, 13(5), 440–455. doi: 10.1080/15250000802329321

Jun, S. (2005). Prosodic typology. In S. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 430–458). Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199249633.003.0016

Jun, S. (2014). Prosodic typology: By prominence type, word prosody, and macro-rhythm. In S. Jun (Ed.), *Prosodic typology II: The phonology of intonation and phrasing* (pp. 520–540). Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199567300.003.0017

Jusczyk, P. W. (1997). *The discovery of spoken language*. Cambridge, MA: The MIT Press.

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in english-learning infants. *Cognitive Psychology*, 39(3–4), 159–207. doi: 10.1006/cogp.1999.0716

Kjelgaard, M. M., & Speer, S. R. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language*, 40(2), 153–194. doi: 10.1006/jmla.1998.2620

Ladd, R. (2008). *Intonational phonology. Lingua* (2nd ed., Vol. 104). New York, NY: Cambridge University Press. doi: 10.1017/CBO9780511808814

Lew-Williams, C., Pelucchi, B., & Saffran, J. R. (2011). Isolated words enhance statistical language learning in infancy. *Developmental Science*, 14(6), 1323–1329. doi: 10.1111/j.1467-7687.2011.01079.x

Mandel, D. R., Jusczyk, P. W., & Nelson, D. G. (1994). Does sentential prosody help infants organize and remember speech information? *Cognition*, 53(2), 155–180. doi: 10.1016/0010-0277(94)90069-8

Männel, C., & Friederici, A. D. (2009). Pauses and intonational phrasing: ERP studies in 5-month-old German infants and adults. *Journal of Cognitive Neuroscience*, 21(10), 1988–2006. doi: 10.1162/jocn.2009.21221

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164(1), 177–190. doi: 10.1016/j.jneumeth.2007.03.024
Massicotte-Laforge, S., & Shi, R. (2015). The role of prosody in infants' early syntactic analysis and grammatical categorization. *Journal of Acoustical Society of America*, 138(October), 441–446. doi: 10.1121/1.4934551
Mehler, J., Jusczyk, P., Lamsertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cogntion*, 29, 143–178. doi: 10.1016/0010-0277(88)90035-2
Michelas, A., & D'Imperio, M. (2015). Prosodic boundary strength guides syntactic parsing of French utterances. *Laboratory Phonology*, 6(1), 119–146. doi: 10.1515/lp-2015-0003
Millotte, S., René, A., Wales, R., & Christophe, A. (2008). Phonological phrase boundaries constrain the online syntactic analysis of spoken sentences. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 34(4), 874–885. doi: 10.1037/0278-7393.34.4.874
Millotte, S., Wales, R., & Christophe, A. (2007). Phrasal prosody disambiguates syntax. *Language and Cognitive Processes*, 22(6), 898–909. doi: 10.1080/01690960701205286
Millotte, S., Morgan, J., Margules, S., Bernal, S., Dutat, M., & Christophe, A. (2010). Phrasal prosody constrains word segmentaion in French 16-month-olds. *Journal of Portuguese Linguistics*, 9–2(10–1), 67–86.
Morgan, J. (1986). *From Simple Input to Complex Grammar*. Cambridge, MA: The MIT Press.
Morgan, J., & Demuth, K. (1996). *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates.
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. doi: 10.1037/0096-1523.24.3.756
Nazzi, T., Iakimova, G., Bertoncini, J., Frédonie, S., & Alcantara, C. (2006). Early segmentation of fluent speech by infants acquiring French: Emerging evidence for crosslinguistic differences. *Journal of Memory and Language*, 54, 283–299. doi: 10.1016/j.jml.2005.10.004
Nespor, M., Shukla, M., van de Vijver, R., Avesani, C., Schraudolf, H., & Donati, C. (2008). Different phrasal prominence realization in VO and OV languages. *Lingue e Linguaggio*, 2, 1–29.
Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris.
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: Bradford Books, The MIT Press.
Saffran, J., Aslin, R., & Newport, E. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294), 1926–1928. doi: 10.1126/science.274.5294.1926
Seidl, A. (2007). Infants' use and weighting of prosodic cues in clause segmentation. *Journal of Memory and Language*, 57(1), 24–48. doi: 10.1016/j.jml.2006.10.004
Shattuck-Hufnagel, S., & Turk, E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, 25(2), 193–247. doi: 10.1007/BF01708572
Shi, R., Werker, J. F., & Morgan, J. L. (1999). Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition*, 72(2), B11–B21. doi: 10.1016/S0010-0277(99)00047-5
Shukla, M., Nespor, M., & Mehler, J. (2007). An interaction between prosody and statistics in the segmentation of fluent speech. *Cognitive Psychology*, 54(1), 1–32. doi: 10.1016/j.cogpsych.2006.04.002
Shukla, M., White, K. S., & Aslin, R. N. (2011). Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. *Proceedings of the National Academy of Sciences*, 108(15), 6038–6043. doi: 10.1073/pnas.1017617108

Snedeker, J., & Trueswell, J. (2001). *Unheeded cues: Prosody and syntactic ambiguity in mother-child communication*. Paper presented at the 26th Boston University Conference on Language Development,Boston, USA.

Snedeker, J., & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language*, 48, 103–130. doi: 10.1016/S0749-596X(02)00519-3

Snedeker, J., & Yuan, S. (2008). Effects of prosodic and lexical constraints on parsing in young children (and adults). *Journal of Memory and Language*, 58, 574–608. doi: 10.1016/j.jml.2007.08.001

Soderstrom, M., Seidl, A., Nelson, D. G., & Jusczyk, P. W. (2003). The prosodic bootstrapping of phrases: Evidence from prelinguistic infants. *Journal of Memory and Language*, 49(2), 249–267. doi: 10.1016/S0749-596X(03)00024-X

Vogel, I., & Raimy, E. (2002). The acquisition of compound vs. phrasal stress: the role of prosodic constituents. *Journal of Child Language*, 29(2), 225–250. doi: 10.1017/S0305000902005020

Warner, N., Otake, T., & Arai, T. (2010). Intonational structure as a word-boundary cue in Tokyo Japanese. *Language and Speech*, 53(1), 107–131. doi: 10.1177/0023830909351235

Weber, A., Grice, M., & Crocker, M. W. (2006). The role of prosody in the interpretation of structural ambiguities: a study of anticipatory eye movements. *Cognition*, 99(2), B63–B72. doi: 10.1016/j.cognition.2005.07.001

Wellmann, C., Holzgrefe, J., Truckenbrodt, H., Wartenburger, I., & Höhle, B. (2012). How each prosodic boundary cue matters: Evidence from German infants. *Frontiers in Psychology*, 3(DEC), 1–13.

Werker, J. F., Cohen, L., Lloyd, V., & Casasola, M. (1998). Acquisition of word-object associations by 14-month-old infants. *Developmental Psychology*, 34(6), 1289–1309. doi: 10.1037/0012-1649.34.6.1289

Zhou, P., Crain, S., & Zhan, L. (2012). Sometimes children are as good as adults: The pragmatic use of prosody in children's on-line sentence processing. *Journal of Memory and Language*, 67(1), 149–164. doi: 10.1016/j.jml.2012.03.005

CHAPTER 3

Early sensitivity and acquisition of prosodic patterns at the lexical level

Anjali Bhatara[1], Natalie Boll-Avetisyan[2], Barbara Höhle[2] & Thierry Nazzi[1]
[1]Université Paris Descartes, Sorbonne Paris Cité & CNRS (Laboratoire Psychologie de la Perception, UMR 8158) / [2]University of Potsdam

This chapter reviews sensitivity to prosodic information at the lexical level during the first year of life, considering crosslinguistic data that bear on both monolingual and bilingual acquisition. First, we discuss infants' early ability to discriminate lexical stress, and how this sensitivity changes across development depending on the native language, reflecting the acquisition of native prosodic properties. Second, we present data establishing the emergence of language-specific lexical stress preferences during that same period, also attesting language-specific acquisition. We then discuss data on the less well studied perception of lexical pitch accent and tone. Finally, we consider the role of lexical stress in early word segmentation abilities, and discuss the representations and processes underlying early prosodic perception at the lexical level.

Introduction

As one of their first steps in language acquisition, infants have to acquire not only their native languages' phoneme inventory, but also their prosodic properties, such as basic linguistic rhythm or sentence-level intonation, and word level prosody. In the present chapter, we focus on how infants process word-level prosodic characteristics and then acquire the properties that their native language(s) use to distinguish between words in the lexicon. There are three different types of word-level prosodic phenomena that can be distinguished, namely lexical stress, lexical pitch accent and lexical tone. In languages with contrastive lexical stress, differences in stress pattern (i.e., differences in prominence marked by a combination of intensity, pitch and duration associated with stressed and

doi 10.1075/tilar.23.03bha

unstressed syllables) can change the meaning or the syntactic class of a word (e.g., in English, /IMport/ is a noun while /imPORT/ is a verb). In languages with lexical pitch accent, the presence or absence of a pitch accent (marked by a steep pitch movement, e.g., a pitch fall in Japanese) and the accent's location in a word are lexically contrastive (e.g., in Japanese, /HAshi/ with falling contour has a pitch accent on /HA/ and means "chopsticks," whereas /haSHI/ with a rising contour has no pitch accent and means "bridge"). In languages with lexical tones, each individual syllable carries pitch contour characteristics (e.g., rising, falling, dipping, level…), and differences in tones change the meaning of a word (for example, in Mandarin, /ma/ can mean "mother," "hemp," "horse," or "scold" depending on the tone). The present chapter will focus on how lexical stress patterns are processed during early infancy as a function of the language(s) in acquisition, illustrating a shift from language-general to language-specific perception. Additionally, we will also discuss, more briefly, early perception of lexical pitch accents and lexical tones.

Infants come to life equipped with a specific sensitivity to prosodic information contained in the speech input. Newborns and a-few-month-old infants can discriminate different prosodic patterns at the sentence/utterance level (Christophe, Guasti, Nespor, Dupoux, & Van Ooyen, 1997; Guasti, Nespor, Christophe, & van Ooyen, 2001; Mehler et al., 1988; Nazzi, Bertoncini, & Mehler, 1998). For example, infants discriminate and prefer the global prosodic features of their target language over that of non-native languages, if the languages belong to different rhythmic classes (Bosch & Sebastián-Gallés, 1997; Mehler et al., 1988; Moon, Panneton-Cooper, & Fifer, 1993; Nazzi, Jusczyk, & Johnson, 2000). Since some of these effects are found at birth, they are likely to result from prosodic acquisition during the last semester of gestation. Prosodic bootstrapping theories have been proposed (Gleitman, 1990; Morgan & Demuth, 1996; Weissenborn & Höhle, 2001), according to which the processing and acquisition of prosodic information provides cues to the acquisition of higher levels of the linguistic hierarchy, such as the lexical and syntactic levels. At present, numerous studies have suggested that prosodic processing plays a role in discovering syntactic units such as clauses and phrases (e.g., Hirsh-Pasek et al., 1987; Jusczyk et al., 1992; Männel, Schipke, & Friederici, 2013; Nazzi, Kemler Nelson, Jusczyk, & Jusczyk, 2000; Seidl & Cristia, 2008; Soderstrom, Seidl, Kemler Nelson, & Jusczyk, 2003; Wellmann, Holzgrefe, Truckenbrodt, Wartenburger, & Höhle, 2012) and specifying the position of the heads of syntactic phrases (e.g., Guasti et al., 2001; Nespor, Guasti, & Christophe, 1996). These studies have shown an early sensitivity to prosodic information that is related to syntactic phrases and clauses and have demonstrated the usefulness of such sensitivity for later language learning (for an in-depth discussion of phrasal prosody in development, see de Carvalho, Dautriche, Millotte, & Christophe, this volume). In addition, numerous studies have demonstrated this same sensitivity and usefulness for prosodic information at the word level, which we discuss below.

From language-general to language-specific processing of lexical stress

In many languages, lexical stress is used contrastively (e.g., English, German, Spanish). In other languages, stress is fixed, always falling on a particular syllable position (e.g., the first syllable of the word in Hungarian; van der Hulst, 1999). Moreover, a few languages do not have any lexical stress (e.g., French: Di Cristo, 1998; and Korean: Jun, 1996), though some syllables may be more accentuated than others (in French, accentuation falls on the last syllable of words in phrase-final positions; Delattre, 1938; Di Cristo, 1998). Importantly, in all of these languages with fixed or no lexical stress, words cannot be contrasted solely by stress. Therefore, sensitivity to lexical stress should be relevant for word recognition for infants learning a language with variable lexical stress, but possibly not for infants learning fixed-stress or no-stress languages. This idea is supported by adult research showing less sensitivity to stress cues among listeners of no-stress (French) and fixed-stress languages (Finnish, Hungarian) when compared with Spanish listeners (Dupoux, Pallier, Sebastián, & Mehler, 1997; Dupoux, Peperkamp, & Sebastián-Gallés, 2001; Peperkamp, Vendelin, & Dupoux, 2010). Hence, infants' sensitivity to lexical stress might follow different developmental trajectories depending on the specific requirements of their native language(s). Next, we review studies exploring the acquisition of native lexical stress, and its impact on infant processing.

Discrimination of word-level stress patterns

Studies exploring sensitivity to stress information at the word level have tested infants in two types of experimental contexts. In the easier contexts, infants are presented with one or several tokens of a single item (e.g., /gaba/) recorded either with a stress-initial (trochaic) or stress-final (iambic) pattern. These studies explore whether infants can discriminate stress patterns using low-level acoustic properties. In the more challenging contexts, infants are presented with lists of segmentally different items (e.g., /datu/, /sapi/) recorded either with a trochaic or iambic pattern. These studies investigate whether infants can abstract and discriminate prosodic patterns over segmental variability.

In the absence of segmental variability, early discrimination of lexical stress has been found in Italian newborns (Sansavini, Bertoncini, & Giovanelli, 1997), English-learning 2-month-olds (Jusczyk & Thompson, 1978), German-learning 4-month-olds (ERP data: Friederici, Friedrich, & Christophe, 2007; behavioral data: Höhle, Bijeljac-Babic, Herold, Weissenborn, & Nazzi, 2009; Höhle, Pauen, Hesse, & Weissenborn, 2014) and Spanish-learning 6-month-olds (Skoruppa et al., 2013). Importantly, it was also found in French-learning infants from 4 to 10 months (Friederici et al., 2007; Höhle et al., 2009; Skoruppa et al., 2009, 2013), even though stress is not used at the lexical level in French. This establishes the existence of early language-independent stress discrimination abilities in the absence of segmental variability.

However, by 9/10 months, language-specific reorganization is observed. French-learning infants' sensitivity to lexical stress declines, even in the absence of segmental variation. In familiarization-test experiments, French-learning 9/10-month-olds discriminated stress patterns following a 2-minute familiarization with one pattern but failed with a 1-minute familiarization, while French-learning 6-month-olds succeeded with only 1 minute (Bijeljac-Babic, Serres, Höhle, & Nazzi, 2012; Höhle et al., 2009; Skoruppa et al., 2009). This suggests that French-learning infants need more time to identify stress patterns at 10 than 6 months, a developmental pattern consistent with the fact that French does not use stress to contrast words.

Do stress pattern discrimination abilities extend to contexts with segmental variability, in which infants have to detect a common prosodic pattern over different words? Such ability appears limited in monolingual newborns, since they can discriminate stress patterns when presented with lists of words varying in their consonants (Sansavini et al., 1997) but not in both their consonants and vowels (Sansavini, Bertoncini, & Giovanelli, 1994). Moreover, both Spanish- and French-learning 6-month-olds have difficulties discriminating stress patterns when presented with lists of segmentally different words (Skoruppa et al., 2013). Clear evidence of discrimination abilities are nevertheless found in older infants, in Spanish-learning 9-month-olds and English-learning 8- and 12-month-olds (Skoruppa, Cristià, Peperkamp, & Seidl, 2011; Skoruppa et al., 2009), while French-learning 9/10-month-olds fail to discriminate (Abboub, Bijeljac-Babic, Serres, & Nazzi, 2015; Skoruppa et al., 2009). Taken together, these results suggest that after 8/9 months of age, monolingual infants learning variable-stress languages process stress efficiently in the presence of segmental variation, while monolingual infants learning a no-stress language like French show difficulties.

Acquisition of lexical stress patterns

The previous section showed that infants' sensitivity to lexical stress contrasts changes during the first year of life, reflecting the probable acquisition of prosodic properties of words from the native language. These findings are complemented by studies on the emergence of preferences for predominant stress patterns in languages that have contrastive stress.

In contexts of no segmental variability, language-specific developmental changes are revealed by the emergence of a listening preference for trochaic over iambic items in German-learning infants between 4 and 6 months of age, reflecting knowledge of the predominance of trochaic words in their input (Höhle et al., 2009). Since no such bias is found at 6 months in French-learning infants, this crosslinguistic pattern of preferences is likely to reflect the acquisition of word level prosodic properties (Höhle et al., 2009).

In the presence of segmental variability, language-specific developmental changes were first revealed by the emergence of a preference for trochaic over iambic words in English-learning infants between 6 and 9 months, reflecting knowledge of the predominance of trochaic words in English (Jusczyk, Cutler, & Redanz, 1993). Similarly, Hebrew-learning infants, whose language is predominantly iambic, show a preference for iambic over trochaic words at the same age (Segal & Kishon-Rabin, 2012). Similar research on Spanish partly replicated these effects, while showing that preferences can be modulated by syllabic structure: Indeed, the predominance of the trochaic pattern (60%) in Spanish disyllabic words does not lead to a trochaic preference in Spanish-learning 9-month-olds presented with CV.CV words. However, when the stimuli contained a heavy (CVC) syllable in word initial position (CVC.CV), a trochaic preference was found, while an iambic preference was found in the opposite case (CV.CVC; Pons and Bosch, 2010).

In summary, word-level prosodic properties of the native language start being acquired by 6 to 10 months, as attested by changes in sensitivity to stress contrasts and the emergence of preferences for predominant stress patterns (for further discussion of the impact of prosody on word learning in development, see Thorson, this volume). We next present the results of a study in which we aimed to specify which of the three acoustic cues that are modulated to create stress (duration, intensity and pitch) might drive early stress pattern preferences, an issue not answered by the above studies, as all of these used natural stimuli in which these cues varied in a natural, hence, uncontrolled manner.

An attempt at dissociating the acoustic correlates of stress

Trochaic words are usually louder and higher-pitched on the initial syllable, whereas iambic words usually have a longer duration on the final syllable (e.g., Hayes, 1995). In all the above-mentioned studies, these three acoustic cues, intensity, pitch (F0), and duration, were co-varying. This leads to the question of what their separate roles could be in marking stress. Could they have different relevance in different languages, depending on the realization of prosody in those languages? In an initial experiment, reported here for the first time, these issues were explored by testing stress pattern preference in French- and German-learning monolingual 6-month-olds, presenting dissociated cues (duration, intensity or pitch) and using the Headturn Preference Procedure (HPP).

As discussed in the previous section, German-learning but not French-learning 6-month-olds have shown a preference for trochaic words (the predominant stress pattern in German), that is, a preference for stress patterns with a louder and/or higher-pitched initial syllable (Höhle et al., 2009). However, German does have iambic words, in which prominence is mainly marked by longer duration on the final syllable. French has no lexical stress, but at the phrasal level and for isolated words, stress is iambic (i.e., marked by longer duration at the end of the phrase or word). Based on these findings and characteristics of the two languages, we predicted that as opposed to the French-learning infants, German-learning infants would show a trochaic bias when pitch or intensity mark stress. Moreover, both the German- and the French-learning infants might either show an iambic bias or no preference when duration marks stress.

We tested 31 infants in the duration condition (20 French, 11 German), 31 in the intensity condition (15 French, 16 German), and 27 in the pitch condition (15 French, 12 German). All infants were presented with 2 lists of segmentally varied disyllabic non-words (e.g., /meli bezu lomi …/). The items were synthesized with MBROLA (Dutoit, Pagel, Pierret, Bataille, & van der Vrecken, 1996), using the fr4 French voice and the de5 German voice. In each condition, the stimuli in the two lists contrasted in stress pattern (trochaic versus iambic), which was acoustically realized by varying one of the three acoustic cues (duration, pitch, or intensity) while the other two were held constant. Half of the infants in each group heard the German voice and half heard the French voice. See Table 1 for stimulus characteristics. These particular values were chosen in order to stay close to the values naturally present in the two languages (as in Bhatara, Boll-Avetisyan, Unger, Nazzi, & Höhle, 2013) with increased pitch variation as occurs in infant-directed speech, similar to the values used in Abboub, Boll-Avetisyan, Bhatara, Höhle, and Nazzi (2016) and Bion, Benavides-Varela, and Nespor (2011).

Table 1. Acoustic characteristics of unaccented and accented syllables (σ) in the bisyllabic trochaic and iambic stimuli, for the 3 experimental conditions

		Duration	**Intensity**	**Pitch**
	Unaccented σ	260 ms	70 dB	200 Hz
Accented σ	Duration condition	460 ms	70 dB	200 Hz
	Intensity condition	260 ms	78 dB	200 Hz
	Pitch condition	260 ms	70 dB	420 Hz

Preliminary results showed no looking time (LT) preference for any of the prosodic patterns in either the French- nor the German-learning infants (Figure 1, left panels). This was true for all conditions: duration ($M_{iambs} = 8.99$ s versus $M_{trochees} = 8.91$ s, p = .82), intensity ($M_{trochees} = 7.53$ s versus $M_{iambs} = 7.59$ s, $p = .88$), and pitch ($M_{trochees =}$ 8.38 s versus $M_{iambs =}$ 8.16 s, $p = .57$).

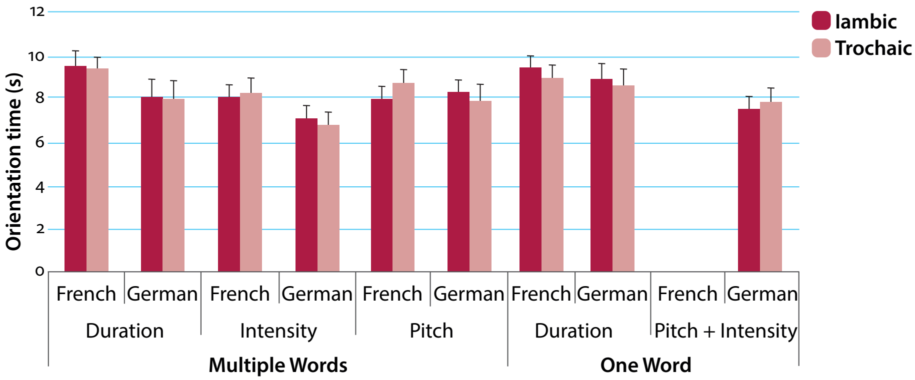


Figure 1. Mean orientation times for each stress pattern for the French- and German-learning infants. Error bars show standard error.

There are at least two possible reasons why infants might have failed to show preferences in the present study, even though German-learning infants of the same age succeeded with natural stimuli (Höhle et al., 2009). The first one might be that infants were presented with lists of phonetically-varied stimuli, requiring that they process the information at a more abstract level than with the natural tokens of a single pseudoword /gaba/ used by Höhle et al. (2009). Therefore, 20 German- and 19 French-learning 6-month-olds were tested on duration contrasts again, this time without segmental variability, each infant being tested with a single pseudoword (a different pseudoword for each infant) presented in each trial as either a trochee or iamb. As in the main study, neither the French nor the German group showed a preference ($M_{trochees} = 8.81$ s versus $M_{iambs} = 9.17$ s, $p = .3$; Figure 1, second panel to the right).

The second reason is that isolated cues might not be strong enough to support the clear identification of the stress patterns required to induce the trochaic bias found at 6 months in German. Accordingly, and given the fact that stressed syllables are marked by a combined increase in both pitch and intensity in German, we tested 18 German-learning 6-month-olds with a single pseudoword (a different word for each infant) varying in both pitch and intensity, presented across trials either with a trochaic or an iambic pattern. Again, the infants showed no preference ($M_{trochees} = 7.88$ s versus $M_{iambs} = 7.56$ s, $p = .48$; Figure 1, right-most panel).

Given previous evidence for stress pattern preferences among German-learning infants at that age (Höhle et al., 2009), these results might suggest that it is yet too difficult for 6-month-olds to associate single stress cues with a specific stress pattern (and possibly even some combinations of 2 cues, as done in our final experiment). This interpretation would be in line with infant studies on the perception of prosodic phrase boundary cues, which also found that older infants could discriminate between sentences on the basis of single prosodic cues while younger ones needed a combination of all cues that naturally co-occur (Seidl, 2007; Seidl & Cristia, 2008; Wellmann et al., 2012). Accordingly, one way to follow up on the present studies would be to test older infants.

The null-results in the last two experiments may also indicate that our stimuli may not have been well-adapted. While previous studies had used either multiple tokens of one pseudoword (Höhle et al., 2009), or multiple naturally recorded (pseudo)words (Jusczyk et al., 1993; Pons & Bosch, 2010; Segal & Kishon-Rabin, 2012), each infant in the present study was presented with a single token, so that there was no variation along any acoustic parameter except for that which determined the stress pattern. As this lack of variability never happens in natural speech, it would not be surprising if infants have difficulties relating these stimuli to their linguistic knowledge (see Rost & McMurray, 2009, for a similar argument made for word learning at 14 months of age). One way to address these issues in future work would be to manipulate natural speech to leave only pitch or only duration variation, using multiple tokens to ensure that infants receive enough variability of input.

Bilingualism and lexical stress

Nowadays, more than half of the world's population is bilingual (Grosjean, 2010). Many recent studies have explored how infants growing up in bilingual families, hearing two languages from birth, detect and rapidly learn the regularities of each of their languages (Byers-Heinlein, Burns, & Werker, 2010; Weikum et al., 2007). These studies reveal complex patterns, in which acquisition does not seem to be slower in bilingual than in monolingual infants (e.g. Bosch & Sebastián-Gallés, 1997), though with respect to phoneme perception, simultaneously learning two linguistic system has sometimes been found to affect early developmental trajectories (Bosch & Sebastián-Gallés, 2003a, 2003b; Liu & Kager, 2015; but c.f. Albareda-Castellot, Pons, & Sebastián-Gallés, 2010; Sundara, Polka, & Molnar, 2008).

Bilingual infants' first main task in language acquisition is separating the two languages to which they are exposed, and there is now evidence that they have the early capacity to use prosodic information to do so. At birth, as for monolinguals

(Nazzi, Bertoncini, & Mehler, 1998), this capacity is restricted to languages differing in rhythmic class for bilingual newborns (Byers-Heinlein et al., 2010). Around 4 months, again as for monolinguals (Butler, Floccia, Goslin, & Panneton, 2011; Molnar, Lallier, & Carreiras, 2014; Nazzi, Jusczyk, & Johnson, 2000), bilingual infants can discriminate languages with similar rhythms (Bosch & Sebastian-Galles, 1997; Molnar, Gervain, & Carreiras, 2013). Thus, at present, apart from some fine-grained differences (Bosch & Sebastián-Gallés, 1997), bilinguals and monolinguals appear to have similar language discrimination abilities, using – at least partially – global prosodic characteristics of the languages in question. However, what about their processing of prosodic information at the word level?

Bilingual infants' discrimination of lexical stress patterns

Two studies explored lexical stress pattern discrimination by bilingual infants, in the absence or presence of segmental variability. In its absence (using the trochaic and iambic /gaba/ German tokens of Höhle et al., 2009), Bijeljac-Babic et al. (2012) found that 10-month-old bilinguals learning French and a language that has variable lexical stress can discriminate stress contrasts, and that they do so better than French-learning monolinguals. This effect was significant only in the subgroup of infants dominant in the language with variable lexical stress, and was marginal in the balanced bilinguals. This suggests that learning a language with lexical stress maintains stress pattern discrimination (depending on language balance).

Do such abilities extend to contexts with segmental variability, in which bilingual infants have to detect a common prosodic pattern across different words? Earlier, we presented evidence that such abilities are limited in monolingual infants up to 6 months of age (Sansavini et al., 1994, 1997; Skoruppa et al., 2013), but improve by the age of 8 to 12 months for infants learning languages with variable lexical stress (Skoruppa et al., 2009; Skoruppa et al., 2011), but not for infants learning only French, which lacks lexical stress (Abboub et al., 2015; Skoruppa et al., 2009). Accordingly, Abboub and colleagues (2015) examined lexical stress perception in the presence of segmental variation in 10-month-old bilinguals learning both French and a second language with variable lexical stress, comparing their performance to that of 10-month-old French-learning monolinguals. They used the same 1-minute of familiarization HPP procedure as Bijeljac-Babic et al. (2012) and the stimuli of Skoruppa et al. (2009). Discrimination was found for the bilingual but not the monolingual infants, with no effect of language balance (contrary to Bijeljac-Babic et al., 2012; see Figure 2). This indicates that these bilingual infants can not only perceive the acoustic cues for the prosodic contrasts, but can also extract and generalize the patterns. This is further supported by the fact that in both bilingual studies, the stimuli presented to the bilinguals were not produced by a speaker of their native language(s), hence suggesting that these bilingual infants could process lexical stress in a foreign language, which might mark stress in at least a partially different manner than in their native language.

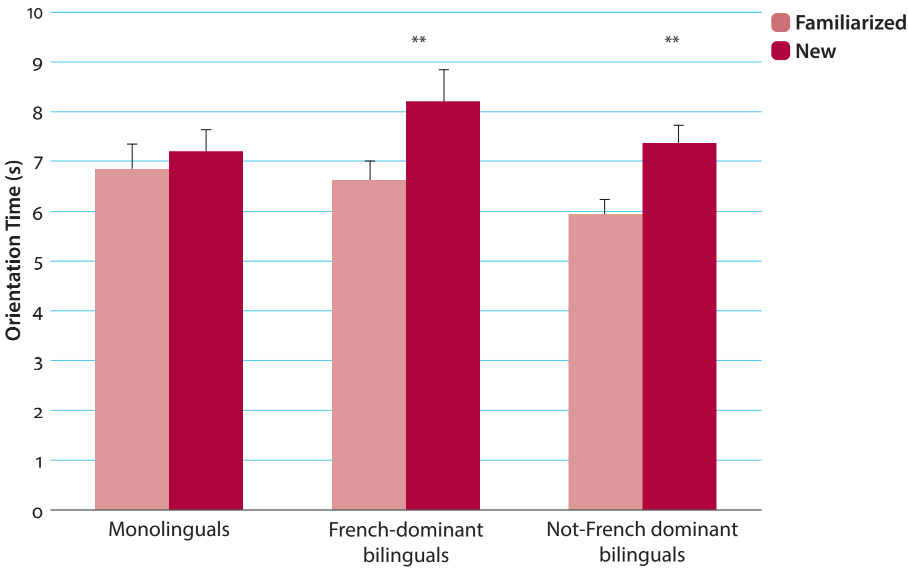


Figure 2. Mean orientation times (and SEs) to the familiarized versus new stress patterns, broken down by linguistic group: French-learning monolinguals, French-dominant bilinguals (60–70% French), and not-French-dominant bilinguals (30–50% French). All bilinguals were learning a language with variable stress (adapted from Abboub et al., 2015).

Bilingual infants' preference for native prosodic patterns

The issue of the acquisition of native prosodic patterns in bilinguals has only been explored so far in the absence of phonetic variation. Bijeljac-Babic, Höhle, and Nazzi (2016) found a trochaic preference in French-German bilinguals at 6 months, the age at which the effect is first found in German-learning infants (Höhle et al., 2009). This effect was independent of the proportion of exposure to the two languages (between 30% and 70% German), suggesting that 30% German exposure for 6 months is enough to develop a preference for the trochaic pattern predominant in German, at least when presented with German non-word stimuli.

This lone study exploring native prosodic pattern acquisition in bilingual infants, suggesting a similar acquisition pattern as in monolinguals, leaves many avenues open to explore. For example, the bilingual infants in Bijeljac-Babic et al. (2016) show a trochaic bias when hearing stimuli in the language for which it is appropriate (German), but we do not know whether this bias is language-specific, or whether it extends to the processing of stimuli in their other language (French), even though it would not be appropriate in that case (at least from a monolingual perspective). Second, it would be useful to explore this question in infants learning other language pairs, perhaps two languages that have different lexical stress patterns. We anticipate different developmental trajectories, depending on the relation between the lexical stress systems of the two native languages.

Beyond lexical stress: Lexical pitch accents and lexical tones

Although the present chapter focuses on the processing and acquisition of lexical stress patterns, among the languages of the world, there are other ways to prosodically contrast words, by either varying the pitch contour applied to the word (as in Japanese), or using different lexical tones (as in Chinese languages). In the following, we more briefly evaluate what we know regarding the processing and acquisition of these prosodic properties in infancy depending on the native language.

Lexical pitch accents

Processing of lexical pitch accents has only been studied in one language, Japanese. Sensitivity to lexical pitch accents is present at least from birth, even in newborns who do not hear a language with lexical pitch accents in their environment, establishing the language-independence of this early ability (Nazzi, Floccia, & Bertoncini, 1998). Evidence suggesting linguistic specialization in discrimination of lexical pitch accent comes from a study exploring 4- and 10-month-old Japanese-learning infants, using both a behavioral method and a brain imaging technique (Near Infra-Red Spectroscopy, NIRS; Sato, Sogabe, & Mazuka, 2010). While infants at both ages could discriminate the pitch accents behaviorally, the NIRS data revealed developmental changes, with left hemispheric lateralization for speech but not pure tone stimuli found only at 10 months, a pattern similar to what is found for adult speakers of Japanese dialects that use pitch accents (Sato, Utsugi, Yamane, Koizumi, & Mazuka, 2013). Future studies on the linguistic specialization of pitch accent perception should be conducted cross-linguistically, keeping in mind that pitch accents, while not used at the lexical level in many languages, are used at higher levels of the prosodic structure.

Lexical tones

The study of lexical tone perception in infancy has been the focus of many recent studies, following the seminal paper of Mattock and Burnham (2006) examining tone perception in Chinese- and English-learning infants. Tone discrimination was found in Chinese-learning infants (Mandarin or Cantonese) starting at 4 months and beyond (Mattock & Burnham, 2006; Mattock, Molnar, Polka, & Burnham, 2008; Yeung, Chen, & Werker, 2013). In infants learning non-tone languages, several studies established early tone discrimination: between 4 and 6 months in English-learning infants (Mattock & Burnham, 2006; Mattock et al., 2008; Yeung et al., 2013), by 6 months in French-learning infants (Cabrera et al., 2015; Mattock et al., 2008), and between 6 and 18 months in Dutch-learning infants (Liu & Kager, 2014). This establishes that young infants are able to discriminate tone contrasts whether or not they are learning a language with tones.

Many of these same studies further show that tone perception later changes as a result of linguistic experience, resulting in a decrease in sensitivity. For infants learning a tone language (Cantonese or Mandarin), it appears that the acquisition of the native tonal system has started by 4 months of age (Yeung et al., 2013). For infants learning non-tone languages, a decline in lexical tone discrimination was found between 6 and 9 months in English-, French- and Dutch-learning infants (Liu & Kager, 2014; Mattock & Burnham, 2006; Mattock et al., 2008; Yeung et al., 2013). However, the study by Liu and Kager (2014), showing a U-shaped trajectory in tone discrimination (with best performance in the 6-month-olds and the 18-month-olds), suggests that the shared use of pitch variation to mark tonal information in tone languages, and intonation across all languages, coupled with the acquisition of the native intonation system, might allow infants learning a non-tonal language to regain some sensitivity to tone contrasts.

There is evidence also that simultaneously learning a non-tone language with a tone language may affect developmental trajectories in both languages. Bilingual Mandarin-English-learning infants at three ages (7.5, 9, and 11 months) were tested on their ability to recognize words in both English and Mandarin that either matched or did not match the pitch of previously presented words (Singh & Foong, 2012). Infants began with an exact-matching strategy in both languages, but by 11 months, they showed language-specific strategies, assigning importance to pitch only in Mandarin.

In summary, the above studies show early language-independent sensitivity to tone contrasts that starts to become language-specific between 4 and 9 months of age. This general developmental trajectory will require some further refinement given emerging evidence that changes in tone sensitivity might not be identical across all tone pairs (e.g., Chen & Kager, 2016). Hence, studies adopting a crosslinguistic approach and a more systematic and controlled use of different tone pairs will be needed to distinguish general features of tone perception and acquisition versus idiosyncratic effects linked to a specific tone pair.

Lexical stress and word form segmentation

One prerequisite to establish a lexicon is for infants to identify the word forms that, attached to a meaning, will constitute words in their native language. This process appears to be non-trivial, as both adult- and infant-directed speech is mostly constituted of sentences rather than isolated words (e.g., Brent & Siskind, 2001). This means that infants will have to extract word forms from fluent speech, an issue known as the segmentation problem. While no single cue exists that systematically marks word boundaries in fluent speech, a number of cues have been identified and found to be used early in development in different languages (including statistical, prosodic, and phonotactic cues), and word segmentation abilities have been found to emerge around 6 to 8 months of age in several languages, including American and Canadian English, German, Dutch, and European and Canadian French. Although a complete review of this topic is beyond the scope of the present chapter (for a recent review, see Goyet, Millotte, Christophe, & Nazzi, 2016, and for a discussion of delay in British English see Floccia et al., 2016), we discuss below how lexical stress, the topic of the present chapter, has been found to be used as an early cue to word form segmentation.

The capacity to use lexical stress for speech segmentation appears early in development, and seems language-specific. By 7 to 10 months, infants whose native language is stress-timed with words being predominantly trochaic (strong-weak; including American English, Canadian English and Dutch) can segment words with a trochaic stress pattern from speech (behavioral HPP studies: Houston, Jusczyk, Kuijpers, Coolen, & Cutler, 2000; Jusczyk, Houston, & Newsome, 1999; Polka & Sundara, 2012; ERP studies: Kooijman, Hagoort, & Cutler, 2005). English-learning infants can also use this prosodic knowledge for segmenting trochaic words from non-native speech (Houston et al., 2000), including Italian, a syllable-timed language (Pelucchi, Hay, & Saffran, 2009), and for segmenting words from nonsense syllable strings (Curtin, Mintz, & Christiansen, 2005; Echols, Crowhurst, & Childers, 1997). Even earlier, at 6 months, English-learning infants can segment trochaic words if the words coincide with utterance boundaries (Johnson, Seidl, & Tyler, 2014).

Importantly though, Dutch- and English-learning infants' ability to segment words with an iambic pattern (the non-predominant stress pattern in these languages) is developed somewhat later, beginning at the age of 10 months (Jusczyk et al., 1999; Kooijman, Hagoort, & Cutler, 2009). This difference in the ability to segment trochaic versus iambic words, mirroring the stress pattern properties of their native language, suggests that the prosodic cues for speech segmentation are acquired from the input language (although studies on languages with a predominant iambic patterns, such as Hebrew, would be needed to confirm this hypothesis).

One interpretation of the above findings is that, at least in stress-based languages with predominant trochaic stress, infants, by 7 to 10 months of age, start using stressed syllables as word onsets (Jusczyk et al., 1999). Alternatively, it has been proposed that this segmentation pattern might result from the fact that by this age, English- and Dutch-learning infants have started using the rhythmic unit of their native language (the trochaic foot) as a segmentation unit (Nazzi, Iakimova, Bertoncini, Frédonie, & Alcantara, 2006). This second hypothesis makes the additional prediction that infants learning syllable-based languages such as French should show a different segmentation pattern, because they would rely on a different rhythmic unit, the syllable, which is the rhythmic unit of French (which is also a language without lexical stress).

Evidence for the "early rhythmic segmentation" hypothesis comes from several studies with European French-learning infants, which converge in showing an advantage of syllabic over bisyllabic word segmentation. In a first study using HPP, French-learning infants, familiarized with isolated words and tested on passages that either did or did not contain the target words, did not segment bisyllabic words until the age of 16 months. However, earlier, at 12 months, they segmented the individual syllables of the bisyllabic words, which supports the view that they use the syllable, the rhythmic unit of French, as an earlier segmentation cue (Nazzi et al., 2006). While later studies (Nazzi, Mersad, Sundara, Iakimova, & Polka, 2014; Nishibayashi, Goyet, & Nazzi, 2015) revealed that French-learning infants, in fact, start segmenting bisyllabic words between 6 and 8 months when familiarized with passages and tested on their recognition of the target words (that is, reversing the order used by Nazzi et al., 2006), it was also found that both 6- and 8-month-olds could segment monosyllabic words, or syllables embedded in bisyllabic words (Goyet, Nishibayashi, & Nazzi, 2013; Nishibayashi et al., 2015). This pattern supports the precedence of syllabic segmentation in the emergence of segmentation abilities in European French. Moreover, ERP paradigms revealed infants' segmentation of bisyllabic words at 12 months when using a procedure of familiarizing infants with isolated words and testing them on text passages (Goyet, de Schonen, & Nazzi, 2010). Lastly, a related study by Polka and Sundara (2012) conducted with Canadian French infants showed that, when familiarized with target bisyllabic words or their initial/final syllables, and then tested on passages containing or not the target words (as in Nazzi et al., 2006), 8-month-olds recognized both the bisyllabic targets and their individual syllables (significant effect for initial syllable; marginal effect for final syllables). These studies not only reveal cross-dialect differences in early word segmentation in European and Canadian French, they also suggest that these Canadian French infants have access to syllabic units while segmenting fluent speech, as predicted by the early rhythmic segmentation hypothesis.

In sum, the evidence thus far suggests that the way infants use prosodic cues to extract word forms from fluent speech (i.e., word segmentation) is at least partly related to the acquisition of the predominant lexical stress pattern, which might itself be related to the acquisition of the rhythmic unit of the native language. In the future, it will be important to test the prediction from this hypothesis that infants learning a language in which words are predominantly iambic, such as Hebrew, will have an initial advantage at segmenting iambic over trochaic words. Future studies should also explore the role of lexical pitch contour and lexical tone in the early word segmentation in languages in which predominant patterns would exist.

Concluding remarks

In the present chapter, we reviewed evidence regarding the early processing and acquisition of prosodic properties at the lexical level. While our focus was on lexical stress patterns, we also briefly reviewed evidence for lexical pitch contours and lexical tones. Overall, the findings suggest that infants come equipped with sensitivity to prosodic information that allows them to be sensitive to (most) prosodic variation used phonologically at the lexical level across languages of the world. We also showed that these early sensitivities rapidly change (as early as 4 months, no later than 9 months), attesting the acquisition of the prosodic properties of the native language(s). We discussed some factors that have been found to play a role in the pattern of performance observed (e.g., the processing of the stimuli at the acoustic versus phonological level; language predominance in the case of bilingual infants), and offered some avenues for future research, in particular the need to conduct direct crosslinguistic studies including more diverse languages and to look at the use of specific acoustic dimensions (given differences across languages in how a given property, e.g. lexical stress, is instantiated acoustically).

Finally, note that some studies have started exploring the implications of these prosodic acquisitions on other linguistic abilities. For example, we reviewed studies exploring the role of lexical prosody in early word form segmentation. Similarly, recent studies explored principles guiding the grouping of syllables into larger chunks based on variation intensity, duration and pitch (the Iambic-

Trochaic Law: Hayes, 1995). Results establish that the ITL is modulated both in adulthood (Bhatara, Yeung, & Nazzi, 2015; Bhatara et al., 2013; Boll-Avetisyan, Bhatara, Unger, Nazzi, & Höhle, 2016; Iversen, Patel, & Ohgushi, 2008; Molnar, Carreiras, & Gervain, 2016) and in infancy (Abboub et al., 2016; Bion et al., 2011; Hay & Saffran, 2012; Molnar et al., 2014; Yoshida et al., 2010), and it has been proposed that this modulation might be related to the acquisition of native prosodic properties either at the phrase level (e.g., Iversen et al., 2008; Molnar et al., 2016) or the lexical level (e.g., Bhatara et al., 2013). Future studies will have to continue exploring how the early acquisition of word-level prosody has cascading effects on the acquisition of higher levels of the linguistic structure.

Funding

This study was conducted with the support of an ANR-DFG grant (ANR-13-FRAL-0010 to RBB; DFG 1960/14-1 and 1960/15-1 to BH), and a LABEX EFL (ANR-10-LABX-0083) grant to TN.

Acknowledgements

Special thanks to the infants and their parents for their kindness and cooperation.

References

Abboub, N., Bijeljac-Babic, R., Serres, J., & Nazzi, T. (2015). On the importance of being bilingual: Word stress processing in a context of segmental variability. *Journal of Experimental Child Psychology*, 132, 111–120. doi:10.1016/j.jecp.2014.12.004

Abboub, N., Boll-Avetisyan, N., Bhatara, A., Höhle, B., & Nazzi, T. (2016). An exploration of rhythmic grouping of speech sequences by French- and German-learning infants. *Frontiers in Human Neuroscience*, 10, 292.

Albareda-Castellot, B., Pons, F., & Sebastián-Gallés, N. (2010). The acquisition of phonetic categories in bilingual infants: New data from an anticipatory eye movement paradigm. *Developmental Science*, 14(2), 395–401. doi: 10.1111/j.1467-7687.2010.00989.x

Bhatara, A., Boll-Avetisyan, N., Unger, A., Nazzi, T., & Höhle, B. (2013). Native language affects rhythmic grouping of speech. *The Journal of the Acoustical Society of America*, 134(5), 3828–3843. doi:10.1121/1.4823848

Bhatara, A., Yeung, H. H., & Nazzi, T. (2015). Foreign language learning is related to better rhythm perception in French speakers. *Journal of Experimental Psychology: Human Perception and Performance*, 41, 277–282.

Bijeljac-Babic, R., Serres, J., Höhle, B., & Nazzi, T. (2012). Effect of bilingualism on lexical stress pattern discrimination in French-learning infants. *PloS One*, 7(2), e30843. doi:10.1371/journal.pone.0030843

Bijeljac-Babic, R., Höhle, B., & Nazzi, T. (2016). Early prosodic acquisition in bilingual infants: the case of the perceptual trochaic bias. *Frontiers in Psychology*, 7, 210.
Bion, R. A., Benavides-Varela, S., & Nespor, M. (2011). Acoustic markers of prominence influence infants' and adults' segmentation of speech sequences. *Language and Speech*, 54(1), 123–140. doi:10.1177/0023830910388018
Boll-Avetisyan, N., Bhatara, A., Unger, A., Nazzi, T., & Höhle, B. (2016). Effects of experience with L2 and music on rhythmic grouping by French listeners. *Bilingualism: Language and Cognition*, 19(5), 971–986.
Bosch, L., & Sebastián-Gallés, N. (1997). Native-language recognition abilities in 4-month-old infants from monolingual and bilingual environments. *Cognition*, 65(1), 33–69. doi:10.1016/S0010-0277(97)00040-1
Bosch, L., & Sebastián-Gallés, N. (2003a). Language experience and the perception of a voicing contrast in fricatives: Infant and adult data. *Proceedings of the 15th International Congress of Phonetic Sciences* – Barcelona, 1987–1990.
Bosch, L., & Sebastián-Gallés, N. (2003b). Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. *Language and Speech*, 46(2–3), 217–243. doi:10.1177/00238309030460020801
Brent, M. R., & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, 81, B33–B44. doi:10.1016/S0010-0277(01)00122-6
Butler, J., Floccia, C., Goslin, J., & Panneton, R. (2011). Infants' discrimination of familiar and unfamiliar accents in speech. *Infancy*, 16(4), 392–417. doi:10.1111/j.1532-7078.2010.00050.x
Byers-Heinlein, K., Burns, T. C., & Werker, J. F. (2010). The roots of bilingualism in newborns. *Psychological Science*, 21(3), 343–348. doi:10.1177/0956797609360758
Cabrera, L., Tsao, F.-M., Liu, H.-M., Li, L.-Y., Hu, Y.-H., Lorenzi, C., & Bertoncini, J. (2015). The perception of speech modulation cues in lexical tones is guided by early language-specific experience. *Frontiers in Psychology*, 6, 1290. doi:10.3389/fpsyg.2015.01290
Chen, A., & Kager, R. (2016). Discrimination of lexical tones in the first year of life. *Infant and Child Development*, 25(5), 426–439.
Christophe, A., Guasti, M. T., Nespor, M., Dupoux, E., & Van Ooyen, B. (1997). Reflections on phonological bootstrapping: Its role for lexical and syntactic acquisition. *Language and Cognitive Processes*, I2(5–6), 585–612. doi:10.1080/016909697386637
Curtin, S., Mintz, T. H., & Christiansen, M. H. (2005). Stress changes the representational landscape: Evidence from word segmentation. *Cognition*, 96(3), 233–262. doi:10.1016/j.cognition.2004.08.005
Delattre, P. (1938). L'accent final en français: Accent d'intensité, accent de hauteur, accent de durée. *French Review*, 12(2), 141–145.
Di Cristo, A. (1998). Intonation in French. In D. Hirst, & A. Di Cristo (Eds.), *Intonation systems: A survey of twenty languages* (pp. 195–218). New York, NY: Cambridge University Press.
Dupoux, E., Pallier, C., Sebastián-Gallés, N., & Mehler, J. (1997). A destressing "deafness" in French? *Journal of Memory and Language*, 36, 406–421. doi:10.1006/jmla.1996.2500
Dupoux, E., Peperkamp, S., & Sebastián-Gallés, N. (2001). A robust method to study stress "deafness." *The Journal of the Acoustical Society of America*, 110(3), 1606–1618. doi:10.1121/1.1380437
Dutoit, T., Pagel, V., Pierret, N., Bataille, F., & van der Vrecken, O. (1996). The MBROLA project: Towards a set of high quality speech synthesizers free of use for non commercial purposes. *Proceedings of the Fourth International Conference on Spoken Language Processing* – Philadelphia, ICSLP '96, Vol. 3, 1393–1396.

Echols, C. H., Crowhurst, M. J., & Childers, J. B. (1997). The perception of rhythmic units in speech by infants and adults. *Journal of Memory and Language*, 36(2), 202–225. doi:10.1006/jmla.1996.2483

Floccia, C., Keren-Portnoy, T., DePaolis, R., Duffy, H., Luche, C. D., Durrant, S., White, L., Goslin, J, & Vihman, M. M. (2016). British English infants segment words only with exaggerated infant-directed speech stimuli. *Cognition*, 148(C), 1–9. doi:10.1016/j.cognition.2015.12.004

Friederici, A., Friedrich, M., & Christophe, A. (2007). Brain responses in 4-month-old infants are already language specific. *Current Biology*, 17(14), 1208–1211. doi:10.1016/j.cub.2007.06.011

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, 1(1), 3–55. doi:10.1207/s15327817la0101_2

Grosjean, F. (2010) *Bilingual: Life and reality*. Cambridge, MA: Harvard University Press. doi:10.4159/9780674056459

Goyet, L., de Schonen, S., & Nazzi, T. (2010). Words and syllables in fluent speech segmentation by French-learning infants: An ERP study. *Brain Research*, 1332, 75–89. doi:10.1016/j.brainres.2010.03.047

Goyet, L., Millotte, S., Christophe, A., & Nazzi, T. (2016). Processing continuous speech in infancy: From major prosodic units to isolated word forms. In J. Lidz, J. Pater, & W. Snyder (Eds.), *Oxford handbook of developmental linguistics* (pp. 133–153). Oxford: Oxford University Press.

Goyet, L., Nishibayashi, L.-L., & Nazzi, T. (2013). Early syllabic segmentation of fluent speech by infants acquiring French. *PLoS ONE*, 8(11), e79646. doi:10.1371/journal.pone.0079646

Guasti, M. T., Nespor, M., Christophe, A., & Van Ooyen, B. (2001). Pre-lexical setting of the head-complement parameter through prosody. In J. Weissenborn, & B. Höhle (Eds.), *Approaches to bootstrapping: Phonological, lexical, syntactic and neurophysiological aspects of early language acquisition* (Vol. 1, pp. 231–248). Amsterdam: John Benjamins. doi:10.1075/lald.23.14gua

Hay, J. S. F., & Saffran, J. R. (2012). Rhythmic grouping biases constrain infant statistical learning. *Infancy*, 17(6), 610–641. doi:10.1111/j.1532-7078.2011.00110.x

Hayes, B. (1995). *Metrical stress theory: Principles and case studies*. Chicago IL: The University of Chicago Press.

Hirsh-Pasek, K., Kemler Nelson, D. G., Jusczyk, P. W., Cassidy, K. W., Druss, B., & Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition*, 26(3), 269–286. doi:10.1016/S0010-0277(87)80002-1

Höhle, B., Bijeljac-Babic, R., Herold, B., Weissenborn, J., & Nazzi, T. (2009). Language specific prosodic preferences during the first half year of life: Evidence from German and French infants. *Infant Behavior & Development*, 32(3), 262–274. doi:10.1016/j.infbeh.2009.03.004

Höhle, B., Pauen, S., Hesse, V. & Weissenborn, J. (2014). Discrimination of rhythmic pattern at 4 months and language performance at 5 years: A longitudinal analysis of data from German-learning children. *Language Learning*, 64, 141–164. doi:10.1111/lang.12075

Houston, D. M., Jusczyk, P. W., Kuijpers, C., Coolen, R., & Cutler, A. (2000). Cross-language word segmentation by 9-month-olds. *Psychonomic Bulletin & Review*, 7(3), 504–509. doi:10.3758/BF03214363

Iversen, J. R., Patel, A. D., & Ohgushi, K. (2008). Perception of rhythmic grouping depends on auditory experience. *The Journal of the Acoustical Society of America*, 124, 2263–2271. doi:10.1121/1.2973189

Johnson, E. K., Seidl, A., & Tyler, M. D. (2014). The edge factor in early word segmentation: Utterance-level prosody enables word form extraction by 6-month-olds. *PLoS ONE*, 9(1), e83546. doi:10.1371/journal.pone.0083546

Jun, S. A (1996): *The phonetics and phonology of Korean prosody: Intonational phonology and the prosodic hierarchy*. New York, NY: Garland

Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development*, 64, 675–687. doi:10.2307/1131210

Jusczyk, P. W., Hirsh-Pasek, K., Nelson, D. G. K., Kennedy, L. J., Woodward, A., & Piwoz, J. (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology*, 24(2), 252–293. doi:10.1016/0010-0285(92)90009-Q

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, 39(3–4), 159–207. doi:10.1006/cogp.1999.0716

Jusczyk, P. W., & Thompson, E. (1978). Perception of a phonetic contrast in multisyllabic utterances by 2-month-old infants. *Perception & Psychophysics*, 23(2), 105–109. doi:10.3758/BF03208289

Kooijman, V., Hagoort, P., & Cutler, A. (2005). Electrophysiological evidence for prelinguistic infants' word recognition in continuous speech. *Cognitive Brain Research*, 24(1), 109–116. doi:10.1016/j.cogbrainres.2004.12.009

Kooijman, V., Hagoort, P., & Cutler, A. (2009). Prosodic structure in early word segmentation: ERP evidence from Dutch ten-month-olds. *Infancy*, 14(6), 591–612. doi:10.1080/15250000903263957

Liu, L., & Kager, R. (2014). Perception of tones by infants learning a non-tone language. *Cognition*, 133(2), 385–394. doi:10.1016/j.cognition.2014.06.004

Liu, L., & Kager, R. (2015). Bilingual exposure influences infant VOT perception. *Infant Behavior and Development*, 38, 27–36. doi:10.1016/j.infbeh.2014.12.004

Männel, C., Schipke, C. S., & Friederici, A. D. (2013). The role of pause as a prosodic boundary marker: Language ERP studies in German 3- and 6-year-olds. *Developmental Cognitive Neuroscience*, 5, 86–94. doi:10.1016/j.dcn.2013.01.003

Mattock, K., & Burnham, D. (2006). Chinese and English infants' tone perception: Evidence for perceptual reorganization. *Infancy*, 10(3), 241–265. doi:10.1207/s15327078in1003_3

Mattock, K., Molnar, M., Polka, L., & Burnham, D. (2008). The developmental course of lexical tone perception in the first year of life. *Cognition*, 106(3), 1367–1381. doi:10.1016/j.cognition.2007.07.002

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. doi:10.1016/0010-0277(88)90035-2

Molnar, M., Carreiras, M., & Gervain, J. (2016). Language dominance shapes non-linguistic rhythmic grouping in bilinguals. *Cognition*, 152(C), 150–159.

Molnar, M., Gervain, J., & Carreiras, M. (2013). Within-rhythm class native language discrimination abilities of Basque-Spanish monolingual and bilingual infants at 3.5 months of age. *Infancy*, 19(3), 326–337. doi:10.1111/infa.12041

Molnar, M., Lallier, M., & Carreiras, M. (2014). The amount of language exposure determines nonlinguistic tone grouping biases in infants from a bilingual environment. *Language Learning*, 64(s2), 45–64. doi:10.1111/lang.12069

Moon, C., Panneton Cooper, R., & Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant Behavior & Development*, 16, 495–500. doi:10.1016/0163-6383(93)80007-U

Morgan, J., & Demuth, K. (1996). Signal to syntax: An overview. In J. Morgan, & K. Demuth (Eds.). *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 1–24.) Mahwah, NJ: Lawrence Erlbaum Associates.

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. doi:10.1037/0096-1523.24.3.756

Nazzi, T., Floccia, C., & Bertoncini, J. (1998). Discrimination of pitch contours by neonates. *Infant Behavior and Development*, 21(4), 779–784. doi:10.1016/S0163-6383(98)90044-3

Nazzi, T., Iakimova, G., Bertoncini, J., Frédonie, S., & Alcantara, C. (2006). Early segmentation of fluent speech by infants acquiring French: Emerging evidence for crosslinguistic differences. *Journal of Memory and Language*, 54(3), 283–299. doi:10.1016/j.jml.2005.10.004

Nazzi, T., Jusczyk, P. W., & Johnson, E. K. (2000). Language discrimination by English-learning 5-month-olds: Effects of rhythm and familiarity. *Journal of Memory and Language*, 43(1), 1–19. doi:10.1006/jmla.2000.2698

Nazzi, T., Kemler Nelson, D. G., Jusczyk, P. W., & Jusczyk, A. M. (2000). Six-month-olds' detection of clauses embedded in continuous speech: Effects of prosodic well-formedness. *Infancy*, 1(1), 123–147. doi:10.1207/S15327078IN0101_11

Nazzi, T., Mersad, K., Sundara, M., Iakimova, G., & Polka, L. (2014). Early word segmentation in infants acquiring Parisian French: Task-dependent and dialect-specific aspects. *Journal of Child Language*, 41(3), 600–633. doi:10.1017/S0305000913000111

Nishibayashi, L. L., Goyet, L., & Nazzi, T. (2015). Early speech segmentation in French-learning infants: Monosyllabic words versus embedded syllables. *Language and Speech*, 58(3), 334–350. doi:10.1177/0023830914551375

Nespor, M., Guasti, M. T., & Christophe, A. (1996). Selecting word order: The rhythmic activation principle. In U. Kleinhenz (Ed.), *Interfaces in phonology* (pp. 1–26). Berlin: Akademie Verlag.

Pelucchi, B., Hay, J. S. F., & Saffran, J. R. (2009). Learning in reverse: Eight-month-old infants track backward transitional probabilities. *Cognition*, 113(2), 244–247. doi:10.1016/j.cognition.2009.07.011

Peperkamp, S., Vendelin, I., & Dupoux, E. (2010). Perception of predictable stress: A cross-linguistic investigation. *Journal of Phonetics*, 38(3), 422–430. doi:10.1016/j.wocn.2010.04.001

Polka, L., & Sundara, M. (2012). Word segmentation in monolingual infants acquiring Canadian English and Canadian French: Native language, cross-dialect, and cross-language comparisons. *Infancy*, 17, 198–232. doi:10.1111/j.1532-7078.2011.00075.x

Pons, F., & Bosch, L. (2010). Stress pattern preference in Spanish-learning infants: The role of syllable weight. *Infancy*, 15(3), 223–245. doi:10.1111/j.1532-7078.2009.00016.x

Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science*, 12(2), 339–349. doi:10.1111/j.1467-7687.2008.00786.x

Sansavini, A., Bertoncini, J., & Giovanelli G. (1994). Newborns discriminate stress patterns in phonetically complex Italian words. *Infant Behavior and Development*, 17, 924–924.

Sansavini, A., Bertoncini, J., & Giovanelli, G. (1997). Newborns discriminate the rhythm of multisyllabic stressed words. *Developmental Psychology*, 33(1), 3–11. doi:10.1037/0012-1649.33.1.3

Sato, Y., Sogabe, Y., & Mazuka, R. (2010). Development of hemispheric specialization for lexical pitch-accent in Japanese infants. *Journal of Cognitive Neuroscience*, 22(11), 2503–2513. doi:10.1162/jocn.2009.21377

Sato, Y., Utsugi, A., Yamane, N., Koizumi, M., & Mazuka, R. (2013). Dialectal differences in hemispheric specialization for Japanese lexical pitch accent. *Brain and Language*, 127(3), 475–483. doi:10.1016/j.bandl.2013.09.008

Seidl, A. (2007). Infants' use and weighting of prosodic cues in clause segmentation. *Journal of Memory and Language*, 57(1), 24–48. doi:10.1016/j.jml.2006.10.004

Seidl, A. & Cristià, A. (2008). Developmental changes in the weighting of prosodic cues. *Developmental Science*, 11(4), 596–606. doi:10.1111/j.1467-7687.2008.00704.x

Segal, O. & Kishon-Rabin, L. (2012). Evidence for language-specific influence on the preference of stress patterns in infants learning an Iambic language (Hebrew). *Journal of Speech, Language, and Hearing Research*, 55, 1329–1341. doi:10.1044/1092-4388(2012/11-0087)

Singh, L., & Foong, J. (2012). Influences of lexical tone and pitch on word recognition in bilingual infants. *Cognition*, 124(2), 128–142. doi:10.1016/j.cognition.2012.05.008

Skoruppa, K., Cristià, A., Peperkamp, S., & Seidl, A. (2011). English-learning infants' perception of word stress patterns. *Journal of the Acoustical Society of America*, 130(1), EL50–EL55. doi:10.1121/1.3590169

Skoruppa, K., Pons, F., Bosch, L., Christophe, A., Cabrol, D., & Peperkamp, S. (2013). The development of word stress processing in French and Spanish infants. *Language Learning and Development*, 9(1), 88–104. doi:10.1080/15475441.2012.693881

Skoruppa, K., Pons, F., Christophe, A., Bosch, L., Dupoux, E., Sebastián-Gallés, N., Peperkamp, S. (2009). Language-specific stress perception by 9-month-old French and Spanish infants. *Developmental Science*, 12(6), 914–919. doi:10.1111/j.1467-7687.2009.00835.x

Soderstrom, M., Seidl, A., Kemler Nelson, D. G., & Jusczyk, P. W. (2003). The prosodic bootstrapping of phrases: Evidence from prelinguistic infants. *Journal of Memory and Language*, 49(2), 249–267. doi:10.1016/S0749-596X(03)00024-X

Sundara, M., Polka, L., & Molnar, M. (2008). Development of coronal stop perception: Bilingual infants keep pace with their monolingual peers. *Cognition*, 108(1), 232–242. doi:10.1016/j.cognition.2007.12.013

van der Hulst, H. (1999). *Word prosodic systems in the languages of Europe*. Berlin: De Gruyter Mouton.

Weikum, W. M., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastian-Gallés, N., Werker, J. F., (2007). Visual language discrimination in infancy. *Science*, 316(5828), 1159–1159. doi:10.1126/science.1137686

Weissenborn, J., & Höhle, B. (2001). *Approaches to bootstrapping: Phonological, lexical, syntactic and neurophysiological aspects of early language acquisition*. Amsterdam: John Benjamins.

Wellmann, C., Holzgrefe, J., Truckenbrodt, H., Wartenburger, I. & Höhle, B. (2012). How each prosodic boundary cue matters: Evidence from German infants. *Frontiers in Psychology*, 3, 580, 1–13.

Yeung, H. H., Chen, K. H., & Werker, J. F. (2013). When does native language input affect phonetic perception? The precocious case of lexical tone. *Journal of Memory and Language*, 68(2), 123–139. doi:10.1016/j.jml.2012.09.004

Yoshida, K. A., Iversen, J. R., Patel, A. D., Mazuka, R., Nito, H., Gervain, J., & Werker, J. F. (2010). The development of perceptual grouping biases in infancy: A Japanese-English cross-linguistic study. *Cognition*, 115(2), 356–361. doi:10.1016/j.cognition.2010.01.005

CHAPTER 4

The role of prosody in early word learning

Behavioral evidence

Jill C. Thorson
University of New Hampshire

Successful word learning in first language acquisition requires three components: segmentation of words from a continuous speech stream, identification of real-word referents or concepts, and mapping of linguistic label to meaning. This review presents behavioral evidence for how prosody is integral to each of these processes, with particular prosodic structures creating optimal word learning environments. Different prosodic strategies aid the three facets of word learning, with rhythm, stress, intonation, and phrasing interacting to facilitate learning. For example, recent behavioral work shows that word learning is enhanced when words appear at prosodic phrase boundaries and when new words carry a more prominent pitch accent. Finally, an integrative account of word learning is discussed in light of these findings.

Introduction

Most people can tell you what their first word was, or at least what their parents told them it was (mine was some form of the word *grapes*). As adults we give little thought to the processes that occur both in the mind and in the environment to make early word learning successful. Words are one of the most basic elements in language that carry meaning, forming the building blocks for communication. The process of word (or sign) learning is not a trivial task, yet nearly every infant successfully overcomes this challenge and learns the fundamental elements of their native language(s). If thought of as a unified process, word learning is a daunting and mysterious part of development. Even as adults, when you hear a foreign language it is difficult to know what a word is, with the gaps in speech not necessarily correlating with the gaps that we see in written form. Beyond the challenge of knowing where a word begins and ends, Quine (1960) aptly describes the difficulty of identifying the referent of a new and foreign

DOI 10.1075/tilar.23.04tho

word. In his example, a native guide exclaims *gavagai* in an unknown language while pointing to a nearby rabbit. The task of the non-native speaker in this exchange is to determine the meaning of this word. While some interpretations of *gavagai* are more likely than others (e.g., *rabbit, look*), the actual possibilities of meanings are infinite (e.g., *that*, *white*, *over there*, *dinner*, etc.), making the word to meaning mapping task far from simple. In first language acquisition, where this process occurs without overt or explicit instruction, it is useful to consider word learning in terms of smaller, simultaneously-learned components coming together to facilitate learning. The infant must be able to extract sounds from the speech signal (*word segmentation*), identify real-word referents and actions (*identification*), and map sound sequences onto meaning (*mapping*). These three components of segmentation, identification, and creating the mapping between words and meaning are not sequential but rather they simultaneously build off one another, each playing an important role in the overall task of infant word learning.

What processes and mechanisms aid infants and toddlers in acquiring their first words? A variety of factors from the linguistic, cognitive, and social domains contribute to early word learning and affect an infant's ability to learn the names of objects and actions in their environment, setting the groundwork for the acquisition of syntax and grammar. These unique mechanisms impact word learning to different degrees, with particular strategies or mechanisms becoming differentially weighted over the course of development (e.g., the role of infant-directed speech). An integrative account of word learning is necessary to account for how each of these factors comes to together to promote infant word learning (Diesendruck, 2009). One critical aspect of speech and language that integrates with linguistic, social, and cognitive factors alike is prosody (see Waxman & Lidz, 2006, for a thorough review on how linguistic, conceptual, and social processes interact to support early word learning).

Prosody is the rhythm and melody (i.e., cadence, intonation) of speech, including patterns of tone, stress, and intonation. Prosodic features are suprasegmental–properties of speech units that are larger than an individual phonetic segment–and typically associate to the syllable (or mora) unit or to higher levels in the prosodic hierarchy (Beckman & Pierrehumbert, 1986; Nespor & Vogel, 1986; Pierrehumbert & Beckman, 1988; Selkirk, 1978, 1986). Prosody is an essential part of successful communication as it conveys lexical, semantic, pragmatic, and affective information (Ladd, 1996). In speech, prosodic variation is signaled by changes in pitch (fundamental frequency or F_0), length of sounds (duration), loudness (intensity), and segmental quality and reduction (spectral characteristics) (Bolinger, 1989; Lehiste, 1970; Shattuck-Hufnagel & Turk, 1996).[1] From these variations in the acoustics, prosody communicates essential information regarding lexical distinctions, meaning, context, and emotion to the listener.

The goal of this chapter is to disentangle the specific role of prosody during early word learning from other co-occurring social and cognitive processes. However, since prosody expresses pragmatic intent and semantic meaning, it naturally extends into non-linguistic mechanisms that influence early word learning. Accordingly, a theoretical account of word learning that includes prosody is naturally integrative. Prosody plays an essential role in this process by aiding the infant in each component of word learning. In this chapter, the three primary sections will demonstrate how rhythm and stress affect word segmentation, how intonation aids with referent identification, and how intonation and phrasal boundaries facilitate the word to meaning mapping. All evidence presented will be exclusively from behavioral (and eye-tracking) studies (for a review of electrophysiological evidence that supports the role of prosody in word segmentation and learning, see Teixidó, François, Bosch, & Männel, this volume). While a truly integrative theoretical account of word learning acknowledges that there are other linguistic, social, and cognitive mechanisms at work during this developmental period, the objective of this chapter is to identify how prosody affects the learning process and to uncover how the infant learner is able to complete the tasks necessary for successful word learning.

Word learning components and the role of prosody

As mentioned, successful word learning is dependent on three primary components that occur simultaneously over development: (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 meaning (action, concept, etc.) (Waxman & Lidz, 2006). The following sections explore the role that prosody plays during each of these early word learning processes.

Segmentation

The breaks or pauses that occur in written language are not present in continuous speech. If looking at a speech waveform, any visible pauses do not necessarily

---

1. How prosody influences language learning in signed languages is outside the scope of this chapter. See recent work by Brentariand colleagues for a discussion on how prosody is acquired inAmerican Sign Language (Brentari, Falk, & Wolford, 2015).

correspond with the beginnings and ends of words, but rather can be word-internal articulation pauses created by particular consonants (e.g., plosives). Additionally, the actual word boundaries may not be signaled with a pause at all, and most words in fluent speech exhibit no acoustic separation with the previous or following word. Since pauses do not reliably signal the beginning and ends of words in fluent speech, how is the infant able to begin to segment speech into units? The infant must use accessible perceptual information to segment a multi-word stream of speech. Linguistic structures begin to emerge, with linguistic and cognitive mechanisms interacting and giving rise to the infant being able to determine where the word boundaries exist. As these processes develop, the once fluent speech unit becomes more detailed and native-language specific in the perception of the infant learner.

An early sensitivity to prosody sets the foundation for its role in later word segmentation and learning. Rhythmic and fundamental frequency information are transmitted in utero, making prosody one of the first facets of language to which the infant is exposed. This early exposure is demonstrated by newborn infants' preference for listening to their mother's voice versus the voice of an unfamiliar woman (DeCasper & Fifer, 1980) and for a familiar story read to them in the womb over an unfamiliar story (DeCasper & Spence, 1986). Infants as young as two days old are sensitive to both native language rhythm and pitch patterns. For example, they can discriminate languages by their broad rhythmic class patterns (Moon, Panneton-Cooper, & Fifer, 1993; Nazzi, Bertoncini, & Mehler, 1998). That is, a newborn infant is able to distinguish a syllable-timed language such as French from 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 (Jusczyk & Thompson, 1978; Nazzi, Floccia, & Bertoncini, 1998). These early perceptual sensitivities to prosody are important for assisting the infant learner in segmenting the speech stream into meaningful units (Werker & Yeung, 2005). Debate still surrounds what types of cues are most influential and primarily utilized for word extraction in infancy–prosodic cues (Johnson & Seidl, 2009) or domain-general mechanisms based on frequency and distributional information (e.g., statistical learning, Aslin, Saffran, & Newport, 1998; Lew-Williams & Saffran, 2012; Saffran, Aslin, & Newport, 1996).

Early sensitivities to stress, rhythm, and melody of language (i.e., prosody) aid the infant in detecting word boundaries. One of the broadest prosodic elements to aid word segmentation is the final boundary, indicated by final lengthening and a clear phrase boundary. For example, research from both French and English shows that a word at the end of a phrase is more easily recognized and also segmented (Johnson, Seidl, & Tyler, 2014; Seidl & Johnson, 2006; for more on how phrasal prosody aids in word segmentation, see de Carvalho, Dautriche, Millotte, & Christophe, this volume). Languages also differ from one another in how they lexically stress words, with patterns varying in the placement of strong and weak speech units (e.g., foot or syllable) (for more on the processing and acquisition of lexical stress across languages, see Bhatara, Boll-Avestisyan, Höhle, & Nazzi, this volume). A disyllable that has a strong-weak stress pattern is called a trochee (e.g., happy, cuddle), while a weak-strong stress pattern is an iamb (e.g., awake, return). Even though languages often contain both types, one pattern usually predominates. In English, the dominant pattern is to place linguistic stress on the first syllable, with 90% of multisyllabic words displaying this type of pattern (Cutler & Carter, 1987). Infants between six and nine months use rhythmic information as a means to group pairs of syllables (Morgan, 1994; Morgan & Saffran, 1995). By nine-months-old, English-learning infants demonstrate a rhythmic bias for familiar and novel trochaic disyllables (Jusczyk, Cutler, & Redanz, 1993; Morgan, 1996).

Typical prosodic patterns and words with predictable stress patterns are more readily segmented. Like adults, English-acquiring nine-month-olds listen longer to stimuli where syllables form strong-weak patterns, demonstrating a preference for strong syllable onsets (Cutler & Butterfield, 1992; Cutler & Norris, 1988; Echols, Crowhurst, & Childers, 1997). Jusczyk, Houston, and Newsome (1999) show that 7.5-month-old infants correctly segment disyllabic words that have a trochaic pattern in English, but not words that are iambic until several months later. In this study, they familiarized infants with isolated repetitions of two words that were either trochaic or iambic (Jusczyk et al., 1999). At test, the infants were played passages that either did or did not include the familiarized disyllabic pattern. Average listening time to the familiarized item versus the test item was measured. Infants listened longer to the trochaic patterns, showing that they were able to segment trochaic disyllables from the speech stream. Follow up experiments were conducted to find out if preference was driven by the strong syllable alone. These experiments showed that the disyllabic strong-weak pattern–not the initial strong syllable alone–guides infant behavior, and that trochaic patterns are preferred even when across word boundaries (e.g., "guitar is"). In this case, infants segment the sequence "taris" not "guitar", preferring the trochaic foot to the whole word iamb. By nine months, infants use lexical stress to start detailing the structure of their language.

Languages vary in their rhythmic patterns and distributions (Saffran, Newport, & Aslin, 1996), making the trochaic preference found in English and also German a learned process developing over the first nine months of life (Höhle, Bijeljac-Babic, Herold, Weissenborn, & Nazzi, 2009; Thiessen & Saffran, 2007). Segmentation abilities continue to improve over development. By 10 months, English-acquiring infants can also segment iambs in a similar manner as trochees

(Gerken & Aslin, 2005). Cross-linguistic research illuminates how other languages take advantage of the rhythmic properties of speech. For example, languages can have fixed stress, such as Finnish and Polish where stress almost always falls on the first syllable of a word, or Czech and Quechua where stress is typically on the penultimate syllable (Jusczyk, 1999). Infants learning these languages have access to predictable patterns that guide perception and segmentation. In the case of French, which is considered to have no lexical stress, there is consistent phrase final accentuation. Following this pattern, French-acquiring infants are shown to segment iambs more readily from the speech stream due to this phrase final accent (Polka, Sundara, & Blue, 2002). Other languages that exhibit less predictable lexical stress (e.g., Cantonese, Spanish) do not provide rhythmic cues to infants in the same way. In these cases, processes related to native phonotactics, transitional probabilities between segments, or syllable weight may provide more productive strategies or interact with prosody in a distinct manner (Pons & Bosch, 2010). Cross-linguistic research is still limited and there is no clear evidence as to which mechanisms are most influential (see Goyet, Millotte, Christophe, & Nazzi, 2016, for a recent review of the role of rhythm and word form segmentation; see also Bhatara, Boll-Avestisyan, Höhle, & Nazzi, this volume). Future work needs to fill in the gaps for how infants use the features of their native language to facilitate early word segmentation, and how these may change over development. Overall, researchers agree that while particular strategies may be more influential than others, there is no one single factor that is sufficient to explain how word segmentation is achieved.

Beyond stress and accentuation patterns that give rise to rhythmic preferences and later word segmentation cues, other aspects of prosody are useful in early word learning. The speech used by caregivers to infants is known as infant-directed speech (IDS) and exhibits unique prosodic characteristics in comparison to typical adult-directed speech (ADS) prosody. IDS has been identified as having an increase in overall pitch, pitch range, segmental and pause duration (i.e., slower speaking rate), and intensity, and is preferred over ADS by young children (Fernald, 1985; Fernald, 1989; Fernald & Simon, 1984). Additionally, IDS has an increased use of specific intonational contours such as the bitonal rising pitch movement–L+H* pitch accent in the ToBI transcription system–used to express focus in English (Yu, Khan, & Sundara, 2014), which is more prominent than a monotonal pitch accent (Baumann & Roth, 2014; Turnbull, Royer, Ito, & Speer, 2014). Fernald and Mazzie (1991) show that these focus-type accents are exaggerated in IDS and may facilitate speech processing for the infant. The slower speaking rate of IDS also has been shown to improve word recognition at 19 months (Song, Demuth, & Morgan, 2010). These types of cues present in IDS aid seven-month-olds' word segmentation in comparison to adult-directed speech prosody

(Thiessen, Hill, & Saffran, 2005). In this study, seven-month-olds were familiarized with novel words that were produced in either adult-directed or infant-directed speech. For each type, the stimuli had prosodic characteristics representative of that register. At test, infants listened longer to words over part words in the IDS condition, but exhibited no difference in listening in the ADS condition. Results indicate that infants can more easily segment whole words from the speech stream when they are produced with IDS prosody.

Interacting with prosody, there are other factors that are involved in early word segmentation. As the infant begins to fill in phonetic information of their native language, they are able to segment words using the phonotactics–permissible combination of sounds or phonemes–of their native language as early as 7.5-months-old (Jusczyk & Aslin, 1995) as well as the phonotactics and prosody together (Mattys & Jusczyk, 2001; Mattys, Jusczyk, Luce, & Morgan, 1999; Morgan, 1996). Also, highly familiar words such as 'mommy' or the child's own name serve as anchors for segmentation (Bortfeld, Morgan, Golinkoff, & Rathbun, 2005). An important breakthrough with regard to how language is acquired is found in the statistical regularity of language. Adjacent transitional probabilities within and between words support word segmentation. Eight-month-old infants use the statistical structure of a synthetic language to form basic word-like units from the syllables, which only vary in the transitional probabilities between syllables (Aslin et al., 1998; Saffran, Aslin, & Newport, 1996). Expectations about word length also interact with statistical regularities and aid the infant in word segmentation (Lew-Williams & Saffran, 2012). Finally, segmentation is not necessary for words produced in isolation. These words are prosodically distinct elements with clear boundaries demarcated by silence that are more readily learned (Brent & Siskind, 2001). How prosody interacts with word segmentation across languages is still being explored, with cross-linguistic research beginning to fill in valuable information about the ways in which infants employ the unique sets of cues needed for segmenting words.

In finding the right methodology to explore a research question, there is a push and pull between data collected in a natural setting versus data collected in a controlled laboratory experiment. In word segmentation studies, the need for control over stimuli and data collection is essential. Many studies often employ novel words or artificial languages as the source of stimuli because they can be more closely controlled to test for specific effects. While this allows researchers to pinpoint particular mechanisms that influence word segmentation, it encourages future studies to more closely approximate real-life language contexts and interactions. Since languages differ in their prosodic structures, cross-linguistic research has offered, and continues to offer, approaches in which to uncover alternative strategies and cues that are used in early word segmentation.

Referent/object identification

Matching meaning to a linguistic label is the ultimate goal of word learning. Referent (and/or object) identification is a key component to this process. In a live context, nouns may refer to an overt referent (and verbs to an action being performed by a referent). An infant must be able to locate the referent (or action) in their environment to create the successful mapping from the linguistic label to its real-world meaning. As with word segmentation, linguistic, conceptual, and social processes all aid in the identification of real-world referents. Alongside prosody, joint attention, eye gaze, and gesture (e.g., pointing) are some of the ways in which a caregiver can help an infant in locating and identifying objects in their environment (Brooks & Meltzoff, 2002; Moore, Angelopoulos, & Bennett, 1999; Samuelson & Smith, 1998; Woodward, 2003). In the case of prosody, even intonation alone can assist in directing the attention of an infant and guide them to the correct referent of a word in their environment. This section will address the role of prosody in referent and object identification and will primarily present data from toddlers and young children (18 months to 5 years). While the ages in this section are generally older than the ones discussed in the word segmentation section, this does not imply that these processes are sequential. Rather, the young language learner is simultaneously being exposed to and learning each facet of word learning.

In adult speech processing, intonation is one aspect of the speech stream that can be used to identify referents in discourse. Adult listeners are able to make use of the accentuation of a noun to determine its referent (Dahan, Tanenhaus, & Chambers, 2002; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). For example, specific pitch accents can speed up processing in contextually appropriate environments. Weber, Braun, and Crocker (2006) show that German adult listeners are able to rapidly take advantage of prosodic information to determine the referent of an adjective-noun combination. Results revealed faster looking to a referent in a contrastive setting when the linguistic label carried a contrastive pitch accent over no pitch accent at all. Like adults, infants are also able to exploit prosodic information to support referent identification.

Young children make use of accented and deaccented lexical information to determine reference. Arnold (2008) looked at the interpretation of accented and deaccented 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. In the experiment, they presented participants with a screen displaying four items, two of which were linguistic cohort competitors that overlapped in their initial segments (e.g., <u>ba</u>gel/<u>ba</u>con). 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 new information is typically accented, and given information deaccented, this experiment manipulated the accentuation of the second (*given*) production of a target word. This study showed that both adults and four- and five-year-olds exhibit a bias to the previously mentioned object when the target in the second utterance was an unaccented noun phrase or pronoun (Arnold, 2008). Accented words in the second utterance did not show any observable bias. Arnold (2008) shows that four- and five-year-olds process deaccented words as identifying previously established referents. By the age of four or five, word learning has already become a more automatic process and referent identification is more accessible. It is important to assess the extent to which prosody interacts with referent identification while the child is still in the early stages of word learning and has had less linguistic experience.

Grassmann and Tomasello (2010) investigated how word stress (i.e., accentuation) 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 the effect of stress and newness. The three conditions were newness only, stress only, and newness and stress combined, with target words presented in short sentences (e.g., *The duck has the* <u>*apple*</u>.). For the newness only condition, a picture of a duck would be presented on screen with the phrase "Look a duck. Look the duck." In a following screen the child would hear "The duck has the apple" (with neither noun accented) and a final picture would appear of a duck and an apple. In the newness only case, the apple was visually new to the context but unaccented. For the stress only condition, both the first and second screen would depict the duck and the apple ("Look the duck and the apple"). For the second picture, the corresponding utterance would accent the word apple ("The duck has the APPLE"). Here, the apple is visually given information in the second screen but it carries an accent. Finally, for the newness and stress condition, the apple was both visually new to the context in the second picture as well as accented ("The duck has the APPLE"). There was no condition that presented the apple as visually given material as well as deaccented. Grassmann and Tomasello (2010) showed that toddlers could be reliably guided to a new target in a scene, but only when the accompanying speech placed a pitch accent on the referenced word. Thus, toddlers only reliably looked at the apple in the newness and stress condition. In an effort to be more ecologically valid, a live experimenter produced the stimuli to the toddler. Although this method may create a more natural setting, one important drawback is that even with training, there is a considerable amount of variability introduced into the stimuli and data. This is true particularly for intonation-based studies, where varying pitch realizations may influence attention. Also, the results contradict studies in visual attention by not showing a novelty bias in the newness only condition.

Typically, by two years old a child will attend to a novel pattern or object (Fantz, 1964; Rose, Gottfried, Melloy-Carminar, & Bridger, 1982). Still, Grassmann and Tomasello (2010) is one of the first studies to look at how both stress and newness of a lexical item influence toddler referent identification in a discourse context. If the presence of prosodic stress impacts this aspect of word learning, a next step is to explore how specific types of contours and pitch accents affect this process. One hypothesis that emerges from this line of inquiry is that pragmatically appropriate pitch accents aid word learning by facilitating word segmentation and conditioning expectations for referent identification.

Recent work by Thorson and Morgan (2014) shows that specific pitch contours interact with new and given information to affect referent resolution. In this study, 18-month-old American English-acquiring toddlers were presented with an online eye-tracking task where the linguistic label of a referent varied in its pitch structure. The familiar word was presented in running speech (in either utterance-medial or -final position) and carried a simple monotonal pitch peak, a bitonal pitch movement, or no pitch excursion. The referential status of the word was either new or given to the discourse context. In contrast to Grassmann and Tomasello (2010), results showed that 18-month-olds' attention to a referent during discourse was driven by the presence of either a pitch movement on the target word or the discourse newness of the referent. Critically, there was increased attention to the referent when it was both contextually new and received a pitch accent, demonstrating how information status and prosody interact to boost attention. Thorson and Morgan (2014) argue that toddlers are more easily able to identify the referent of a word that is more prominent or salient in the context, with intonation increasing salience.

An infant's ability to identify a referent in their environment is an important part of creating the word-to-meaning mapping. Building on research in adult speech processing, recent work with young children demonstrates that there exists an interaction between the prosody of a linguistically labeled referent and its identification in context. More prominent prosodic patterns, such as bitonal pitch accents and IDS prosodic properties, aid in guiding toddler attention to a referent. While other contextual factors are simultaneously helping the infant parse their environment, prosody functions as a way for the caregiver to direct and engage the infant with their surroundings.

Mapping linguistic form to meaning

Beyond making the initial connection between a word and its referent, the ultimate objective of a learner is to create an enduring link between the word and its meaning. Words come in many grammatical varieties and can refer not only to objects, but also to actions or concepts. The infant learner must use multiple cues and mechanisms to determine these connections and create the mapping necessary for successful learning. Since prosody critically interacts with the pragmatics of a social scene (the optimal environment for wording learning to occur), it is important to consider how social-pragmatics affects the word-to-meaning mapping. Tomasello and Akhtar (1995) investigated how the social-pragmatics of a specific context affects word-to-object or word-to-action mapping. This experiment showed that a 27-month-old child exhibits a bias for mapping novel words to objects (over actions), but that this effect could be manipulated by altering the discourse setting (Akhtar, Carpenter, & Tomasello, 1996; Tomasello & Akhtar, 1995). One way in which they manipulated the discourse setting in the second study of Tomasello and Akhtar (1995) was by varying how the experimenter engaged with a nameless object during free-play with the child. In the Action-Highlighted condition the experimenter would repeatedly demonstrate a nameless action with a novel object while saying the novel word ("Widget, Jason, widget" while producing the action). In contrast, the action was only produced once in the Object-Highlighted condition as verbal labeling of the object continued. The children mapped the novel word with the either the verb (Action-Highlighted) or the noun (Object-Highlighted) depending on the context. With respect to intonation, these studies did not address in what manner the utterances were spoken to the participants. Also, since these were live play sessions, a substantial amount of acoustic variability is naturally introduced into the data. This demonstrates the ongoing challenge of striking a balance between ecological validity and experimental control. These studies on novel word learning and social-pragmatics demonstrate that young children are sensitive to changes in discourse newness when assigning meaning to a word. Importantly, substantial research shows that prosody is used to express discourse newness as well as to shape cognitive and social expectations. Accordingly, by looking at how prosody varies within a discourse context and how these variations change patterns of learning, we can better understand the role of prosody in the creation of word-to-meaning mappings.

Like in referent identification, the accentuation of specific words aids a young child during novel word learning. Grassmann and Tomasello (2007) tested two-year-olds' word learning abilities by manipulating the accentuation of novel words during a 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). The stimuli were a set of sentences that each consisted of a German determiner followed by a nonce noun and then a nonce verb. Earlier studies have shown that children 'fast map' novel names with novel nouns (Golinkoff, Hirsh-Pasek, Bailey, & Wenger, 1992; Mervis & Bertrand, 1994). This study took one step further and empirically demonstrated that children can use prosodic highlighting (i.e., accentuation) of new material to match a novel word with a new referent in a live discourse setting (Grassmann & Tomasello, 2007). Another noteworthy aspect of this study is that toddlers were able to match the novel word with the new object when the noun was accented, but they were unable to match an accented novel verb label with a new action. Why this particular inability occurs in this experiment is debatable and may be partly due to the ways in which verb comprehension is tested. Since the earlier studies by Akhtar and colleagues did not take into consideration prosodic highlighting as a factor, the goal of Grassmann and Tomasello (2007) was to analyze the role that prosody plays in relaying intonational cues to what is given and/or new information in a discourse context.

In addition to accentuation, lexical stress influences the association between novel word forms and meaning. In English, switching between trochaic and iambic stress patterns differentiates a certain set of nouns from verbs (e.g., *récord*-noun, *recórd*-verb). Curtin, Campbell, and Hufnagle (2012) exploited this alternation and predicted that 16-month-olds would more readily map iambic novel words to actions over objects, and trochaic novel words to objects. Their predictions were confirmed and they showed that toddlers have developed a bias for iambic disyllables mapping onto actions by 16 months of age. Graf Estes and Bowen (2013) explored the interaction of lexical stress and phonotactics on early word learning. In their study, 19-month-olds were presented with two objects and a corresponding label that varied in its stress (trochaic or iambic) and its phonotactic probability (low or high). Like previous findings, they found that infants only learned words that had trochaic stress, but they additionally showed that the words must also have high phonotactic probabilities. Iambic words were not learned, nor were words with low phonotactic probabilities (Graf Estes & Bowen, 2013).

Aspects of prosodic phrasing also support the process of extracting a word from the speech signal and mapping it to the appropriate referent. Shukla, White, and Aslin (2011) showed that infants as young as six months are able to segment a word from the speech stream and map it with a visual referent if speech is prosodically organized. In their study, infants were only able to map a word to its referent if the word form was aligned with a prosodic phrase boundary (Shukla et al., 2011). Learning was facilitated by the more complex context of producing a word form in continuous speech, showing that infants prefer this prosodically structured input.

More recently, Thorson (2015) investigated the role that specific pitch accents (monotonal H* and bitonal L + H*) play in novel word learning. In an eye-tracking task, two-year-olds were exposed to contexts that presented novel words paired with novel animals. The novel word-referent pair was presented in either a contrastive or noncontrastive setting (i.e., directly in contrast (or not) to a previously presented referent), and was uttered with no pitch accent, a monotonal H* pitch accent, or a bitonal L + H* pitch accent. Learning occurred for both of the pitch accent conditions (H* and L + H*). Interestingly, there was increased looking to the novel animal during the test phase–indicating word to object mapping–when the word was learned with a bitonal pitch accent and during a contrastive situation. This study made use of the L + H* pitch accent, but did not use IDS like many other novel word learning experiments (Thorson, 2015). Still, the contrastive and bitonal L + H* condition proved to be the most salient learning condition, creating a more optimal environment for linking a novel word to a referent. Future work needs to further tease apart the role of contextually appropriate pitch accent usage and how this impacts the early word learning process.

Infant-directed speech (IDS) prosody has been shown to facilitate not only word segmentation, but also the mapping of linguistic labels to meaning. Recent work claims that 21-month-olds more readily learn object labels from IDS than ADS, even if the label was produced in a sentence during learning rather than in isolation (Ma, Golinkoff, Houston, & Hirsh-Pasek, 2011). By 27 months, children were able to learn the novel object label from ADS in addition to IDS (Ma et al., 2011). This demonstrates changes in how children learn across development, with IDS being a preferred strategy at a younger age. In a habituation-based word learning paradigm, Graf Estes and Hurley (2013) tested how 17-month-old toddlers mapped novel words onto novel objects when the novel words were presented in isolation. They found that children only learned the word-to-meaning mapping when stimuli were produced in IDS and where there was prosodic variation across tokens. In the case where prosodically identical IDS tokens were used during habituation, the infants failed to learn the association. Graf Estes and Hurley (2013) illustrates how prosodic variation and salient IDS contours aid the infant learner in creating the word-to-meaning mappings.

Novel word learning tasks reveal how lexical stress and intonation affect the word-to-meaning mapping for young toddlers. Language-dominant lexical patterns, prosodic characteristics of IDS, and prominent accentuation movements all aid in the learning process. The extent to which each strategy is influential and at what point in development each occurs is still under investigation. Future work needs to continue to explore which mechanisms and strategies are most critical at different phases of development, with an emphasis on cross-linguistic study. It is important to look at languages that have differing structures in order to reveal which mechanisms may be universal and which may be language specific.

Integrative approaches to word learning

There are a number of different theoretical approaches and mechanisms that aim to explain how infants and young children acquire words. These accounts are based in the cognitive, linguistic, and social domains, with researchers positing unique descriptions and rationales for how language acquisition occurs, in an effort to demystify word learning. In addition, there are accounts that pull from cognitive, linguistic, and social theories. These integrative theories recognize the complex nature of early word learning and combine different mechanisms into more unified accounts (e.g., Diesendruck, 2009; Yu & Ballard, 2007).

Unique cognitive, social, and linguistic processes and mechanisms influence the way in which children learn words. In the cognitive domain, some argue that attention and memory lie at the root of successful word learning (Samuelson and Smith, 1998). On the linguistic side, others have argued that the manner in which children receive speech input is essential to the fundamental language process, such as the role of discourse novelty (Diesendruck, Markson, Akhtar, & Reudor, 2004) or word accentuation (Grassmann & Tomasello, 2007). Meanwhile, a social-pragmatic account relies on sensitivity to mental states and describes how a child learns words in communicative contexts and follows referential and speaker intent. This theory follows a theory of mind type argument, where the child must be able to relate to the mental state of the speaker during the learning of a novel word (Akhtar et al., 1996). While each of these mechanisms explains a part of the word learning process, integrative accounts attempt to incorporate multiple mechanisms into one overall theoretical approach.

Word learning entails association of linguistic labels with real-world referents, and it combines linguistic, social, and cognitive processes to guide early word recognition and learning. In approaches such as the emergentist coalition model (Hirsh-Pasek, Golinkoff, & Hollich, 2000) and Bloom's (2000) mind, concepts, and syntax theory, each approach involves multiple mechanisms that must combine to create an environment for successful word learning. Prosody integrates with multiple elements within the linguistic domain as well as the social and cognitive domains. It relays semantic, pragmatic, and affective information and naturally crosses cognitive, social, and linguistic mechanistic boundaries. As demonstrated by its role in word segmentation, referent identification, and the mapping of linguistic form to meaning, the inclusion of prosody into the study of early word learning bolsters an integrative theoretical approach to word learning.

Acknowledgements

Many thanks to the volume editors, series editors, and the anonymous reviewers for their comments on an earlier draft of this manuscript. Additional thanks to Katherine Demuth and Glenda Molina Onario for their helpful feedback. The author was supported on a Postdoctoral Fellowship from a National Institutes of Health training grant awarded to Sargent College, Boston University (T32DC013017-01A1).

References

Akhtar, N., Carpenter, M., & Tomasello, M. (1996). The role of discourse novelty in early word learning. *Child Development*, 67, 635–645. doi:10.2307/1131837

Arnold, J. (2008). THE BACON not the bacon: How children and adults understand accented and unaccented noun phrases. *Cognition*, 108, 69–99. doi:10.1016/j.cognition.2008.01.001

Aslin, R. N., Saffran, J., & Newport, E. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, 9(4), 321–324. doi:10.1111/1467-9280.00063

Baumann, S., & Roth, A. (2014). Prominence and coreference: On the perceptual relevance of F0 movement, duration and intensity. *Proceedings of Speech Prosody 7* – Dublin, 227–231.

Beckman, M., & Pierrehumbert, J. B. (1986). Intonational structure in English and Japanese. *Phonology Yearbook*, 3, 255–310. doi:10.1017/S095267570000066X

Bloom, P. (2000). *How children learn the meanings of words*. Cambridge, MA: MIT Press.

Bolinger, D. L. (1989). *Intonation and its uses: Melody in grammar and discourse*. Stanford, CA: Stanford University Press.

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. doi:10.1111/j.0956-7976.2005.01531.x

Brent, M. R., & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, 81, B33–B44. doi:10.1016/S0010-0277(01)00122-6

Brentari, D., Falk, J., & Wolford, G. (2015). The acquisition of prosody in American Sign Language. *Language,* 91(3), 144–168.

Brooks, R., & Meltzoff, A. N. (2002). The importance of eyes: How infants interpret adult looking behavior. *Developmental Psychology*, 38, 958–966. doi:10.1037/0012-1649.38.6.958

Curtin, S., Campbell, J., & Hufnagle, D. (2012). Mapping novel labels to actions: How the rhythm of words guides infants' learning. *Journal of Experimental Child Psychology*, 112, 127–140. doi:10.1016/j.jecp.2012.02.007

Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language*, 31, 218–236. doi:10.1016/0749-596X(92)90012-M

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech & Language*, 2, 133–142. doi:10.1016/0885-2308(87)90004-0

Cutler, A., & Norris, D. G. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 113–121.

Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language*, 47, 292–314. doi:10.1016/S0749-596X(02)00001-3

DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science*, 208, 1174–1176. doi:10.1126/science.7375928

DeCasper, A. J., & Spence, M. J. (1986). Prenatal maternal speech influences newborn's perception of speech sounds. *Infant Behavior and Development*, 9, 133–150. doi:10.1016/0163-6383(86)90025-1

Diesendruck, G. (2009). Mechanisms of word learning. In E. Hoff & M. Shatz (Eds.), *Blackwell handbook of language development* (pp. 257–276). Oxford: Blackwell.

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. doi:10.1111/j.1467-7687.2004.00320.x

Echols, C. H., Crowhurst, M. J., & Childers, J. B. (1997). Perception of rhythmic units in speech by infants and adults. *Journal of Memory and Language*, 36, 202–225. doi:10.1006/jmla.1996.2483

Fantz, R. L. (1964). Visual experience in infants: Decreased attention to familiar patterns relative to novel ones. *Science*, 146, 668–670. doi:10.1126/science.146.3644.668

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development*, 8, 181–195. doi:10.1016/S0163-6383(85)80005-9

Fernald, A. (1989). Intonation and communicative intent in mother's speech to infants: Is the melody the message? *Child Development*, 60(6), 1497–1510. doi:10.2307/1130938

Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology*, 27, 209–221. doi:10.1037/0012-1649.27.2.209

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology*, 20, 104–113. doi:10.1037/0012-1649.20.1.104

Gerken, L., & Aslin, R. N. (2005). Thirty years of research on infant speech perception: The legacy of Peter W. Jusczyk. *Language Learning and Development*, 1(1), 5–21. doi:10.1207/s15473341lld0101_3

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. doi:10.1037/0012-1649.28.1.99

Goyet, L., Millotte, S., Christophe, A., & Nazzi, T. (2016). Processing continuous speech in infancy: From major prosodic units to isolated word forms. In J. Lidz, W. Synder, & J. Pater (Eds.), *Oxford Handbook of Developmental Linguistics* (pp. 133–1560). Oxford: Oxford University Press. doi: 10.1093/oxfordhb/9780199601264.001.0001

Graf Estes, K., & Bowen, S. (2013). Learning about sounds contributes to learning about words: Effects of prosody and phonotactics on infant word learning. *Journal of Experimental Child Psychology*, 114, 405–417. doi:10.1016/j.jecp.2012.10.002

Graf Estes, K., & Hurley, K. (2013). Infant-directed prosody helps infants map sounds to meaning. *Infancy*, 18(5), 797–824. doi:10.1111/infa.12006

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. doi:10.1016/j.pragma.2010.04.019

Hirsh-Pasek, K., Golinkoff, R. M., & Hollich, G. (2000). An emergentist coalition model for word learning: Mapping words to obejcts is a product of the interaction of multiple cues. In R. M. Golinkoff, K. Hirsh-Pasek, L. Bloom, L. B. Smith, A. L. Woodard, N. Akhtar, M. Tomasello, & G. Hollich (Eds.), *Becoming a word learner: A debate on lexical acquisition* (pp. 179–186). New York: Oxford University Press. doi:10.1093/acprof:oso/9780195130324.003.006

Johnson, E. K., & Seidl, A. H. (2009). At 11 months, prosody still outranks statistics. *Developmental Science*, 12(1), 131–141. doi:10.1111/j.1467-7687.2008.00740.x

Johnson, E. K., Seidl, A. H., & Tyler, M. D. (2014). The edge factor in early word segmentation: Utterance-level prosody enables word form extraction by 6-month-olds. *Plos One*, 9(1), e83546. doi:10.1371/journal.pone.0083546

Jusczyk, P. W. (1999). How infants begin to extract words from speech. *Trends in Cognitive Science*, 3(9), 323–328. doi:10.1016/S1364-6613(99)01363-7

Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, 29(1), 1–23. doi:10.1006/cogp.1995.1010
Jusczyk, P. W., Houston, D., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, 39, 159–207. doi:10.1006/cogp.1999.0716
Jusczyk, P. W., Cutler, A., & Redanz, N. (1993). Preference for the predominant stress patters of English words. *Child Development*, 64, 675–687. doi:10.2307/1131210
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. doi:10.3758/BF03208289
Höhle, B., Bijeljac-Babic, R., Herold, B., Weissenborn, J., & Nazzi, R. (2009). Language specific prosodic preferences during the first half year of life: Evidence from German and French infants. *Infant Behavior and Development*, 32(3), 262–274. doi:10.1016/j.infbeh.2009.03.004
Ladd, D. R. (1996). *Intonational phonology*. Cambridge: Cambridge University Press.
Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: The MIT Press.
Lew-Williams, C., & Saffran, J. (2012). All words are not created equal: Expectations about word length guide infant statistical learning. *Cognition*, 122, 241–246. doi:10.1016/j.cognition.2011.10.007
Ma, W., Golinkoff, R., Houston, D., & Hirsh-Pasek, K. (2011). Word learning in infant- and adult-directed speech. *Language Learning and Development*, 7, 209–225. doi:10.1080/15475441.2011.579839
Mattys, S. L., & Jusczyk, P. W. (2001). Do infants segment words or recurring contiguous patterns? *Journal of Experimental Psychology: Human Perception and Performance*, 27(3), 644–655.
Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology*, 38, 465–494. doi:10.1006/cogp.1999.0721
Mervis, C. B., & Bertrand, J. (1994). Acquisition of the novel name-nameless category (N3C) principle. *Child Development*, 65, 1646–1662. doi:10.2307/1131285
Moon, C., Panneton-Cooper, R., & Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant Behavior and Development*, 16, 495–500. doi:10.1016/0163-6383(93)80007-U
Moore, C., Angelopoulos, M., & Bennett, P. (1999). Word learning in the context of referential and salience cues. *Developmental Psychology*, 35(1), 60–68. doi:10.1037/0012-1649.35.1.60
Morgan, J. L. (1994). Converging measures of speech segmentation in preverbal infants. *Infant Behavior and Development*, 17, 389–403. doi:10.1016/0163-6383(94)90031-0
Morgan, J. L. (1996). A rhythmic bias in preverbal speech segmentation. *Journal of Memory and Language*, 35, 666–688. doi:10.1006/jmla.1996.0035
Morgan, J. L., & Saffran, J. R. (1995). Emerging integration of sequential and suprasegmental information in preverbal speech segmentation. *Child Development*, 66, 911–936. doi:10.2307/1131789
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. doi:10.1016/S0163-6383(98)90044-3
Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris.

Pierrehumbert, J., & Beckman, M. (1988). *Japanese tone structure. Linguistics Inquiry Monograph, 15*. Cambridge, MA: The MIT Press.

Polka, L., Sundara, M., & Blue, S. (2002). The role of language experience in word segmentation: A comparison of English, French and bilingual infants. *Paper presented at the 143rd Meeting of the Acoustical Society of America: Special session in memory of Peter Jusczyk*, Pittsburgh, USA.

Pons, F., & Bosch, L. (2010). Stress pattern preference in Spanish-learning infants: The role of syllable weight. *Infancy*, 15(3), 223–245. doi:10.1111/j.1532-7078.2009.00016.x

Quine, W. V. O. (1960). *Word and object*. Cambridge, MA: The MIT Press.

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. doi:10.1037/0012-1649.18.5.704

Saffran, J., Aslin, R. N., & Newport, E. (1996). Statistical learning by 8-month-olds. *Science*, 274(5294), 1926–1928. doi:10.1126/science.274.5294.1926

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. doi:10.1111/j.1467-8624.1998.tb06136.x

Seidl, A., & Johnson, E. K. (2006). Infant word segmentation revisited: Edge alignment facilitates target extraction. *Developmental Science*, 9(6), 565–573. doi:10.1111/j.1467-7687.2006.00534.x

Selkirk, E. O. (1978). On prosodic structure and its relation to syntactic structure. In T. Fretheim (Ed.), *Nordic prosody II* (pp. 111–140). Trondheim: Tapir.

Selkirk, E. O. (1986). On derived domains in sentence phonology. *Phonology*, 3, 371–405. doi:10.1017/S0952675700000695

Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, 25(2), 193–247. doi:10.1007/BF01708572

Shukla, M., White, K. S., & Aslin, R. N. (2011). Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. *Proceedings of the National Academy of Sciences*, 108(15), 6038–3043. doi:10.1073/pnas.1017617108

Song, J. Y., Demuth, K., & Morgan, J. (2010). Effects of the acoustic properties of infant-directed speech on infant word recognition. *Journal of the Acoustical Society of America*, 128(1), 389–400. doi:10.1121/1.3419786

Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information during spoken language comprehension. *Science*, 268, 1632–1634. doi:10.1126/science.7777863

Thiessen, E. D., & Saffran, J. R. (2007). Learning to learn: Infants' acquisition of stress based strategies for word segmentation. *Language, Learning, and Development*, 3(1), 73–100.

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, 7, 49–67. doi:10.1207/s15327078in0701_5

Thorson, J. C. (2015). *The development of intonation and information structure: Perceptual and productive investigations*. Unpublished PhD Dissertation. Brown University.

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

Tomasello, M., & Akhtar, N. (1995). Two-year-olds use pragmatic cues to differentiate reference to objects and actions. *Cognitive Development*, 10, 201–224. doi:10.1016/0885-2014(95)90009-8

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

Waxman, S. R., & Lidz, J. L. (2006). Early word learning. In W. Damon, & R. M. Lerner (Series Eds.) & R. Siegler, & D. Kuhn (Volume Eds.), *Handbook of child psychology, Vol. 2. Cognition, perception, and language* (pp. 299–335). Hoboken, NJ: John Wiley.

Weber, A., Braun, B., & Crocker, M. W. (2006). Finding referents in time: Eye-tracking evidence for the role of contrastive accents. *Language and Speech*, 49(3), 367–392. doi:10.1177/00238309060490030301

Werker, J. F., & Yeung, H. (2005). Infant speech perception bootstraps word learning. *Trends in Cognitive Science*, 9(11), 519–527. doi:10.1016/j.tics.2005.09.003

Woodward, A. L. (2003). Infants' developing understanding of the link between looker and object. *Developmental Science*, 6(3), 297–311. doi:10.1111/1467-7687.00286

Yu, C., & Ballard, D. H. (2007). A unified model of early word learning: Integrating statistical and social cues. *Neurocomputing*, 70, 2149–2165. doi:10.1016/j.neucom.2006.01.034

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

CHAPTER 5

The role of prosody in early speech segmentation and word-referent mapping

Electrophysiological evidence

Maria Teixidó[1], Clément François[1,2,3], Laura Bosch[1,3,4] & Claudia Männel[5,6]

[1]University of Barcelona, Department of Cognition, Development and Educational Psychology / [2]Cognition and Brain Plasticity Group (Bellvitge Biomedical Research Institute) IDIBELL / [3]Institut de Recerca Sant Joan de Déu (IRSJD) / [4]Institut de Neurociències, University of Barcelona / [5]Max Planck Institute for Human Cognitive and Brain Sciences, Department of Neuropsychology and Department of Neurology / [6]University of Leipzig, Medical Faculty, Clinic for Cognitive Neurology

This chapter reviews electrophysiological studies on early word-form segmentation and word-referent mapping, with a focus on the role of prosody in these early abilities closely related to vocabulary acquisition. First, we will review event-related brain potential (ERP) studies on word segmentation showing the impact of lexical stress cues, infant-directed speech (IDS) properties and melodic information on word-form extraction. Then, we will review research on word-referent mapping, revealing the scarcity of ERP studies specifically exploring the contribution of prosody in this domain. Throughout the chapter we will emphasize how electrophysiological methods offer a more fine-grained perspective of the brain processes supporting segmentation and mapping abilities, often revealing infants' sensitivities to auditory input before overt responses from behavioral methods can be obtained.

Introduction

Early in development, infants face the twin problems of segmenting the speech stream to extract linguistically relevant units and mapping these possible word-forms to their referents, the objects and events in the surrounding world. Developmental research has demonstrated the pivotal role of prosody in this process.

DOI 10.1075/tilar.23.05tei

Prosodic features, such as language rhythm, intonation and lexical stress patterns, can provide infants with a starting point in early word learning (see Jusczyk, 1999; Goyet, Millotte, Christophe, & Nazzi, 2016; for reviews). Following the seminal work by Jusczyk and Aslin (1995), a large number of behavioral studies have revealed infants' gradually refined abilities to exploit prosodic cues, as well as other available cues from the input, to successfully segment the speech stream and build their first vocabulary.

Behavioral research methods involve overt attentional measures from video-recordings or eye-tracking procedures. In contrast, neuroimaging techniques, such as event-related brain potentials (ERPs), can register brain responses underlying early acquisition processes and even reveal infants' abilities where behavioral methods might not be sensitive enough. Following a brief introduction on the ERP methodology, this chapter offers an overview of infant ERP studies on the role of prosody in early word segmentation and word-referent mapping. More specifically, we will focus, in turn, on lexical stress patterns, some properties of infant-directed speech (IDS) such as accentuation and repetition, and melodic information, as cues to early word segmentation. Then, we will review word-referent mapping studies, pointing to the scarcity of ERP data specifically exploring the role of prosody in this domain. Throughout the sections of this chapter, we will emphasize how ERP data significantly extend our knowledge gathered from behavioral language acquisition research.

The method of event-related brain potentials (ERPs) in language acquisition research

Behavior-independent neuroimaging techniques, like the ERP methodology based on electroencephalography (EEG), offer insights into the mechanisms driving infants' behavior and their developmental progression. The ERP methodology has largely contributed to our understanding of early language acquisition processes and continues to be used as a powerful research tool complementing behavioral techniques (for an overview, see Conboy, Rivera-Gaxiola, Silva-Pereyra, & Kuhl, 2008; Friederici, 2005; Friederici & Männel, 2013; Kuhl, 2010; Männel & Friederici, 2008).

The human brain constantly produces electrical activity. EEG captures this brain activity by recording ongoing scalp-level voltage fluctuations that originate from neurons' postsynaptic potentials. Electrophysiological research relies on the fact that external stimulation modulates EEG data. However, the effect of a single stimulation represents only a small part of the ongoing brain activity, such that the event-related portion of EEG recordings associated with a stimulus of interest needs to be extracted. Repeated stimulation and subsequent averaging across stimulus-time-locked EEG-epochs result in stable ERPs representing average stimulus processing and ruling out non-relevant brain activity.

The schematic ERP waveform in Figure 1 displays a sequence of positive- and negative-going stimulus-triggered voltage changes. The labeling of the different ERP components indicates their polarity and the latency of maximum amplitude relative to the onset of the stimulation. The N1(00) component, for example, refers to a negativity that can be observed around 100 ms after stimulus onset. In general, ERP components occurring early after stimulus onset (around 50–200 ms) are mainly modulated by sensory input features, while those occurring later (after 300 ms) reflect cognitive processes to a much larger degree.

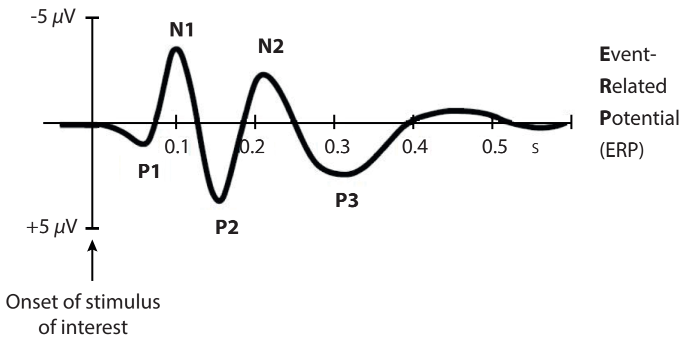


Figure 1. Schematic ERP waveform representing the average response to a class of stimuli. Positive- and negative-going voltage changes are labeled according to their amplitude polarity and latency from stimulus onset (adapted from Männel, 2008).

Different ERP paradigms have been developed to study early language acquisition during passive listening. For example, the Mismatch Negativity paradigm tests auditory discrimination of short stimuli, such as tones, phonemes or syllables, and involves the unexpected presentation of deviant stimuli in a series of repeated, standard stimuli. Familiarization-test paradigms are more suitable when longer or more complex stimuli are of interest, as is the case in word segmentation studies. The latter usually include a familiarization or learning phase, followed by a test phase with familiar and non-familiar elements, so that brain responses time-locked to the onset of the target stimuli can reveal familiarity recognition indicating successful word-form segmentation. Comparing ERPs elicited by different stimuli allows for inferences about the effect of experimental manipulations (Kutas, van Petten, & Kluender, 2006). Differences in amplitude can be the result of variations in processing demands or efficiency across experimental conditions, whereas latency differences can indicate the slowing down or speeding up of cognitive processing.

The ERP methodology, as a non-invasive and highly infant-suitable neuroimaging technique, has been successfully applied in early word segmentation and word-mapping studies. While its spatial resolution is weak, compared to hemodynamic techniques such as functional near-infrared spectroscopy (fNIRS), it has an excellent temporal resolution, comparable to whole-head magneto-encephalography (MEG). Thus, complementing behavioral data, ERP data have largely contributed to a better understanding of the time course of early word segmentation and mapping. The following sections will review available data from ERP studies, with a focus on research specifically exploring the role of prosodic factors on these crucial abilities for language acquisition.

Word segmentation

Finding the words in fluent speech is a fundamental ability in early language acquisition, linked to later vocabulary outcomes (Junge, Kooijman, Hagoort, & Cutler, 2012; Newman, Ratner, Jusczyk, Jusczyk, & Dow, 2006; Singh, Steven Reznick, & Xuehua, 2012). Although single-word utterances are not exceptional in the input to children, most words do not occur in isolation (Brent & Siskind, 2001). Infants thus need to segment the speech stream to extract the individual word-forms, exploiting any possible cues to word boundaries present in the speech input (see Thorson, this volume; de Carvalho, Dautriche, Millotte, & Christophe, this volume, for reviews). Behavioral studies indicate that word segmentation ability is present at around 7 to 10 months of age (Houston, Jusczyk, Kuijpers, Coolen, & Cutler, 2000; Jusczyk, Houston, & Newsome 1999; Polka & Sundara, 2012). Word segmentation can be observed even earlier, by 6 months of age, for monosyllabic word-forms (Bosch, Figueras, Teixidó, & Ramon-Casas, 2013; Höhle & Weissenborn, 2003), and for words following an already familiar name or anchor word (Bortfeld, Morgan, Golinkoff, & Rathbun, 2005). In this context, ERP studies have provided relevant information on the segmentation process, not only by tracking the online processing, but also by offering a direct measure of the factors favoring or constraining early word-form segmentation. In the next sections, we first describe the electrophysiological indicators of word-form segmentation, followed by a review of ERP studies addressing (a) the role of lexical stress patterns in speech segmentation, (b) the effect of specific features of IDS in modulating the developmental onset of this ability, and (c) the contribution of melodic information (song) to the word segmentation process.

ERP indicators of word segmentation

Kooijman, Hagoort, and Cutler (2005) designed the first ERP study on word segmentation with Dutch-learning 10-month-olds. They adapted the behavioral paradigm from Jusczyk and Aslin (1995), familiarizing infants with repetitions of strong-weak words and testing them on passages containing either the familiarized or novel words. Infants' ERPs in the familiarization phase showed a negative-going effect, with more negative ERP responses to the last repetitions compared to the first ones. This "familiarity" or repetition effect appeared from 200–500 ms post word onset at fronto-central recording sites. At test, a negative ERP effect over the left hemisphere was found from 350–500 ms, with larger amplitudes for familiarized target words than unfamiliarized words. The authors concluded that these negative ERP components (later termed N200–500, see Figure 2) were related to word-form familiarity through repetition, and recognition of the familiarized words, thus, indicating that segmentation of the target word-forms embedded in sentences had occurred.

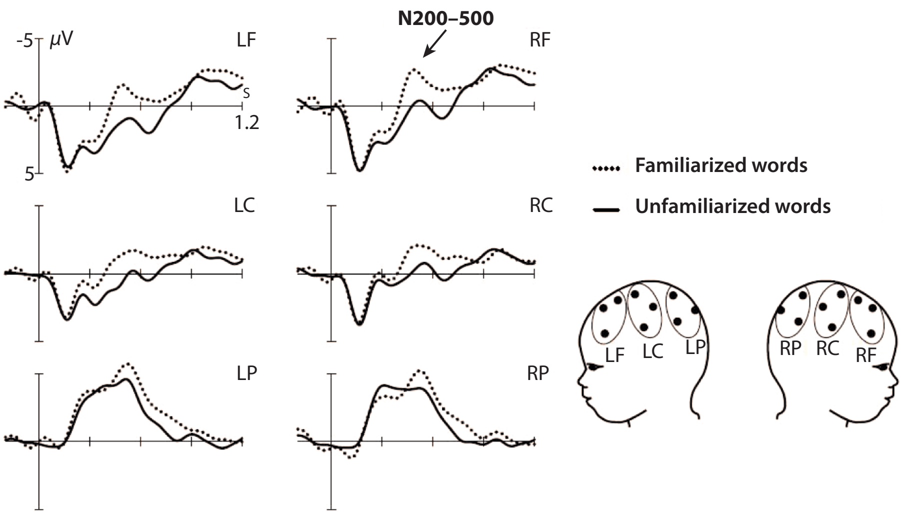


Figure 2. ERP responses of 12-month-olds to familiarized and unfamiliarized test words in a familiarization-test paradigm (modified from Männel & Friederici, 2013). L/RF = left/right frontal region, L/RC = left/right central region, L/RP = left/right posterior region.

Comparable negative-going ERP responses have been reported in studies on word recognition in older infants, both for known words (Mills, Coffey-Corina, & Neville, 1993, 1997; Thierry, Vihman, & Roberts, 2003) and for repetitions of pseudowords (von Koss Torkildsen et al., 2009). The N200–500 component can thus be taken as an ERP indicator of word segmentation. In the next sections, we review ERP evidence of the impact of different prosodic factors on infants' word segmentation abilities.

Language-specific prosodic cues to word segmentation: lexical stress patterns

Behavioral evidence has revealed that differences in the predominant stress pattern of languages may be responsible for the variation in the onset and the robustness of early word segmentation abilities (see Bhatara, Boll-Avetisyan, Höhle, & Nazzi, this volume). For instance, the metrical segmentation strategy, derived from the lexical stress properties of Dutch as the native language, would predict no difficulties in extracting disyllabic word-forms following the predominant stress pattern (e.g., strong-weak in Dutch) and less successful segmentation for the less frequent, weak-strong pattern. But behavioral evidence had failed to yield positive evidence of the expected strong-weak segmentation in 7.5-month-old Dutch-learning infants (Kuijpers, Coolen, Houston, & Cutler, 1998). A follow-up ERP study, however, adopting the same familiarization-test paradigm previously used with 10-month-olds (Kooijman et al., 2005), clarified this issue and confirmed successful segmentation of strong-weak words in 7-month-olds (Kooijman, Junge, Johnson, Hagoort, & Cutler, 2013). Here, the N200–500 component emerged during familiarization, and, in addition, two different ERP patterns indicating familiarity recognition appeared at test. Some infants showed a right-frontal positive ERP effect at 350–450 ms to familiarized versus unfamiliarized words, while others showed a left-posterior negative ERP effect at 430–530 ms. When children's language outcomes were followed up at 3 years, those previously showing a negative-going ERP response obtained higher scores in vocabulary measures. Thus, polarity differences in the ERP could be taken to reflect different strategies underlying infants' word segmentation, linked to maturational variation and related to later vocabulary outcomes.

Evaluating the effect of lexical stress on word segmentation, Kooijman, Hagoort and Cutler (2009) examined whether Dutch-learning 10-month-olds would also be able to segment words that did not follow the predominant stress pattern, namely weak-strong words. ERPs of the familiarization revealed similar N200–500 effects as reported by Kooijman et al. (2005), a familiarity effect that started while infants were still hearing the weak (first) syllable. Results of the test phase revealed a more variable pattern linked to the strong syllables of the target words. While infants were able to recognize repetitions of whole words occurring in isolation, they still relied on the strong syllable in more complex sentence contexts, failing to correctly segment the weak-strong words. These results thus differ from the ones obtained by Kooijman et al. (2005), showing that 10-month-old infants recognized the whole strong-weak word and not only the strong syllable.

Together, these findings highlight the important role of the predominant lexical-stress pattern of the native language in facilitating word segmentation.

We now turn to cross-linguistic differences in word segmentation, looking at French as a syllable-timed language that does not show word-initial stress, but phrase-final lengthening. Goyet, Schonen, and Nazzi (2010) ran an ERP study to further evaluate the delayed disyllabic weak-strong word segmentation reported in a previous behavioral study with French 12-month-olds (Nazzi, Iakimova, Bertoncini, Frédonie, & Alcantara, 2006). Again, 12-month-old infants were familiarized with weak-strong words and then tested on sentences containing familiarized and novel words. Results from the familiarization phase showed an enhanced N200–500 occurring at 300–450 ms post word onset for later repetitions. Similar results were found at test, with the N200–500 effects of word-form familiarity resembling the ones previously reported by Kooijman et al. (2005). Importantly, however, here they indicate successful word segmentation of weak-strong words by French-learning infants, while the 10-month-old Dutch-learning infants in Kooijman et al. (2009) had failed at segmenting words with this specific stress pattern. While confirming French-learning infants' ability to segment disyllabic weak-strong words at the tested age of 12 months, it remains to be shown whether similar results (i.e., whole-word disyllabic segmentation and no independent final-syllable segmentation) can be found at an earlier age. Moreover, ERP data from infants acquiring other syllable-timed languages with variable lexical stress patterns (e.g., Catalan and Spanish) are still missing, limiting the overall conclusions about the function of cross-linguistically different prosodic features in infants' emerging segmentation abilities.

In summary, the described ERP studies show that word-level prosody affects early word segmentation in several significant ways. First, the fact that the language of interest has a more or less predominant word-stress structure seems to affect the onset of word segmentation. Second, words following the predominant pattern of a given language are segmented earlier than words that do not follow this pattern. Third, cross-linguistic rhythm differences can set some constraints on infants' ability to segment words longer than the smallest, monosyllabic elements. Beyond any cross-linguistic differences reported in the literature on infant word segmentation, it is important to highlight that 10-month-old infants show stable ERP responses indicating successful word segmentation; the N200–500 component, which has been previously reported for word-form familiarity. This finding suggests that infants at this early age are able to build up word-form representations from only a few word repetitions in the speech input. Moreover, ERP studies have identified a developmental transition between 7 and 10 months of age, from a positive-going to a negative-going familiarity effect in the ERP, suggesting different underlying mechanisms modulated by advancing brain maturation and sustained language exposure.

Properties of infant-directed speech (IDS) supporting word segmentation: Prosodic emphasis on sentence-embedded words

Across languages, adults have been shown to use infant-directed speech (IDS) when communicating with infants (Fernald et al., 1989). IDS is described as a speech register with especially salient prosodic features, such as more exaggerated pitch contours and emphasis on specific words, produced with higher pitch and intensity levels and longer duration. Moreover, IDS features a slower speech rate, shorter mean length of utterance, more prosodic repetition, higher frequency of specific words and syntactic patterns, and simpler grammatical structures (Fernald, 1992). The specific acoustic features of IDS have been found to facilitate language acquisition, and, importantly, word segmentation (Thiessen, Hill, & Saffran, 2005; see also Cristia, 2013; Soderstrom, 2007, for reviews).

While infants show an initial strong preference for IDS over ADS, this preference for prosodically enhanced speech seems to reduce during the first year of life (Newman & Hussain, 2006). Moreover, parents have been shown to adjust their IDS register to infants' age, such that IDS features are especially pronounced during infants' first months of life, but become less salient later on (Kitamura & Burnham, 2003; Stern, Spieker, Barnett, & MacKain, 1983). Thus, it seems that both infants' use of IDS in language learning and IDS features provided in the speech input change as a function of infants' age. In this context, Männel and Friederici (2010, 2013), using electrophysiological measures, systematically explored the age-dependent role of different speech cues that are typically used by caregivers in word-learning situations: prosodic accentuation and repetition (see Aslin, Woodward, LaMendola, & Bever, 1996; Bernstein-Ratner, 1996; Fernald & Mazzie, 1991; Fernald & Morikawa, 1993). The authors tested German-learning infants at 6, 9 and 12 months of age in a familiarization-test paradigm and manipulated the degree of prosodic modulation and number of repetitions of sentence-embedded target words. Infants were first familiarized with different sentences containing the same target word, presented with or without an additional, naturally produced prosodic emphasis (for acoustic analyses, see Männel & Friederici, 2013). The test phase consisted of tokens of the familiar target word and an unfamiliar control word. ERP responses to sentence-embedded targets collected during familiarization differed depending on infants' age. At 6 months, infants showed a fronto-central positive shift at 200–400 ms post word onset for accentuated versus unaccentuated words, but no effect of repetition. At 9 months, infants also showed an early occurring positive-going ERP effect of prosodic realization and, in addition, an N200–500 word repetition effect. Finally, 12-month-old infants did not show an effect of prosodic realization on target word processing, but a sustained N200–500 as an effect of repetition. These results suggest that at younger ages, prosodic input cues drive infants' word processing, while with increasing age repetition takes more effect. When infants were subsequently tested on their word recognition, ERP responses to familiarized and unfamiliarized words were compared as a function of prosodic realization during familiarization. As displayed in Figure 3, test results revealed that 6-month-olds were only able to recognize words that had received an additional prosodic emphasis during familiarization, showing a positive-going familiarity response in the ERP, starting at around 500 ms post word onset. Nine-month-old infants showed two familiarity effects in the ERP: an N200–500 to familiarized versus unfamiliarized words independently of previous prosodic realization; and a late negativity, only for words that were accentuated in the familiarization phase. Finally, 12-month-old infants also showed a sustained N200–500 effect that occurred for all familiarized compared to unfamiliarized words, independently of accentuation. Taken together, these data suggest that infants recognize familiar words, relying on different speech input cues provided by the speaker when *teaching* infants new words. While younger infants primarily benefit from the additional prosodic emphasis on sentence-embedded words, older infants primarily rely on word repetition, and recognize familiar words independently of their prosodic realization during the learning situation.

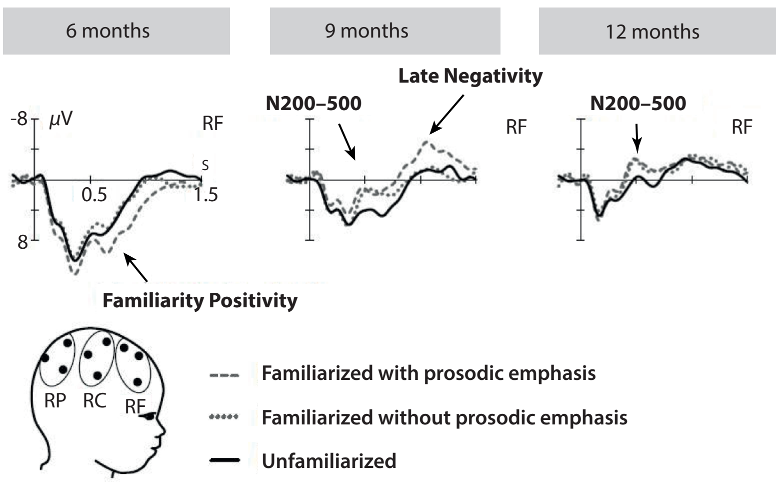


Figure 3. Developmental changes of ERP responses indicating word-form recognition as a function of prosodic manipulation during familiarization (adapted from Männel & Friederici, 2013). RF = right frontal region.

The ERP study by Männel and Friederici (2013) further contributes to the idea that prosodic information in the speech input enables infants' initial word segmentation, and confirms previous behavioral research showing that infants at 6 months of age make use of prosodic cues aligned with word boundaries to segment disyllabic words (Shukla, White, & Aslin, 2011). Moreover, Männel and Friederici (2013) observed a developmental positivity-negativity transition in the ERP responses indicating word recognition, with the latter indicating a more mature process (see also Kooijman et al., 2013; Kooijman et al., 2005). This transition in the ERP markers of infant word segmentation can be attributed to infants' growing linguistic experience and parallel brain maturation. More specifically, at around 6 months of age, myelination and neurogenesis of the auditory cortex undergo tremendous changes (Moore & Guan, 2001; Moore & Linthicum, 2007) that can influence the latency and polarity of early sensory ERP components (see Eggermont & Moore, 2012). These maturational changes may also affect the polarity of infants' word recognition responses. Thus, the current example demonstrates how ERP studies additionally deliver information about the underlying brain mechanisms that trigger infants' word segmentation abilities.

From IDS to song: How early can melodic cues contribute to word segmentation?

Very early in life, infants are exposed to both language and music (Brandt, Gebrian, & Slevc, 2012), and both domains share important similarities in terms of structural information and involved cerebral resources (Heffner & Slevc, 2015; Patel, 2008). But, is music, and specifically song, relevant for speech segmentation and word learning? Song and IDS registers exhibit several similar acoustic and structural characteristics (Fernald, 1992; Papousek, Papousek, & Bornstein, 1985). For example, many of the IDS features, such as exaggerated pitch contours, emphasis on specific words, higher pitch and intensity levels, longer duration, slower rate, shorter utterances and repetition, are also found in songs for infants. Infants show a clear preference for ID songs over AD songs (Unyk, Trehub, Trainor, & Schellenberg, 1992), and ID melodic information has been shown to specifically benefit speech processing as well as language acquisition (Gervain & Werker, 2013; Lebedeva & Kuhl, 2010; Schön et al., 2008; Thiessen & Saffran, 2009).

Assessing a beneficial effect of song on word segmentation, Schön and colleagues (2008) compared segmentation of a spoken, flat-contour stream and a sung stream in adults. While the level of performance was at chance after exposure to the spoken stream, successful word recognition was observed following exposure to the sung stream. These results suggest that redundancy in statistical musical and linguistic structures benefits segmentation. Based on these findings in adults (Schön et al., 2008) and evidence showing that prosodic cues facilitate segmentation in 7-month-old infants (Thiessen et al., 2005) and even neonates (Bosseler, Teinonen, Tervaniemi, & Huotilainen, 2016), it has been recently tested whether melodic information of ID song may benefit speech segmentation in newborns (François et al., 2017). The authors recorded EEG in a group of 2-day-old sleeping neonates during a familiarization-test paradigm with continuous artificial speech streams that were either spoken or sung. In the spoken condition, the continuous streams consisted of 100 repetitions of four pseudowords, with transitional probabilities between adjacent syllables as the only cues to word boundaries. In the sung condition, the language streams were identical except that each syllable had a unique pitch from the diatonic scale. Therefore, the resulting sung streams contained two types of overlapping cues signaling word boundaries: the statistical cues (drop in transitional probabilities between adjacent syllables) together with the pitch interval changes that systematically occurred at word boundaries only. In each condition, spoken and sung, newborns were first exposed to a 3.5 min familiarization stream followed by a test phase in which test items that violated the statistically-defined syllabic order of the words were inserted at random positions allowing for an evaluation of the implicit detection of structural violations.

During familiarization, a comparison between the ERPs of the first and the second half revealed the presence of a fronto-central negative deflection at 400–600 ms post word onset in both spoken and sung conditions. However, while in the spoken condition the fronto-central negativity reached maximum amplitude during the second half, its amplitude was already at maximum during the first half in the sung condition, suggesting a faster segmentation process. To determine whether familiarization was sufficient to establish pseudoword representations and detect violations of syllabic order in the test phase, ERPs to the newly learned pseudowords were compared to those elicited by the violation items. Evidence of successful detection of structural violations was found in the sung condition, but not in the spoken condition. Compared to the newly learned pseudowords, test items violating the syllabic structure elicited an early positivity in the 250–400 ms time window. This positivity resembled the mismatch response consistently observed in infants for speech and non-speech stimuli (see Cheour-Luhtanen et al., 1995; Mueller, Friederici, & Männel, 2012; Stefanics et al., 2009). Thus, the test results suggest that sung streams facilitated the building of robust word-form memory traces and triggered strong enough expectations to correctly detect online structural violations. In addition, these results confirm the importance of the fast-emerging fronto-central negativity during the learning phase for subsequent word-form recognition, as neonates were sensitive to structural violations in the sung condition, but not in the spoken one. These results provide direct evidence of the benefit of songs during the initial steps of language acquisition, thus highlighting the relevance of both speech prosody and melodic information to break the speech code and begin identifying basic linguistic units. Moreover, the benefit of song on speech segmentation supports the idea that musical interventions might be a powerful tool to facilitate language acquisition (Flaugnacco et al., 2015; François, Grau-Sánchez, Duarte, & Rodriguez-Fornells, 2015; Zhao & Kuhl, 2016).

Word-referent mapping

So far, we have focused on research evaluating how infants extract word-forms from fluent speech, greatly aided by prosodic cues. Word learning, however, goes beyond word segmentation, because newly segmented word-forms will eventually become labels for objects, people, and events in children's environment. Thus, establishing word-referent associations and storing these as memory representations are crucial steps children have to accomplish in building a first receptive vocabulary. Behavioral studies have shown that starting at 6–8 months of age, infants are able to successfully establish word-to-world mappings in laboratory settings (Gogate, Bolzani, & Betancourt, 2006; Matatyaho & Gogate, 2008), and even at 2 months of age when presented with distinct syllable-object motion pairings (Gogate, Prince, & Matatyaho, 2009). Regarding semantic representations, Bergelson and Swingley (2012) even found that 6-month-old infants already know the meaning of very frequent words, having established some word-object associations from their daily experience. These abilities have also been explored with the ERP method in infants at different ages, offering further insight into the underlying brain mechanisms and their developmental progression. In the following sections, we will first review ERP evidence of infants' early word knowledge and their ability to establish novel word-object associations. We will then turn to the discussion of the role of prosody in infants' word-referent mapping skills, revealing the scarcity of ERP studies in this field and the need for further neuroimaging research.

ERP indicators of word-referent mapping

Friedrich and Friederici (2004) were the first to investigate infants' brain mechanisms underlying the processing of words in meaningful contexts. To tackle the link between words and their referents by means of ERPs, the authors used a cross-modal picture-word priming paradigm with German-speaking 19-month-olds and adults (see also Friedrich & Friederici 2005a, 2005b, 2008, 2011, 2015; for review, see Männel, 2008). Participants were presented with pictures of objects and accompanying IDS words that correctly or incorrectly labeled the objects.

Crucially, the objects' presentation preceded the label onset, such that the picture content could prime subsequent word processing. Thus, if participants knew the objects (from previous experimental familiarization or existing memory representations), this knowledge could trigger the expectation of the corresponding word and thus ease phonological word-form processing and lexical-semantic processing upon the label's presentation.

The adult ERP data collected in the picture-word priming paradigm revealed two processes indicating successful word processing by priming from the preceding picture context (Friedrich & Friederici, 2004): First, facilitated phonological word-form processing as reflected in a frontal N200–500 effect (i.e., word-form familiarity) for picture-matching words compared to non-matching ones, proving established word-object associations. And second, successful lexical-semantic word comprehension as reflected in an N400 effect (i.e., semantic integration efforts) for words not matching the picture compared to matching ones, due to semantic expectancy violations given referential connections between words and the semantic picture context. The N400 describes a negativity that occurs at around 400 ms relative to stimulus onset and is more pronounced the more semantically unfamiliar, unexpected, or non-matching an event is, given the current semantic context or semantic knowledge (see Holcomb, 1993). Evidence of the N400 in infancy research would thus be a valid indicator of infants' emerging ability to understand the referential meaning of words. It is important to note that although both N200–500 and N400 effects are partly overlapping in their time of occurrence, they can still be differentiated based on their different scalp distributions: The N200–500 typically occurs fronto-centrally, while the N400 occurs broadly distributed over the scalp, with a centro-parietal focus.

Crucially, using the same cross-modal picture-word priming paradigm in 19-month-olds, Friedrich and Friederici (2004) observed comparable ERP effects in infants as reported for adults. Infants did not only show ERP indicators of word-form familiarity from cross-modal phonological priming (N200–500), but also the establishment of referential meaning (N400) (Figure 4). When analyzing these ERP effects in light of children's lexical knowledge (as reported by their parents), there was no correlation of children's receptive vocabulary and the N200–500 effect, but a correlation with the N400-like effect. In other words, only infants with higher comprehension abilities showed an N400 similar to the one found in adults, confirming the interpretation of this infant ERP component as indicating referential processes.

Following this seminal work, Friedrich and colleagues conducted a series of ERP experiments in German exploring the developmental progression of the emerging ability to form referential connections during early infancy. In these follow-up ERP studies, the authors were able to show that even 14-month-olds displayed recognition of word-forms (N200–500) and were able to form referential

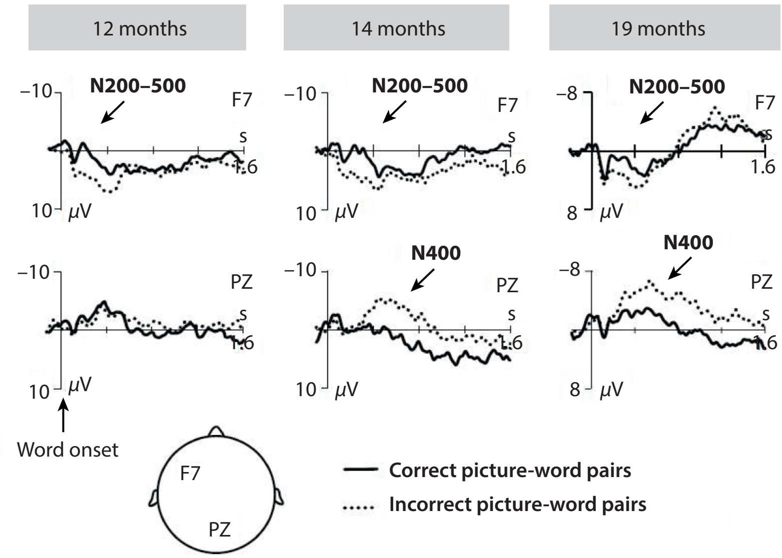


Figure 4. ERP responses for word processing in a cross-modal picture-word priming paradigm at electrodes F7 and PZ (adapted from Friederici, 2005; Friedrich & Friederici 2004; Friedrich & Friederici, 2005a, 2005b).

connections between objects and labels (N400) similarly to older infants and adults (Friedrich & Friederici, 2005b, 2008; Figure 4). However, younger infants at 12 and 6 months of age, while initiating word-object associations both phonologically (N200–500) and semantically (N400) during the familiarization phase, only showed phonologically-driven associations (N200–500) when tested later, suggesting memory limitations preventing full mapping representations (Friedrich & Friederici, 2005a, 2011; Figure 4). Finally, 3-month-old infants were found to generate word-form representations during familiarization (N200–500), as well as an early perceptually-driven form of association between objects and words (Friedrich & Friederici, 2015). The latter ERP effect appeared as centro-parietal negativity at 500–1000 ms post word onset to matching versus non-matching picture-word pairs. This late centro-parietal negativity might indicate a precursor learning mechanism that enables infants to establish first connections between the perceptual representations of objects and words, forming their first proto-words (see Nazzi & Bertoncini, 2003). However, 3-month-olds did not reveal any sign of retention of these associations when tested one day later, making evident the difficulties infants at this age have in retaining this information beyond the immediate encoding stage. Together, these results suggest that the neural mechanisms underlying the building of a first vocabulary are present from early on, while the capacity for creating full referential connections and their successful retention is still highly constrained below 14 months of age.

In summary, the existing ERP studies on word-referent mapping across infancy have reliably revealed two ERP components: First, a fronto-centrally distributed N200–500 that indicates word-form familiarity and evidences word-object associations at a phonological level. Second, a centro-parietal N400 that indicates semantic integration efforts and evidences word-object mapping at a semantic level. Importantly, the association and reference processes underlying these components seem to work semi-independently, and show a different developmental onset. Thus, the reported findings impressively illustrate how the use of electrophysiological measures at early ages not only enables us to successfully delineate the developmental onset and progression of infants' mapping abilities, but also to differentiate the underlying processes.

The contribution of prosody to word-referent mapping: From behavioral to ERP studies

Comparable to behavioral research exploring the role of prosodic features in infants' successful word segmentation, there are an impressive number of behavioral studies addressing the effect of prosody in word-object mapping. Factors such as accentuation (Grassmann & Tomasello, 2007, 2010), lexical stress (Curtin, Campbell, & Hufnagle, 2012; Graf-Estes & Bowen, 2013), specific pitch accents (Thorson & Morgan, 2015), and the role of IDS (Graf-Estes & Hurley, 2013; Ma, Golinkoff, Houston, & Hirsh-Pasek, 2011) have been analyzed and connected to word mapping abilities (see Thorson, this volume). For example, Graf-Estes and Hurley (2013) examined whether the pronounced prosodic features of IDS would enhance word-object associations in 17.5-month-olds. The authors habituated infants to pairs of novel objects and novel words, either in IDS or ADS. Looking-time measures revealed that infants were only able to map words onto objects for labels produced in IDS, but not in ADS. Crucially, the advantage of IDS only took effect when the realization of IDS labels involved prosodic variability. Focusing on a different prosodic property, Curtin et al. (2012) found that lexical stress cues can also guide label-referent associations. In their study, English-learning 16-month-olds were already showing a bias to associate weak-strong patterns with action labels, a language-specific bias consistent with the stress pattern properties of their native language. Thus, behavioral research has delivered significant indications of prosodic cues also driving infants' initial word mapping skills. However, a systematic examination of how prosodic input features affect referential processes by means of electrophysiological measures is still missing. As exemplified above, ERP studies on infants' word-referent mapping have used IDS material (Friedrich & Friederici, 2004, 2005a, 2005b, 2006, 2008, 2011). Yet, these studies have never directly contrasted the effect of IDS and ADS, or manipulated other prosodic speech input features, despite promising behavioral evidence. Thus, ERP studies aiming at exploring the impact of prosody on infants' developing mapping skills are clearly needed.

Another line of ERP research that needs to be strengthened is the investigation of word segmentation and mapping as a joint process, also assessing the impact of speech input features. Naturally, when infants start to segment word-forms from fluent speech, they eventually conceive them as labels to referents in their surrounding world. Recent behavioral research has shown that from an early age, infants can simultaneously segment and map simple, statistically defined word-forms, but only if prosody favors this process (Shukla et al., 2011). Investigating simultaneous segmentation and mapping in a more ecological learning situation, Teixidó and Bosch (2014) tested 4-, 6- and 9-month-old infants on natural language stimuli. Interestingly, the authors observed successful word segmentation and mapping only for infants at 9 months, but not at younger ages. Here, a follow-up ERP study has set out to explore the underlying neural mechanisms driving the success (or failure) of infants' segmentation and mapping skills at young ages (Männel, Teixidó, Bosch, Friederici, & Friedrich, 2017). Importantly, this study manipulated the prosodic realization of target words by placing them at different sentence positions. Preliminary ERP results indicate that while positioning did not affect infants' segmentation ability, it impacted on infants' mapping success, because only targets at prosodically salient edge positions enabled mapping. This ERP study represents a relevant first research step, yet there are clearly more studies needed that assess the neural mechanisms supporting infants' emerging mapping skills and the role played by prosody at different developmental stages.

Conclusions

In this chapter, we have reviewed the role of prosody in infants' word learning, evaluating their emerging abilities of word segmentation and word-referent mapping. Here we have focused on electrophysiological studies that capture the underlying neural mechanisms of infants' emerging abilities, accessible from an early age, when behavioral assessments can be challenging. Current evidence from ERP studies suggests that there are at least two different electrophysiological indicators related to infant word learning. First, an N200–500 component, that is, a word-form familiarity effect characterized by a fronto-lateral negativity in the ERP at 200–500 ms post word onset, elicited when a given phonological sequence is familiar. Second, an N400 component, that is, a semantic (violation) effect characterized by a centro-parietal negativity at around 400 ms post word onset, elicited when the meaning of a word does not match the semantic (picture) context. Importantly, the developmental onset and the amplitude of these two ERP components can predict later language development, particularly children's vocabulary. Crucial for developmental research, both ERP components can be utilized as indicators of infants' emerging word learning abilities to draw a more fined-grained picture of the progression of word-form segmentation and the establishment of referential meaning across infancy. Over and above the illustration of infants' advancing word learning skills by means of ERPs, the data presented here have highlighted the idea that prosody is a key factor in getting infants started on detecting word forms in fluent speech and conceiving them as referents to their surrounding world.

Acknowledgements

The present chapter was supported by the Spanish *Ministerio de Ciencia e Innovación* (projects PSI-2014–55105-P to L.B. and PSI-2015–69132-P to C.F.), the German Research Foundation (project MA 6897/2–1 to C.M.) and the Max Planck Society (C.M.). We would like to thank the editors and two anonymous reviewers for their helpful comments and editorial improvements.

References

Aslin, R. N., Woodward, J. Z., LaMendola, N. P., & Bever, T. G. (1996). Models of word segmentation in fluent maternal speech to infants. In J. L. Morgan, & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 117–134). Mahwah, NJ: Lawrence Erlbaum Associates.

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences*, 109(9), 3253–3258. doi:10.1073/pnas.1113380109

Bernstein-Ratner, N. (1996). From signal to syntax: But what is the nature of the signal? In J. L. Morgan, & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 135–150). Mahwah, NJ: Lawrence Erlbaum Associates.

Bortfeld, H., Morgan, J. L., Golinkoff, R. M., & Rathbun, K. (2005). Mommy and me: Familiar names help launch babies into speech-stream segmentation. *Psychological Science*, 16(4), 298–304. doi:10.1111/j.0956-7976.2005.01531.x

Bosch, L., Figueras, M., Teixidó, M., & Ramon-Casas, M. (2013). Rapid gains in segmenting fluent speech when words match the rhythmic unit: Evidence from infants acquiring syllable-timed languages. *Frontiers in Psychology*, 4(106). doi:10.3389/fpsyg.2013.00106

Bosseler, A. N., Teinonen, T., Tervaniemi, M., & Huotilainen, M. (2016). Infant-directed speech enhances statistical learning in newborn infants: An ERP study. *PloS one*, 11(9), e0162177.

Brandt, A. K., Gebrian, M., & Slevc, R. (2012). Music and early language acquisition. *Frontiers in Psychology*, 3(327). doi:10.3389/fpsyg.2012.00327

Brent, M. R., & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, 81(2), B33–B44. doi:10.1016/S0010-0277(01)00122-6

Cheour-Luhtanen, M., Alho, K., Kujala, T., Sainio, K., Reinikainen, K., Renlund, M., Aaltonen, O., Eerola, O., & Näätänen, R. (1995). Mismatch negativity indicates vowel discrimination in newborns. *Hearing Research*, 82(1), 53–58. doi:10.1016/0378-5955(94)00164-L

Conboy, B. T., Rivera-Gaxiola, M., Silva-Pereyra, J., & Kuhl, P. K. (2008). Event-related potential studies of early language processing at the phoneme, word, and sentence levels. *Early Language Development*, 5, 23–64. doi:10.1075/tilar.5.04con

Cristia, A. (2013). Input to language: The phonetics and perception of infant-directed speech. *Language and Linguistics Compass*, 7(3), 157–170. doi:10.1111/lnc3.12015

Curtin, S., Campbell, J., & Hufnagle, D. (2012). Mapping novel labels to actions: How the rhythm of words guides infants' learning. *Journal of Experimental Child Psychology*, 112(2), 127–140. doi:10.1016/j.jecp.2012.02.007

Eggermont, J. J., & Moore, J. K. (2012). Morphological and functional development of the auditory nervous system. In L. A. Werner, R. R. Fay, & A. N. Popper (Eds.), *Human auditory development* (pp. 61–105). New York, NY: Springer. doi:10.1007/978-1-4614-1421-6_3

Fernald, A. (1992). Human maternal vocalizations to infants as biologically relevant signals: An evolutionary perspective. In J. H. Berkow (Ed.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 391–428). London: Oxford University Press.

Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology*, 27(2), 209–221. doi:10.1037/0012-1649.27.2.209

Fernald, A., & Morikawa, H. (1993). Common themes and cultural variations in Japanese and American mothers' speech to infants. *Child Development*, 64(3), 637–656. doi:10.2307/1131208

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*, 16, 477–501. doi:10.1017/S0305000900010679

Flaugnacco, E., Lopez, L., Terribili, C., Montico, M., Zoia, S., & Schön, D. (2015). Music training increases phonological awareness and reading skills in developmental dyslexia: A randomized control trial. *PloS one*, 10(9), e0138715. doi:10.1371/journal.pone.0138715

François, C., Grau-Sánchez, J., Duarte, E., & Rodriguez-Fornells, A. (2015). Musical training as an alternative and effective method for neuro-education and neuro-rehabilitation. *Frontiers in Psychology*, 6(475). doi:10.3389/fpsyg.2015.00475.

François, C., Teixidó, M., Takerkart, S., Agut, T., Bosch, L., & Rodriguez-Fornells, A. (2017). Enhanced neonatal brain responses to sung streams predict vocabulary outcomes by age 18 months. *Scientific Reports*, 7(1), 12451. doi:10.1038/s41598-017-12798-2

Friederici, A. D. (2005). Neurophysiological markers of early language acquisition: from syllables to sentences. *Trends in Cognitive Sciences*, 9(10), 481–488. doi:10.1016/j.tics.2005.08.008

Friederici, A. D., & Männel, C. (2013). Neural correlates of the development of speech perception and comprehension. In K. N. Ochsner, & S. M. Kosslyn (Eds.), *The Oxford handbook of cognitive neuroscience* (Vol. 1, pp. 171–192). New York, NY: Oxford University Press.

Friedrich, M., & Friederici, A. D. (2004). N400-like semantic incongruity effect in 19-month-olds: Processing known words in picture contexts. *Journal of Cognitive Neuroscience*, 16(8), 1465–1477. doi:10.1162/0898929042304705

Friedrich, M., & Friederici, A. D. (2005a). Phonotactic knowledge and lexical-semantic processing in one-year-olds: Brain responses to words and nonsense words in picture contexts. *Journal of Cognitive Neuroscience*, 17(11), 1785–1802. doi:10.1162/089892905774589172

Friedrich, M., & Friederici, A. D. (2005b). Lexical priming and semantic integration reflected in the event-related potential of 14-month-olds. *Neuroreport*, 16(6), 653–656. doi:10.1097/00001756-200504250-00028

Friedrich, M., & Friederici, A. D. (2006). Early N400 development and later language acquisition. *Psychophysiology*, 43(1), 1–12. doi:10.1111/j.1469-8986.2006.00381.x

Friedrich, M., & Friederici, A. D. (2008). Neurophysiological correlates of online word learning in 14-month-old infants. *Neuroreport*, 19(18), 1757–1761. doi:10.1097/WNR.0b013e328318f014

Friedrich, M., & Friederici, A. D. (2011). Word learning in 6-month-olds: Fast encoding–weak retention. *Journal of Cognitive Neuroscience*, 23(11), 3228–3240. doi:10.1162/jocn_a_00002

Friedrich, M., & Friederici, A. D. (2015). The origins of word learning: Brain responses of three-month-olds indicate their rapid association of objects and words. *Developmental Science*, 20(2), e12357.

Gervain, J., & Werker, J. F. (2013). Prosody cues word order in 7-month-old bilingual infants. *Nature Communications*, 4, 1490–1496. doi:10.1038/ncomms2430

Gogate, L. J., Bolzani, L. H., & Betancourt, E. A. (2006). Attention to maternal multimodal naming by 6- to 8-month-old infants and learning of word-object relations. *Infancy*, 9(3), 259–288. doi:10.1207/s15327078in0903_1

Gogate, L. J., Prince, C. G., & Matatyaho, D. J. (2009). Two-month-old infants' sensitivity to changes in arbitrary syllable–object pairings: The role of temporal synchrony. *Journal of Experimental Psychology: Human Perception and Performance*, 35(2), 508.

Goyet, L., de Schonen, S., & Nazzi, T. (2010). Words and syllables in fluent speech segmentation by French-learning infants: An ERP study. *Brain Research*, 1332, 75–89. doi:10.1016/j.brainres.2010.03.047

Goyet, L., Millotte, S., Christophe, A., & Nazzi, T. (2016). Processing continuous speech in Infancy: From major prosodic units to isolated word forms. In W. Snyder, & J. Pater (Eds.), *The Oxford handbook of developmental linguistics*, (Vol. 1, pp. 133–156). New York, NY: Oxford University Press.

Graf-Estes, K., & Hurley, K. (2013). Infant-directed prosody helps infants map sounds to meanings. *Infancy*, 18(5), 797–824. doi:10.1111/infa.12006

Graf-Estes, K., & Bowen, S. (2013). Learning about sounds contributes to learning about words: Effects of prosody and phonotactics on infant word learning. *Journal of Experimental Child Psychology*, 114(3), 405–417. doi:10.1016/j.jecp.2012.10.002

Grassmann, S., & Tomasello, M. (2007). Two-year-olds use primary sentence accent to learn new words. *Journal of Child Language*, 34(03), 677–687. doi:10.1017/S0305000907008021

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. doi:10.1016/j.pragma.2010.04.019

Heffner, C. C., & Slevc, L. R. (2015). Prosodic structure as a parallel to musical structure. *Frontiers in Psychology*, 6(1962). doi:10.3389/fpsyg.2015.01962

Höhle, B., & Weissenborn, J. (2003). German-learning infants' ability to detect unstressed closed-class elements in continuous speech. *Developmental Science*, 6(2), 122–127. doi:10.1111/1467-7687.00261

Holcomb, P. J. (1993). Semantic priming and stimulus degradation: Implications for the role of the N400 in language processing. *Psychophysiology*, 30(1), 47–61. doi:10.1111/j.1469-8986.1993.tb03204.x

Houston, D. M., Jusczyk, P. W., Kuijpers, C., Coolen, R., & Cutler, A. (2000). Cross-language word segmentation by 9-month-olds. *Psychonomic Bulletin and Review*, 7(3), 504–509. doi:10.3758/BF03214363

Junge, C., Kooijman, V., Hagoort, P., & Cutler, A. (2012). Rapid recognition at 10 months as a predictor of language development. *Developmental Science*, 15(4), 463–473. doi:10.1111/j.1467-7687.2012.1144.x

Jusczyk, P. W. (1999). How infants begin to extract words from speech. *Trends in Cognitive Sciences*, 3(9), 323–328. doi:10.1016/S1364-6613(99)01363-7

Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, 29(1), 1–23. doi:10.1006/cogp.1995.1010

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, 39(3), 159–207. doi:10.1006/cogp.1999.0716

Kitamura, C., & Burnham, D. (2003). Pitch and communicative intent in mother's speech: Adjustments for age and sex in the first year. *Infancy*, 4(1), 85–110. doi:10.1207/S15327078IN0401_5

Kooijman, V., Hagoort, P., & Cutler, A. (2005). Electrophysiological evidence for prelinguistic infants' word recognition in continuous speech. *Cognitive Brain Research*, 24, 109–116. doi:10.1016/j.cogbrainres.2004.12.009

Kooijman, V., Hagoort, P., & Cutler, A. (2009). Prosodic structure in early word segmentation: ERP evidence from Dutch ten-month-olds. *Infancy*, 14(6), 591–612. doi:10.1080/15250000903263957

Kooijman, V., Junge, C., Johnson, E. K., Hagoort, P., & Cutler, A. (2013). Predictive brain signals of linguistic development. *Frontiers in Psychology*, 4(25). doi:10.3389/fpsyg.2013.00025

Kuhl, P. K. (2010). Brain mechanisms in early language acquisition. *Neuron*, 67(5), 713–727. doi:10.1016/j.neuron.2010.08.038

Kuijpers, C., Coolen, R., Houston, D., & Cutler, A. (1998). Using the head-turning technique to explore cross-linguistic performance differences. In C. Rovee-Collier, L. Lipsitt, & H. Hayne (Eds.), *Advances in infancy research* (Vol. 12, pp. 205–220). Norwood, NJ: Ablex.

Kutas, M., van Petten, C., & Kluender, R. (2006). Psycholinguistics electrified II (1994–2005). In M. Traxler, & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (Vol. 1, pp. 83–143). San Diego, CA: Academic Press. doi:10.1016/B978-012369374-7/50018-3

Lebedeva, G. C., & Kuhl, P. K. (2010). Sing that tune: Infants' perception of melody and lyrics and the facilitation of phonetic recognition in songs. *Infant Behavior and Development*, 33(4), 419–430. doi:10.1016/j.infbeh.2010.04.006

Ma, W., Golinkoff, R. M., Houston, D. M., & Hirsh-Pasek, K. (2011). Word learning in infant- and adult-directed speech. *Language Learning and Development*, 7(3), 185–201. doi:10.1080/15475441.2011.579839

Männel, C. & Friederici, A. D. (2008). Event-related brain potentials as a window to children's language processing: From syllables to sentences. In I. A. Sekerina, E. M. Fernández, H. Clahsen (Eds.), *Developmental psycholinguistics: On-line methods in children's language processing* (pp. 29–72). Amsterdam: John Benjamins. doi:10.1075/lald.44.04man

Männel, C. & Friederici, A. D. (2010). Prosody is the key: ERP studies on word segmentation in 6- and 12-month-old children. *Journal of Cognitive Neuroscience*, Supplement, 261.

Männel, C., & Friederici, A. D. (2013). Accentuate or repeat? Brain signatures of developmental periods in infant word recognition. *Cortex*, 49(10), 2788–2798. doi:10.1016/j.cortex.2013.09.003

Männel, C., Teixidó, M., Bosch, L., Friederici, A. D. & Friedrich, M. (November, 2017). Sentence prosody cues object category learning at 6 months. Paper presented at the Ninth Annual Meeting of the Society for the Neurobiology of Language (SNL), Baltimore, USA.

Männel, C. (2008). The method of event-related brain potentials in the study of cognitive processes. A tutorial. In A. D. Friederici, & G. Thierry (Eds.), *Early language development: Bridging brain and behaviour*, (pp. 1–22). Amsterdam: John Benjamins. doi:10.1075/tilar.5.03man

Matatyaho, D. J., & Gogate, L. J. (2008). Type of maternal object motion during synchronous naming predicts preverbal infants' learning of word–object relations. *Infancy*, 13(2), 172–184. doi:10.1080/15250000701795655

Mills, D. L., Coffey-Corina, S. A., & Neville, H. J. (1993). Language acquisition and cerebral specialization in 20-month-old infants. *Journal of Cognitive Neuroscience*, 5(3), 317–334. doi:10.1162/jocn.1993.5.3.317

Mills, D. L., Coffey-Corina, S. A., & Neville, H. J. (1997). Language comprehension and cerebral specialization from 13 to 20 months. *Developmental Neuropsychology*, 13(3), 397–445. doi:10.1080/87565649709540685

Moore, J. K., & Guan, Y. L. (2001). Cytoarchitectural and axonal maturation in human auditory cortex. *Journal of the Association for Research in Otolaryngology*, 2(4), 297–311. doi:10.1007/s101620010052

Moore, J. K., & Linthicum Jr, F. H. (2007). The human auditory system: A timeline of development. *International Journal of Audiology*, 46(9), 460–478. doi:10.1080/14992020701383019

Mueller, J. L., Friederici, A. D., & Männel, C. (2012). Auditory perception at the root of language learning. *Proceedings of the National Academy of Sciences*, 109(39), 15953–15958. doi:10.1073/pnas.1204319109

Nazzi, T., & Bertoncini, J. (2003). Before and after the vocabulary spurt: Two modes of word acquisition? *Developmental Science*, 6(2), 136–142. doi:10.1111/1467-7687.00263

Nazzi, T., Iakimova, G., Bertoncini, J., Frédonie, S., & Alcantara, C. (2006). Early segmentation of fluent speech by infants acquiring French: Emerging evidence for crosslinguistic differences. *Journal of Memory and Language*, 54(3), 283–299. doi:10.1016/j.jml.2005.10.004

Newman, R. S., & Hussain, I. (2006). Changes in preference for infant-directed speech in low and moderate noise by 4.5- to 13-month-olds. *Infancy*, 10(1), 61–76. doi:10.1207/s15327078in1001_4

Newman, R., Ratner, N. B., Jusczyk, A. M., Jusczyk, P. W., & Dow, K. A. (2006). Infants' early ability to segment the conversational speech signal predicts later language development: A retrospective analysis. *Developmental Psychology*, 42(4), 643–655. doi:10.1037/0012-1649.42.4.643

Papousek, M., Papousek, H., & Bornstein, M. (1985) The naturalistic vocal environment of young infants: On the significance of homogeneity and variability in parental speech. In T. M. Field, & N. A. Fox (Eds.), *Social perception in infants* (pp. 269–297). Norwood, NJ: Ablex.

Patel, A. D. (2008). Science & music: Talk of the tone. *Nature*, 453(7196), 726–727. doi:10.1038/453726a

Polka, L., & Sundara, M. (2012). Word segmentation in monolingual infants acquiring Canadian English and Canadian French: Native language, cross-dialect, and cross-language comparisons. *Infancy*, 17(2), 198–232. doi:10.1111/j.1532-7078.2011.00075.x

Schön, D., Boyer, M., Moreno, S., Besson, M., Peretz, I., & Kolinsky, R. (2008). Songs as an aid for language acquisition. *Cognition*, 106(2), 975–983. doi:10.1016/j.cognition.2007.03.005

Shukla, M., White, K. S., & Aslin, R. N. (2011). Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. *Proceedings of the National Academy of Sciences*, 108(15), 6038–6043. doi:10.1073/pnas.1017617108

Singh, L., Steven Reznick, J., & Xuehua, L. (2012). Infant word segmentation and childhood vocabulary development: A longitudinal analysis. *Developmental Science*, 15(4), 482–495. doi:10.1111/j.1467-7687.2012.01141.x

Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review*, 27, 501–532. doi:10.1016/j.dr.2007.06.002

Stefanics, G., Háden, G. P., Sziller, I., Balázs, L., Beke, A., & Winkler, I. (2009). Newborn infants process pitch intervals. *Clinical Neurophysiology*, 120(2), 304–308. doi:10.1016/j.clinph.2008.11.020

Stern, D. N., Spieker, S., Barnett, R. K., & MacKain, K. (1983). The prosody of maternal speech: Infant age and context related changes. *Journal of Child Language*, 10(01), 1–15. doi:10.1017/S0305000900005092

Teixidó, M., & Bosch, L. (2014). Early segmentation and word mapping in a natural language: Evidence from 4-, 6- and 9-months-olds. Paper presented at the XIX Biennial International Conference on Infant Studies (ICIS), Berlin, Germany.

Thierry, G., Vihman, M., & Roberts, M. (2003). Familiar words capture the attention of 11-month-olds in less than 250 ms. *Neuroreport*, 14(18), 2307–2310. doi:10.1097/00001756-200312190-00004

Thiessen, E. D., & Saffran, J. R. (2009). How the melody facilitates the message and vice versa in infant learning and memory. *Annals of the New York Academy of Sciences*, 1169(1), 225–233. doi:10.1111/j.1749-6632.2009.04547.x

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, 7(1), 53–71. doi:10.1207/s15327078in0701_5

Thorson, J. C. & Morgan, J. L. (2015). Acoustic correlates of information structure in child and adult speech. In E. Grillo & K. Jepson (Eds.), *Proceedings of the 39th Annual Boston University Conference on Language Development* (pp. 411–423). Somerville, MA: Cascadilla Press.

Unyk, A. M., Trehub, S. E., Trainor, L. J., & Schellenberg, E. G. (1992). Lullabies and simplicity: A cross-cultural perspective. *Psychology of Music*, 20(1), 15–28. doi:10.1177/0305735692201002

von Koss Torkildsen, J., Hansen, H. F., Svangstu, J. M., Smith, L., Simonsen, H. G., Moen, I., & Lindgren, M. (2009). Brain dynamics of word familiarization in 20-month-olds: Effects of productive vocabulary size. *Brain and Language*, 108(2), 73–88. doi:10.1016/j.bandl.2008.09.005

Zhao, T. C., & Kuhl, P. K. (2016). Musical intervention enhances infants' neural processing of temporal structure in music and speech. *Proceedings of the National Academy of Sciences*, 113(19), 5212–5217.

PART 2

Learning to produce prosody

CHAPTER 6

Set in time

Temporal coordination of prosody and gesture in the development of spoken language production

Heather Leavy Rusiewicz[1] & Núria Esteve-Gibert[2,3]
[1]Duquesne University / [2]Aix Marseille Université, CNRS, LPL UMR7309 / [3]Universitat de Barcelona

The purpose of this chapter is to discuss the relationship between prosody, gesture, and spoken language production from a developmental perspective. The coordination of prosody and manual movements from theoretical, behavioral and neuroanatomical perspectives will be explored. We focus on how gesture landmarks coordinate with prosodic prominence from infants to school-aged children, and we compare these results with adult patterns. The reviewed literature shows that gesture movements, prosody, and speech abilities in general, develop in a parallel and dynamic way, the emergence of a specific ability in one modality entraining the emergence of a parallel ability in another modality. Lastly, implications of the coordination of prosody and gesture for children with communication disorders will be explored.

Introduction

Prosodic characteristics of speech are essential not only for the production and flow of spoken language, but also for the perception and processing of speech and language. This chapter focuses mainly on prosodic prominence (i.e., stress and accent) and its relationship to manual movements in the developing system. Certainly, the interactions between gestures and spoken language are vast and varied. The empirical study of the interactive nature of speech and gesture continues to illuminate many interesting features in children's early years, such as the impact of gesture on early word learning (e.g., Capone & McGregor, 2005; Goodwyn, Acredolo, & Brown, 2000) and later mathematical concepts (e.g., Cook, Duffy, & Fenn, 2013). Gesture use can predict language milestones (e.g.,

DOI 10.1075/tilar.23.06rus

Butcher & Goldin-Meadow, 2000) and language impairments (e.g., Crais, Watson, & Baranek, 2009; Sauer, Levine, & Goldin-Meadow, 2010). Yet, the majority of these developmental relationships are focused on the cognitive-linguistic function of gestures, not the temporal linkages of the speech and manual systems from a motor behavior perspective. It is now well established that speech and gesture are tightly aligned in time, and that prosodically prominent events are the anchoring point for these synchronized movements. Exploring the alignment and interactive nature of prosodic stress and manual movements can offer insight into the development of both motor control (i.e., movements involved in communication) and the linguistic abilities that infants and children use to communicate with others in a meaningful way.

Theoretical underpinnings

Take a moment to view a speaker on the television/computer or a conversational partner. Observe their gestures while they speak. Regardless of the frequency or type of gestures that they are using or the content of their speech, do you notice any patterns in the way that their hand movements and speech units synchronize? You certainly may notice that the two systems line up according to the meaning of the gesture and the meaning of the words. However, what you will also likely notice is that manual gestures coordinate in time with the points of prosodic stress within the speaker's utterances.

Some theories suggest that spoken language and gesture align as a function of linguistic processes, such as lexical access (e.g., de Ruiter 1998, 2000; Krauss, Chen, & Gottesman, 2000). Others have theorized that speech and gesture are aligned as a function of underlying motoric pulses (Iverson & Thelen, 1999; McNeill, 1992; Port, 2003; Rusiewicz, 2011; Tuite, 1993). According to this latter group, movements are entrained (i.e., coordinated) as a function of the peaks (i.e., pulses) of an ongoing rhythm (Merker, Madison, & Eckerdal, 2009). Pulse-based entrainment is rooted in dynamic systems theory (DST) and is frequently applied in many human behaviors, as in the performance of music or perception and attentional processes (Ito, 2012; Large & Jones, 1999). Thus, a stressed syllable can draw in the movement of the hands or even other body parts, such as the torso and eyebrows. According to the Theory of Entrained Manual and Speech Systems (TEMSS), the analogous pulse of movement in the speech stream (i.e., a syllable produced with increased physiologic and acoustic efforts) entrains the pulse of other body movements in both space and time (Rusiewicz, 2011). As illustrated in Figure 1, the more automatic and proximal pulse of the speech system (i.e., prosodic prominence) "pulls in" and entrains other body movements both in time and space. The TEMSS also proposes that the body movements can impact the magnitude, timing, and spatial properties of the speech movements (e.g., wide open hand shapes will lead to more open mouth configurations than closed fist shapes). Thus, the TEMSS proposes that body movements and the movement in the speech stream influence and impact each other. Though the TEMSS does not overtly predict developmental differences compared to an adult system, the basic tenets of this theory regarding entrained speech and manual movements remain applicable for children.

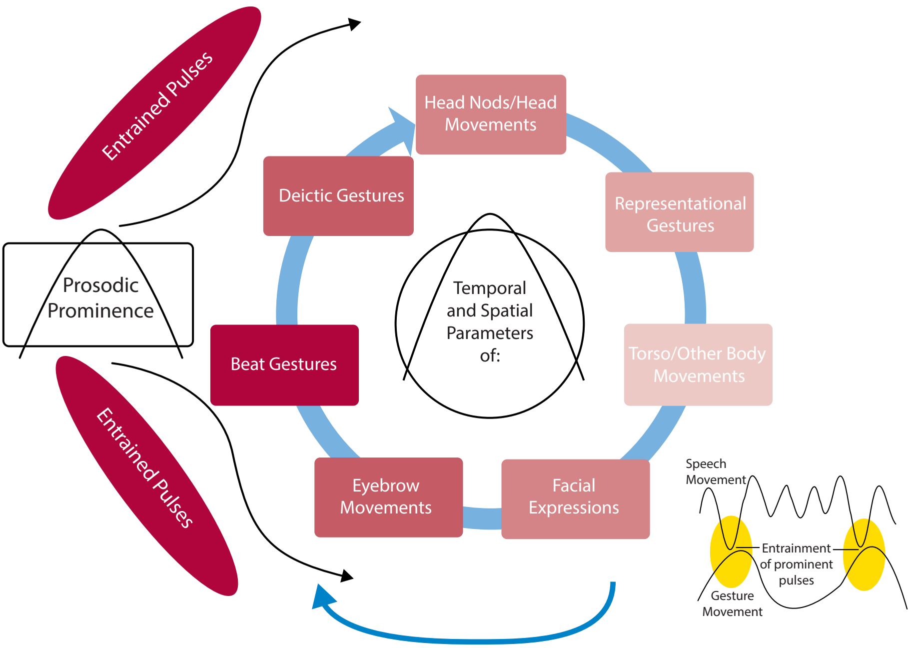


Figure 1. Representation of the spatiotemporal entrainment of speech and other body movements aligning with prosodically prominent points within the speech stream as a function of underlying, internal, pulse-based entrainment. The hypothesized degree of entrainment and subsequent synchronization of body movements with prosodic prominence is illustrated by darker shading for more automatic, proximal, and robust degrees of entrainment. The temporal alignment of theses pulses is illustrated on the right.

What does it mean to be temporally coordinated?

It has long been observed that body movements such as hand gestures and the suprasegmental properties of speech temporally align in adults (e.g., Birdwhistell, 1952; Dittman & Llewellyn, 1969; Kendon, 1980). Speech and manual movements are coordinated as a function of prosodically prominent syllables, which act as the anchoring point in speech with which the specific events in the gesture movement align. Thus, the prominent phases in the gesture movement (the 'gesture stroke[1]' or the 'gesture apex[2]') occur within the limits of the stressed syllables in speech (see Figure 2). The evidence for this tight temporal coordination comes from studies using different elicitation methodologies and data analysis systems (e.g. De Ruiter, 1998; Esteve-Gibert & Prieto, 2013; Leonard & Cummins, 2011; Rochet-Capellan, Laboissièrre, Galvan, & Schwartz, 2008; Rusiewicz, Shaiman, Iverson, & Szuminsky, 2013; Rusiewicz, Shaiman, Iverson, & Szuminsky, 2014). Leonard and Cummins (2011), for instance, asked adults to read a story and analyzed the gesture-speech alignment patterns using a 3-D motion tracking system. They found that the apex of beat manual gestures (the point in time in which the hand changes its direction in the vertical or lateral movement) always occurred within the stressed syllables. In a study conducted by Rusiewicz et al. (2013), participants produced sentences that could have words with and without contrastive stress (e.g., *No, the LIFEguard/lifeGUARD is above the square*) while they produced a pointing gesture. Data were obtained using a modified Theremin that measured voltage as the hand approximated the capacitance sensors. Results showed that prosodic stress and its position affected the time of pointing gesture onset and the time required to reach the apex of the pointing gesture movement produced with and without contrastive stress. Interestingly, recent evidence suggests that the position of prosodic edges (intonational phrase boundaries) contribute to defining the specific alignment of the apex within the accented syllable. These results show that the gesture apex is retracted (occurring earlier than expected) when a gesture movement is coordinated with a stressed syllable that is immediately followed by a phrase boundary, since otherwise the speaker would not have time to accommodate the rest of the gesture movement within the limits of the speech signal (Esteve-Gibert, Borràs-Comes, Asor, Swerts, & Prieto, 2017; Krivokapić, Tyede, Tirone, & Goldenberg, 2016). Thus, the position of prosodic prominence is an anchoring point in speech for gesture prominence coordination, but its specific alignment seems to depend also on the position of prosodic edges.

Not only do speech and manual pulses align, but the spatiotemporal parameters of one system can alter the spatiotemporal parameters of the second system.

---

1. In iconic and metaphoric gestures, the gesture stroke is the movement that actually depicts the object or the action; in deictic gestures it is the interval of time that determines the referent; in beat gestures, it is the interval that goes from the movement transition point to the end of the rest position.

2. The apex is defined as the specific temporal point in which the gesture reaches its kinetic goal. In deictic gestures, for instance, it occurs when the arm and the finger are maximally extended.

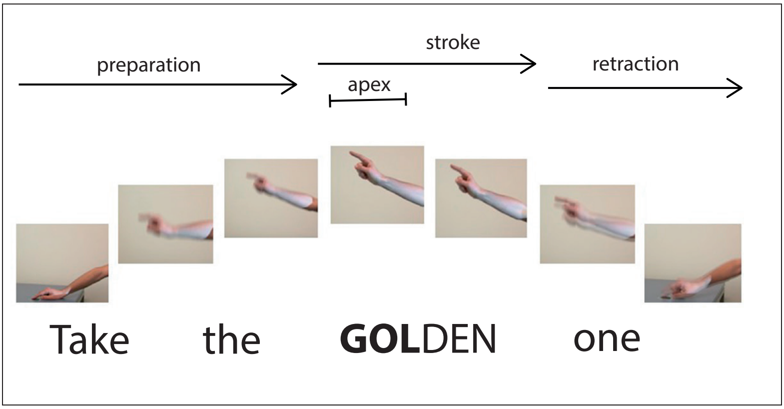


Figure 2. Schematic representation depicting the temporal alignment of prosodic and gesture landmarks. In the sentence, capital letters indicate the target word, and bold letters indicate the accented syllable.

For instance, Rusiewicz et al. (2014) again used voltage traces measured by capacitance to demonstrate that as speech (i.e., syllables within short phrases with and without contrastive stress) was lengthened due to delayed auditory feedback (DAF) perturbations (i.e., hearing one's own speech with a 200 ms delay), the corresponding pointing gesture was also lengthened. This finding mirrors earlier work by McNeill (1992) in which DAF resulted in perturbations of both the speech and gesture systems in adults. Additional support for the effect of pulses in one system affecting the timing and magnitude of movements in the second system comes from an intriguing series of experiments by Krahmer and Swerts (2007). Participants were instructed to either stress *Amanda* or *Malta* in sentences like *Amanda went to Malta* (in Dutch in the original experiment), while at the same time producing a manual beat gesture, an eyebrow movement, or a head nod. As one would expect, the perception of prominence was enhanced when a visual beat occurred on the target syllable. However, the production of a beat changed the acoustic parameters of syllables, even when the speaker was consciously attempting to produce the vowel as unstressed. These data provide motivation to continue the reflection and investigation of *visual prosody* in the form of hand gestures, facial expressions, head nods, etc. in both adults and children.

Developmental sequence of gestures

While the emergence of the prosodic-meaning patterns is detailed in Esteve-Gibert and Prieto (this volume) and Frota and Butler (this volume), the aim of this section is to lay a foundation for understanding how different types of gestures

Table 1. Comparison of gestural, speech, linguistic, and prosodic milestones (information compiled from: Bates & Dick, 2002; Demuth, this volume; Ejiri & Masataka, 2001; Esteve-Gibert & Prieto, this volume; Frota & Butler, this volume; Green, Moore, Higasikawa, & Steeve, 2000; Green, Moore, & Reilly, 2002; Hulit, Fahey, & Howard, 2015; Iverson & Thelen, 1999; Patel & Brayton, 2009; Patel & Grigos, 2006; Smith & Zelaznik, 2004; Steve, 2010)

Age	*Communication / Pragmatics*	*Speech/Oral*	*Gesture*	*Prosody*	*Prosody-gesture coordination*
Newborn-6 months	Social smiles, prolonged eye contact	Sucking, crying, vegetative sounds, cooing emerges	Reflexive grasping/non-communicative facial expressions	Intonation variation emerges with cooing and crying	Hand to mouth linkages
6–8 months	Responds to name / goal-directed actions / joint attention	Sound play, rhythmic canonical babbling / receptively understands some words	Reaching, rhythmic waving and rhythmic hand banging	Prosodic cues expressing positive and negative emotions	Rhythmical hand movements timed with canonical babbling
8–14 months	Intentional use of communicative strategies (language/gesture) emerges	Variegated babbling, first words follow onset / emergence of first words	Gestural routines / deictic pointing gestures (declarative and requestive functions)	Early speech acts expressed through intonation / emergence of the distinction between stressed and unstressed syllables	Non-coupling of pointing and speech productions (not co-occurring in time).
14–18 months	Reasoning of the others' intentional behavior / emergence of speech acts	Jaw movements reach distinguished from chewing/ one-word stage / word truncation	Early iconic gestures	Speech acts expressed through intonation / basic 'unmarked' prosodic words	Pointing apex aligned with the stressed syllable in pointing-speech combinations

Table 1. (*Continued*)

Age	*Communication / Pragmatics*	*Speech/Oral*	*Gesture*	*Prosody*	*Prosody-gesture coordination*
18–24 months	Symbolic and mental representations	Two-word combinations / jaw movements move toward adult like motor patterns	Iconic gestures	Full inventory of pitch accents and boundary tones / language-specific prosodic words	Co-speech iconic gestures, both modalities expressing the same information
Preschool-age (3–5 yrs)	False-belief understanding / adult-like turn-taking skills	Utterance length and complexity greatly increase /able to tell short stories / refining of the speech motor control	Beat gestures	Intonation signaling information structure (i.e. focus)	Head and body movements timed with focalized speech events
School-age children (6–11 yrs)	Second-order belief representations	Developed narrative and discourse skills / adult-like speech motor control at 14 years of age	Beat gestures / metaphorical gestures	Intonation expressing complex meanings (contrast-of-beliefs) / adult-like rhythm metrics	Beat gestures timed rhythmically with speech in narratives

emerge across different stages of development, a crucial pre-requisite to understanding how these gestures are later combined with speech.

From a format point of view, all gestures are very similar, and require similar motor accuracy. However, they differ greatly in function, and the more complex the pragmatic dimension represented by the gesture, the later it is observed in development. Crucially, the types of gestures that infants are able to produce in each developmental stage are strongly related to the prosodic and speech patterns with which they will be aligned. Thus, the emergence pattern of the different gesture types seems to depend on the child's linguistic, communicative, and cognitive development (see Table 1 for a comparison of the emergence of the different communicative, speech, and gesture milestones across development).

Only some days after being born, young infants can produce facial expressions that caregivers interpret as having emotional value (e.g. Barrett & Campos, 1987). Very young infants signal changes in their states through facial activity, and these facial patterns become specific expressions of emotions, because adults link these patterns with emotions during the course of infant-caregiver social interactions (Cole & Moore, 2015). Then, during the first year of life, infants produce hand actions (during play/rest activities, while manipulating objects, that are directed to adults, and that have some form of iconicity) that seem to develop dynamically. Mathew and Manjula (2015) examined the hand actions of nine 3- to 12-month-old infants and found that most of the hand actions follow a non-linear developmental pattern. Thus, only those hand actions that were observed already at 3 months of age develop in a linear way (e.g., 'spread', 'grasp', and 'reach'), and, for most of the actions involving manipulation of an object, there was an explicit age within the developmental period in which their frequency rate significantly increased or decreased.

Deictic gestures are the first communicative gestures that infants produce. Deictic gestures locate objects or events in space, the location being concrete or abstract (normally performed by extending the index finger, although other body parts like the head, chin or nose can be used). These gestures are first observed at around 9–10 months of age, are one of the clearest signs that infants have developed the ability to engage in referential intentional communication (Esteve-Gibert & Prieto, this volume), and can take the form of showing, giving, and pointing gestures. At an early stage, infants point in order to use the adult as a means to obtain a desired object or event (proto-imperative pointing), or to use the object as a means to obtain the desired adult's attention (proto-declarative pointing) (Bates, Camaioni, & Volterra, 1975; Volterra, Caselli, Capirci, & Pizzuto, 2005). Around 12 months of age infants are found to point with an imperative function (to request an object from the adult), an expressive function (to share their interest about an object with the adult), and an informative function (Tomasello, Carpenter, & Liszkowski, 2007). In fact, the infants' pragmatic development determines the functions that infants can express through pointing gestures. Thus, at earlier stages, the most basic speech action functions will be observed, and these will be more complex as they become more pragmatically skilled (see Esteve-Gibert and Prieto, this volume, for a review on the prosody-pragmatics interface). The infants' production of deictic gestures to refer to objects and events is found to precede their use of words to refer to these same objects and events. Deictic gestures thus seem to bootstrap the acquisition of nouns in the speech modality (Iverson & Goldin-Meadow, 2005; Özçalışkan & Goldin-Meadow, 2005).

The next gestures to emerge are representational gestures, which can be iconic (depicting concrete objects or events) or metaphoric (depicting concepts and ideas in an abstract way). Earlier forms of representational gestures emerge around 15 months of age, like the no-gesture, the bye-bye gesture, a gesture representing tall, or drinking from a cup (Acredolo & Goodwyn, 1988; Iverson, Capirci, & Caselli, 1994). The rate and type of representational gestures are found to increase significantly as children become more advanced in their lexical and cognitive development, i.e., around 26 months of age (Özçalışkan, Gentner, & Goldin-Meadow, 2013; Özçalışkan & Goldin-Meadow, 2011). Again, the emergence of representational gestures signals the development of parallel linguistic and grammatical skills. This pattern, however, seems to differ in iconic gestures compared to deictic gestures: while in deictic gestures, the use of gesture for reference precedes the use of speech, the use of iconic gestures to refer to verbs seems to be preceded by the use of speech for the same purpose (Özçalışkan et al., 2013).

The last gestures to emerge ontogenetically are beat gestures. Beat gestures are bi-phasic, rhythmic, co-speech movements that highlight items in the discourse and that bear no semantic relation to speech. There is very little research on when and how children start producing these gestures (Colletta et al., 2015; Mathew, Yuen, Shattuck-Hufnagel, Ren, & Demuth, 2014; Nicoladis, Mayberry, & Genesee, 1999), although it seems that their rhythmic skills need to be quite advanced. Nicoladis et al. (1999) found that the children's mean length of utterance needs to be at least 2.5 morphemes, while some studies suggest that it is not only until much later (6–10 years of age) that beat gestures accompany explanations and narrations in an adult-like way (Colletta et al., 2015; Mathew et al., 2017). Adult-like rhythmic prosodic skills seem to emerge quite late in development (Post & Payne, this volume), and the children's production of beat gestures seems to parallel this pattern. However, to date, and as far as we know, it is still unknown whether beat gestures pave the way for the emergence of prosodic rhythm, or whether the development of prosodic rhythm is a prerequisite for beat gestures to emerge.

Temporal coordination of prosody and gesture in development

Iverson and Thelen (1999) offer the sole theoretical model that explicitly considers the interplay of prosody and gestures during development. Like others, their postulation relies upon the assumption that speech and gesture are aligned in time and that these behaviors are coordinated as specified in DST. The authors stated it is "through rhythmical activity, and later through gesture, the arms gradually entrain the activity of the vocal apparatus […] this mutual activation increases as vocal communication through words and phrases becomes more practiced, leading to a tight synchrony of speech and gesture in common communicative intent" (p. 36).

Iverson and Thelen propose four phases in the dynamics of gesture-speech entrainment in infants. During the first months of life, infants show *initial linkages* between oral and manual systems: they bring their hands to the facial area and introduce their fingers into their mouth, they open their mouth when a pressure is applied to their palms (the so-called Babkin reflex), and at around 2 months of age they start bringing objects to their mouth. At around 3–4 months of age, and especially from 6 months onwards, they show *emerging control* between the two systems: infants start producing rhythmical movements of the hands and arms that are timed together with vocal cooing and babbling. Later (at 10–11 months), infants show *flexible coupling* of gesture and speech, controlling and using the two modalities asymmetrically to convey the intended meaning (gestures being more frequent than speech). Finally, *synchronization coupling* emerges, and infants coordinate both modalities for intentional communication in an adult-like way (i.e., coordinating the most prominent part in gesture with the most prominent part in speech).

Iverson and Thelen's (1999) proposal presents a tantalizing postulation for the coordination of speech and gestures in the earliest years of development (before the age of two). This emphasis on infancy/toddlerhood mirrors the empirical literature on this topic, which offers a relative dearth of information on preschool and school-aged children compared to children between birth to two and adults. Theoretically, it is quite certain that the "changing patterns of coordination of mouth and hand in the first years" (p. 36) remain in a state of flux throughout the refinement of speech, language, and other motor behaviors throughout childhood and even adulthood, especially for those with communication disorders. In other words, as children acquire and master new language forms (e.g., increased utterance length and complexity, mastery of narratives and discourse, etc.) the entrainment, synchronization, and mutual influence of pulses with the speech stream and hand movements will also go through dynamic fluctuations (see Table 1). In this area, research points to a predictable and tight unity of prosodic stress and gestures in both adults and children, as well as the need to continue this line of theoretical and empirical work for those with typical spoken language skills and those with communication deficits.

Vocal-motor entrainment in early infancy

There is consistent evidence that infants increase their rate of upper limb rhythmic movements when they start babbling canonically (e.g. Ejiri & Masataka, 2001; Iverson & Fagan, 2004). The canonical babbling period emerges in the middle of the first year of life and consists of combinations of consonant-vowel syllables that are well-formed and have adult-like spectral and temporal properties (Kent & Bauer, 1985; Oller, Wieman, Doyle, & Ross, 1976). This is a crucial stage in development, because it provides the raw material for the production of early words (Vihman, this volume). Ejiri and Masataka (2001) videotaped 4 Japanese infants from 6 to 11 months of age while interacting with their mothers. They analyzed the video recordings in order to examine and analyze the temporal overlap between vocalizations and rhythmic activities of the limbs during the babbling stage. The authors found that vocalizations frequently co-occurred with rhythmic actions; these coordinated behaviors increased immediately before and during the month in which canonical babbling was initiated, and babbling vocalizations were more adult-like (shorter syllables and shorter duration of formant-frequency transitions) when they were accompanied by rhythmic manual movements.

The early coordinated vocal-limbic rhythmic movements seem to be the precursors of an integrated gesture-speech system in which gesture prominences align with prosodic prominences. Iverson and Fagan (2004) arrived at this conclusion after analyzing the video-recordings of 47 infants between the ages of 6 and 9 months and observing that: (1) infants coordinated vocalizations with single-arm rhythmic movements more often than with both-arm movements; (2) the proportion of coordinated right-arm movements was higher than that of left-arm movements, paralleling adult-like behaviors; and (3) infants' motor activities were synchronous with or slightly anticipated vocalization onsets. At this age, young infants have not started producing communicative gestures yet (i.e. pointing gestures), nor have they entered the single-word period. However, this early tight entrainment between vocal and motor systems suggests that gesture and speech develop in a parallel way and might form an integrated system from the very beginning of language and communicative development. Infants produce oral rhythmic movements coordinated with manual rhythmic movements, and this is taken as the emergence of the pulse-entrainment rules that will govern the coordination between prosodic and gestural prominences.

Prosodic-gestural coordination in infancy and toddlerhood

As previously explained in this chapter, gesture and speech are temporally coordinated because the timing of the gesture movement is determined by the prosodic structure in speech (the position of the accented syllables and of intonational boundaries). In development, the first instances of adult-like temporal coordination of prosodic and gesture landmarks are observed with the emergence of pointing gestures. It should be noted that the first pointing gestures that infants produce are not combined with vocalizations (i.e. they do not co-occur with babbling or word-like vocalizations). It is around 12–15 months of age, when children enter the one-word stage, that infants combine pointing gestures with meaningful speech (Butcher & Goldin-Meadow, 2000; Esteve-Gibert & Prieto, 2014; Igualada, Bosch, & Prieto, 2015; Murillo & Belinchón, 2012; Murillo & Capilla, 2016).

As soon as infants combine pointing gestures with meaningful speech (intentional vocalizations, word-like sounds, or words that refer intentionally to objects or events), these pointing-speech combinations already show adult-like pulse-entrainment rules. Butcher and Goldin-Meadow (2000) analyzed the spontaneous pointing gesture and speech productions of six English-learning infants when they were at the one-word period and until they started producing two-word combinations. Their results show that infants combined pointing gestures with speech at the end of the one-word period, although it was not until the two-word period that infants temporally aligned the gesture stroke, i.e., the interval of prominence in the pointing movement, with the corresponding word in speech.

More recent analyses of the temporal patterns of the early pointing-speech coordination suggest that there is evidence of prosody-gesture coordination in infants already at the single-word period. Esteve-Gibert and Prieto (2014) examined infants from the babbling period to the two-word period to see the temporal alignment patterns of pointing-speech combinations. They recorded 11- to 19-month-old infants while interacting naturally with their caregivers. Four temporal measures were annotated: (a) the distance between the gesture onset and the onset of the vocalization or word, (b) the distance between the stroke onset and the beginning of vocalization or word, (c) the distance between the stroke onset and the onset of the accented syllable, and (d) the position of the gesture apex with respect of the accented syllable. The results of the analyses showed that infants produced fine temporal alignment of gesture-prosodic prominent events from the onset of pointing-speech combinations. More precisely, when infants combined a pointing gesture with a vocalization or a word, the onset of the gesture movement preceded the onset of the corresponding speech. This happened at all ages analyzed, although the variability and temporal distance between one event and the other became more adult-like as infants were older. Also, at all ages, the onset of the stroke phase in pointing gestures coincided very precisely with the onset of the accented syllable in the corresponding vocalization or word (and since most of the words were monosyllabic, the stroke onsets coincided with vocalization or word onsets). Finally, and interestingly, infants produced the apex of the pointing gesture before the end of the corresponding accented syllable, and they did so at all ages, although the variability and the temporal distance between one event and the other was more adult-like when infants were more advanced in their language development.

There are several studies indicating that the infants' production of temporally aligned pointing-speech combinations is not incidental or independent of linguistic purposes. First, it has been observed that 9-month-old infants (and well before the emergence of pointing-speech combinations) are sensitive to the temporal alignment between the pointing gesture movements and the prosodically prominent events in speech. Using a habituation/test paradigm, Esteve-Gibert, Prieto, and Pons (2015) found that 9-month-olds can perceive if a pointing gesture is temporally aligned with a disyllabic word in an adult-like way (the apex of the pointing gesture occurring within the limits of the accented syllable). Second, Igualada et al. (2015) observed that 12-month-old infants control the production of pointing-speech combinations as a function of the communicative needs of the situational context. These researchers observed that infants aligned a pointing gesture with speech more often when they needed to attract the attention of an adult towards an object. This study showed that the infants' rate of pointing-speech combinations is higher when the situation is more demanding from a communicative and attentional point of view.

Despite these recent findings, one important caveat needs to be made. The influence of the prosodic structure on the timing of the gesture movements cannot be evaluated at this early stage as it is in adulthood. At the early stages of word production, the nature of the prosodic words, prosodic structure, and prosodic phonology rules differ widely from those of adults (Demuth, this volume; Kehoe, this volume; Vihman, this volume). Consequently, the anchoring points in speech where gestures movements can align are also different. Further studies should examine the alignment patters of pointing gestures with speech in later stages of the children's prosodic development, especially when the production of adult-like prosodic structure units are more developed.

The preschool years and beyond

The development of the prosodic-gesture coordination after the early toddler years is less clear and less often studied. Yet, it is particularly interesting to study the developmental sequence of beat gestures, discourse skills, and prosodic modulation in the preschool years and beyond. As reviewed earlier and summarized with other phases of development in Table 1, beat gestures are the last form of gesture to be developed. Though studies have investigated the coordination of many gesture types, manual movements, and other body movements with prosodic prominence, beat gestures arguably share the most obvious and potentially strongest link with prosody. Traditionally, beat gestures (i.e., batons) are defined as marking the rhythmic timing of spoken language without conveying semantic information. Beat gestures are used to organize information in discourse and to mark prominent elements in the utterances. Adult listeners use the alignment of manual beat gestures with speech to process sentences in which the position of the phrase boundary is key to disambiguate the meaning in sentences, such as in Italian sentences like *[Canta solo] [se è felice] vs. [Canta] [solo se è felice]* 'He sings alone if he is happy' vs. 'He sings only if he is happy' (Guellaï, Langus, & Nespor, 2014). Also, speakers mark contrastively focused elements in an utterance by means of prosodic cues but also through beat gestures produced with the head (Esteve-Gibert et al., 2017; Kim, Cvejic, & Davis, 2014).

At preschool ages, children seem to process the highlighting function of beat gestures, since they recall words better if they are accompanied by a beat gesture (Igualada, Esteve-Gibert, & Prieto, 2017). In production, beat gestures seem to emerge consistently between the ages of 5 and 7 years (Capone & McGregor, 2004; Colletta et al., 2015; Igualada, Wray, Prieto, & Norbury, 2016; Mathew et al., 2014; McNeill, 1992; Tellier, 2008), though they are still used to a lesser degree than other gesture types at this age (Iverson & Braddock, 2011). Mathew et al. (2014) analyzed the rate of beat gestures that 6-year-old children produced, along with the rhythmic marking of speech. Interestingly, the authors compared a narration task (a story-retelling activity) with an explanation task (child and mother planning a trip), and took into account two types of beat gestures: those that were 'independent' (with non-propositional hand shape) and those that were 'embedded' (with propositional hand shape; for instance, integrated within an iconic gesture). Results showed that 6-year-old children produced more beat gestures in the explanation task and that they followed the adult-like temporal alignment patterns. The production of beat gestures in narratives seems to be observed later in development. Colletta et al. (2015) examined the gestures produced within a story retell task of a cartoon that was under 3 minutes in length, in 5- and 10-year old French-, American English- and Italian-speaking children. As expected, children increased the length and complexity of narratives with increased age. Interestingly, the amount of gesture use did not differ for 5 and 10 year olds. However, what did change with age, in relation to gesture production, was the number of beat gestures that accented the narrative, such that the 10-year-old children produced significantly more beat gestures compared to the 5-year-old children.

When examining prosodic and language development at these ages, prosodic control stabilizes (Patel & Brayton, 2009; Patel & Grigos, 2006), discourse and narrative skills increase (Owens, 2016), and metalinguistic awareness expands significantly (Bernstein, 1989). Indeed, similarly to the developmental parallels between early motoric, speech, and language milestones, the overlay of higher-level language development, prosodic stabilization, and the emergence of rhythmic beat gestures is not unexpected. Additional support for the continued development and refinement of gestures in later stages of spoken language development is evident in the literature on individuals speaking a second language (L2). Beat gestures are more frequently used by individuals who are more fluent in their L2 compared to individuals who are less fluent in their L2 (Nicoladis et al., 1999). Likewise, beat gestures were found to align with prosodically prominent syllables in L2 research (e.g., McCafferty, 2006). Nicoladis et al. (1999) speculated that children increase their use of beat gestures in their L2 when their mean length of utterance and overall prosodic variation increases. Thus, the entrainment of prosodic stress and other body movements for children in later stages of development, for bilingual speakers and for those with communication disorders and in disordered systems, remains a ripe area for future empirical work.

Implications for children with speech sound disorders

Despite increased interest in the patterns of gesture and speech combinations in children with language impairments and autism spectrum disorders (e.g., De Marchena & Eigsti, 2010; Ellawadi & Weismer, 2014; Evans, Alibali, & McNeill, 2001; Igualada et al., 2016; Iverson & Braddock, 2011; Mainela-Arnold, Alibali, Hostetter, & Evans, 2006), few researchers have investigated the interaction of speech and manual movements within disordered populations from a motoric perspective of entrainment (Corriveau & Goswami, 2009). Yet, spatiotemporal entrainment of speech and manual systems exhibited in typical populations (e.g., Kelso, Tuller, & Harris, 1983) may be at the foundation of anecdotal reports of interventions capitalizing on manual movements for the clinical management of individuals with childhood apraxia of speech (e.g., DeThorne, Johnson, Walder, & Mahurin-Smith, 2009; Rusiewicz & Rivera, 2017). In fact, manual gestures are frequently and seamlessly integrated into treatment routines for speech sound disorders, both for pacing rate of speech (e.g., slow tapping leads to slower speech) and for cueing more accurate speech production (e.g., releasing the fingers from a fist outward to mimic the spatiotemporal properties of /p/). It is also noteworthy that prosody may be an area of relative strength for individuals with severe motor speech disorders and low intelligibility (e.g., Patel, 2004) and that eliciting prosodic prominence may lead to greater intelligibility (e.g., Connaghan & Patel, 2013). Peppé (this volume) presents information about the role of prosody in disordered communication. Likewise, the continued study of the coordination of prosody with gesture may offer possible therapeutic benefits for individuals with communication disorders. In fact, Fujii and Wan's (2014) sound envelope processing (SEP) hypothesis posits that rhythmic, pulse-based entrainment and synchronization may stimulate cortical and subcortical networks, facilitating speech and language processing and subsequent rehabilitation. For instance, the use of hand gestures to enhance the prosodic characteristics of speech, specifically emphatic stress and rate of speech, for children with speech sound disorders may then also lead to greater communicative success due to improved speech sound accuracy and overall intelligibility. Additionally, continued work in this area may elucidate the specific impact of movement within music and rhythmic approaches to speech and language interventions for children with severe communication deficits. Such approaches include Melodic Intonation Therapy, which exaggerates the prosodic characteristics of speech by overlaying singing for simple phrases while tapping with one's hand at the same time (Albert, Sparks, & Helm, 1973) and méthode verbo tonale, which capitalizes on the relationship of body movements and intonation to improve the ability to perceive, distinguish, and produce speech sounds for individuals who are deaf and hard-of-hearing (e.g., Renard, 1979). In short, continued exploitation of entrained prosodic prominence and gesture movements will likely lead to meaningful clinical change for children with communication disorders.

Conclusions

The last 20 years have seen a tremendous surge of interest in gestures. As we continue to increase our understanding of the underlying mechanism, which interconnects gestures and spoken language production, the essential role of prosody is once again affirmed. The investigation of the coordination of prosodic characteristics of speech and body movements, such as hand gestures, will provide fundamental information about the dynamic nature of linguistic and motoric processes in typically developing children, as well as extend to better diagnosis and intervention for children with speech and other communication disorders.

Dittmann and Llewellyn's (1969) comment from almost 50 years ago, "the really interesting question which this research has raised are those of why body movements should be located as they are in the rhythmic stream of speech"

(p. 104), still remains of high interest to this day. This interest spans beyond theoretical conjecture to various studies of the shared neurological, behavioral, and even evolutionary substrates of the speech and manual systems. Indeed, there is much evidence that cortical areas share activation for both observation and production of mouth and hand movements (see Yang, Andric, & Mathew, 2015, for a complete review). Although the precise neurological connections of prosodic structure and hand movements are unknown to date, it can be conjectured that the fundamental motoric pulse shared between prosodically prominent speech gestures and manual gestures likely is rooted in shared neurological substrates. For instance, the significance of hand-mouth connections in human mirror neuron systems is found in Gentilucci and colleagues' research that concluded that grasping different size objects affects the kinematic and acoustic magnitude of syllables produced by both adults and children (e.g., Gentilucci & Dalla Volta, 2007; Gentilucci, Santunione, Roy, & Stefanini, 2004).

The prosodic and gesture systems develop in a parallel and dynamic way, and this is reflected in how these systems are temporally coordinated throughout development. Pre-lexical infants show pulse-entrainment rules regulating the prosodic-gesture coordination, and at the one-word stage, infants control the alignment of prosodic and gesture prominences. The prosodic-gesture entrainment continues to develop in later stages, beat gestures being produced rhythmically with speech in order to organize and highlight information in the discourse when children are around 5–7 years of age. Thus, continued study of the underlying mechanism of the coordination of prosody and gesture will support our understanding of the dynamic nature of prosody in both developmental and disordered processes.

Funding

This work has benefited from support from the Labex BLRI (ANR-11-LABX-0036) grant under the project title Investments of the Future A*MIDEX (ANR-11-IDEX-0001–02), a grant from the Spanish Ministry of Economy, Industry, and Competitivity (FJCI-2015-26845), and the Childhood Apraxia of Speech Association of North America Research Grant.

Acknowledgements

We are grateful to the reviewers, whose thoughtful comments strengthened this work. We also appreciate the work of Sarah Leech for proofreading and managing references.

References

Acredolo, L., & Goodwyn, S. (1988). Symbolic gesturing in normal infants. *Child Development*, 59, 450–466. doi:10.2307/1130324

Albert, M. L., Sparks, R. W., & Helm, N. A. (1973). Melodic intonation therapy for aphasia. *Archives of Neurology*, 29(2), 130–131. doi:10.1001/archneur.1973.00490260074018

Barrett, K., & Campos, J. J. (1987). Perspectives on emotional development II: A functionalist approach to emotions. In Osofsky J. D. (Ed.), *Handbook of infant development* (pp. 555–578). Oxford: John Wiley & Sons.

Bates, E., Camaioni, L., & Volterra, V. (1975). The acquisition of performatives prior to speech. *Merrill-Palmer Quarterly of Behavior and Development*, 21(3), 205–226.

Bates, E., & Dick, F. (2002). Language, gesture, and the developing brain. *Developmental Psychobiology*, 40, 293–310. doi:10.1002/dev.10034

Bernstein, D. (1989). *Language development: The school age years. Language and communication disorders in children*. Columbus, OH: Merrill/MacMillon.

Birdwhistell, R. L. (1952). *Introduction to kinesics. An annotation system for analysis of body motion and gesture*. Louisville, KY: University of Louisville Press.

Butcher, C., & Goldin-Meadow, S. (2000). Gesture and the transition from one- to two-word speech: When hand and mouth come together. In McNeill D. (Ed.), *Language and gesture* (pp. 235–257). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511620850.015

Capone, N. C., & McGregor, K. K. (2004). Gesture development: A review for clinical and research practices. *Journal of Speech, Language, and Hearing Research*, 47(1), 173–186.

Capone, N. C., & McGregor, K. K. (2005). The effect of semantic representation on toddlers' word retrieval. *Journal of Speech, Language, and Hearing Research*, 48, 1468–1480. doi:10.1044/1092-4388(2005/102)

Cole, P. M., & Moore, G. A. (2015). About face! Infant facial expression of emotion. *Emotion Review*, 7(2), 116–120. doi:10.1177/1754073914554786

Colletta, J. M., Guidetti, M., Capirci, O., Cristilli, C., Demir, O. E., Kunene-Nicolas, R. N., & Levine, S. (2015). Effects of age and language on co-speech gesture production: An investigation of French, American, and Italian children's narratives. *Journal of Child Language*, 42(1), 122–145. doi:10.1017/S0305000913000585

Connaghan, K. P., & Patel, R. (2013). Impact of prosodic strategies on vowel intelligibility in childhood motor speech impairment. *Journal of Medical Speech-Language Pathology*, 20(4), 133–139.

Cook, S. W., Duffy, R. G., & Fenn, K. M. (2013). Consolidation and transfer of learning after observing hand gesture. *Child Development*, 84(6), 1863–1871. doi:10.1111/cdev.12097

Corriveau, K. H., & Goswami, U. (2009). Rhythmic motor entrainment in children with speech and language impairments: Tapping to the beat. *Cortex*, 45(1), 119–130. doi:10.1016/j.cortex.2007.09.008

Crais, E. R., Watson, L. R., & Baranek, G. T. (2009). Use of gesture development in profiling children's prelinguistic communication skills. *American Journal of Speech-Language Pathology*, 18, 95–108. doi:10.1044/1058-0360(2008/07-0041)

De Marchena, A., & Eigsti, I. M. (2010). Conversational gestures in autism spectrum disorders: Asynchrony but not decreased frequency. *Autism Research*, 3(6), 311–322. doi:10.1002/aur.159

De Ruiter, J. P. (1998). *Gesture and speech production*. PhD dissertation. Katholieke Universiteit, Nijmegen.
De Ruiter, J. P. (2000). The production of gesture and speech. In D. McNeill (Ed.), *Language and gesture* (pp. 284–311). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511620850.018
DeThorne, L. S., Johnson, C. J., Walder, L., & Mahurin-Smith, J. (2009). When "Simon Says" doesn't work: Alternatives to imitation for facilitating early speech development. *American Journal of Speech-Language Pathology*, 18(2), 133–145. doi:10.1044/1058-0360(2008/07-0090)
Dittmann, A. T., & Llewellyn, L. G. (1969). Body movement and speech rhythm in social conversation. *Journal of Personality and Social Psychology*, 11(2), 98–106. doi:10.1037/h0027035
Ellawadi, A. B., & Weismer, S. E. (2014). Assessing gestures in young children with autism spectrum disorder. *Journal of Speech, Language, and Hearing Research*, 57(2), 524–531. doi:10.1044/2013_JSLHR-L-12-0244
Ejiri, K., & Masataka, N. (2001). Co-occurrence of preverbal vocal behavior and motor action in early infancy. *Developmental Science*, 4(1), 40–48. doi:10.1111/1467-7687.00147
Esteve-Gibert, N., Borràs-Comes, J., Asor, E., Swerts, M., & Prieto, P. (2017). The timing of head movements: the role of prosodic heads and edges. *Journal of the Acoustical Society of America*, 141(6), 4727–4739.
Esteve-Gibert, N., & Prieto, P. (2013). Prosodic structure shapes the temporal realization of intonation and manual gesture movements. *Journal of Speech, Language, and Hearing Research*, 56(3), 850–865. doi:10.1044/1092-4388(2012/12-0049)
Esteve-Gibert, N., & Prieto, P. (2014). Infants temporally coordinate gesture-speech combinations before they produce their first words. *Speech Communication*, 57, 301–316. doi:10.1016/j.specom.2013.06.006
Esteve-Gibert, N., Prieto, P., & Pons, F. (2015). Nine-month-old infants are sensitive to the temporal alignment of prosodic and gesture prominences. *Infant Behavior and Development*, 38, 126–129. doi:10.1016/j.infbeh.2014.12.016
Evans, J. L., Alibali, M. W., & McNeil, N. M. (2001). Divergence of verbal expression and embodied knowledge: Evidence from speech and gesture in children with specific language impairment. *Language and Cognitive processes*, 16(2–3), 309–331. doi:10.1080/01690960042000049
Fujii, S., & Wan, C. Y. (2014). The role of rhythm in speech and language rehabilitation: The SEP hypothesis. *Frontiers in Human Neuroscience*, 8, 1–15. doi:10.3389/fnhum.2014.00777
Gentilucci, M., & Dalla Volta, R. (2007). The motor system and the relationships between speech and gesture. *Gesture*, 7(2), 159–177. doi:10.1075/gest.7.2.03gen
Gentilucci, M., Santunione, P., Roy, A. C., & Stefanini, S. (2004). Execution and observation of bringing a fruit tothe mouth affect syllable pronunciation. *European Journal of Neuroscience*, 19(1), 190–202. doi:10.1111/j.1460-9568.2004.03104.x
Goodwyn, S. W., Acredolo, L. P., & Brown, C. A. (2000). Impact of symbolic gesturing on early language development. *Journal of Nonverbal Behavior*, 24(2), 81–103. doi:10.1023/A:1006653828895
Green, J. R., Moore, C. A., Higashikawa, M., & Steeve, R. W. (2000). The physiologic development of speech motor control: Lip and jaw coordination. *Journal of Speech, Language and Hearing Research*, 43(1), 239–255. doi:10.1044/jslhr.4301.239

Green, J. R., Moore, C. A., & Reilly, K. J. (2002). The sequential development of jaw and lip controlfor speech. *Journal of Speech, Language, and Hearing Research*, 45(1), 66–79. doi:10.1044/1092-4388(2002/005)

Guellaï, B., Langus, A., & Nespor, M. (2014). Prosody in the hands of the speaker. *Frontiers in Psychology*, 5, 00700.

Hulit, L. M., Fahey, K. R., & Howard, M. R. (2015). *Born to talk: An introduction to speech and language development*. Boston, MA: Pearson.

Igualada, A., Bosch, L., & Prieto, P. (2015). Language development at 18 months is related to multimodal communicative strategies at 12 months. *Infant Behavior and Development*, 39, 42–52. doi:10.1016/j.infbeh.2015.02.004

Igualada, A., Esteve-Gibert, N., & Prieto, P. (2017). Beat gestures improve word recall in 3- to 5-year old children. *Journal of Experimental Child Psychology*, 156, 99–112.

Igualada, A., Wray, C., Prieto, P., & Norbury, C. (2016). Synchronous gesture-prosody alignments in narrations by children with specific language impairment (SLI). Paper presented at the *VIIIth International Conference of Language Acquisition (AEAL)*, Palma, Spain.

Ito, J .P. (2012). Focal impulses and expressive performance. In M. Aramaki, M. Barthet, R. Kronland-Martinet, & S. Ystad (Eds.), *From sounds to music and emotions* (pp. 480–489). Berlin: Springer.

Iverson, J. M., & Braddock, B. A. (2011). Gesture and motor skill in relation to language in children with language impairment. *Journal of Speech, Language, and Hearing Research*, 54(1), 72–86. doi:10.1044/1092-4388(2010/08-0197)

Iverson, J. M., Capirci, O., & Caselli, M. C. (1994). From communication to language in two modalities. *Cognitive Development*, 9(1), 23–43. doi:10.1016/0885-2014(94)90018-3

Iverson, J. M., & Fagan, M. K. (2004). Infant vocal-motor coordination: Precursor to the gesture-speech system? *Child Development*, 75(4), 1053–1066. doi:10.1111/j.1467-8624.2004.00725.x

Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science*, 16(5), 367–371. doi:10.1111/j.0956-7976.2005.01542.x

Iverson, J. M., & Thelen, E. (1999). Hand, mouth and brain. The dynamic emergence of speech and gesture. *Journal of Consciousness Studies*, 6, 19–40.

Kelso, J. S., Tuller, B., & Harris, K. S. (1983). A "dynamic pattern" perspective on the control and coordination of movement. In P. F. MacNeilage (Ed.), *The production of speech* (pp. 137–173). New York, NY: Springer. doi:10.1007/978-1-4613-8202-7_7

Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. In M. R. Key (Ed.), *The relationship of verbal and nonverbal communication* (pp. 207–227). The Hague: Mouton.

Kent, R. D., & Bauer, H. R. (1985). Vocalization of one-year-olds. *Journal of Child Language*, 12, 491–526. doi:10.1017/S0305000900006620

Kim, J., Cvejic, E., & Davis, C. (2014). Tracking eyebrows and head gestures associated with spoken prosody. *Speech Communication*, 57, 317–330. doi:10.1016/j.specom.2013.06.003

Krahmer, E., & Swerts, M. (2007). The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language*, 57(3), 396–414. doi:10.1016/j.jml.2007.06.005

Krauss, R. M., Chen, Y., & Gottesman, R. F. (2000). Lexical gestures and lexical access: A process model. In D. McNeill (Ed.), *Language and gesture* (pp. 261–283). New York, NY: Cambridge University Press. doi:10.1017/CBO9780511620850.017

Krivokapić, J., Tiede, M., Tyrone, M., & Goldenberg, D. (2016). *Speech and manual gesture coordination in a pointing task*. Paper presented at the meeting Framing speech: Celebrating 40 years of inquiry with Stefanie Shattuck-Hufnagel, Boston, USA.

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people tracktime-varying events. *Psychological Review*, 106(1), 119–159. doi:10.1037/0033-295X.106.1.119

Leonard, T., & Cummins, F. (2011). The temporal relation between beat gestures and speech. *Language and Cognitive Processes*, 26(10), 1457–1471. doi:10.1080/01690965.2010.500218

Mainela-Arnold, E., Evans, J. L., & Alibali, M. W. (2006). Understanding conservation delays in children with Specific Language Impairment: Task representations revealed in speech and gesture. *Journal of Speech, Language, and Hearing Research*, 49, 1267–1279. doi:10.1044/1092-4388(2006/091)

Mathew, M., & Manjula, R. (2015). Patterns in the emergence of hand actions in typically developing infants. *Journal of Communication Disorders, Deaf Studies & Hearing, Aids*, 3, 140.

Mathew, M., Yuen, I., Shattuck-Hufnagel, S., Ren, A., & Demuth, K. (2014). *The development of prosodic gestures during parent-child discourse interactions*. Paper presented at the 14th Conference on Laboratory Phonology, Tokyo, Japan.

Mathew, M., Yuen, I., & Demuth, K. (2017). Talking to the beat: Six-year-olds' use of stroke-defined non-referential gestures. *First Language*. doi:10.1177/0142723717734949

McCafferty, S .G. (2006). Gesture and the materialization of second language prosody. *IRAL-International Review of Applied Linguistics in Language Teaching*, 44(2), 197–209. doi:10.1515/IRAL.2006.008

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago, IL: The University of Chicago Press.

Merker, B., Madison, G., & Eckerdal, P. (2009). On the role and origin of isochrony in human rhythmic entrainment. *Cortex*, 45, 1–17. doi:10.1016/j.cortex.2008.06.011

Murillo, E., & Belinchón, M. (2012). Gestural-vocal coordination: Longitudinal changes and predictive value on early lexical development. *Gesture*, 12(1), 16–39. doi:10.1075/gest.12.1.02mur

Murillo, E., & Capilla, A. (2016). Properties of vocalization-and gesture-combinations in the transition to first words. *Journal of Child Language*, 43(4), 890–913.

Nicoladis, E., Mayberry, R. I., & Genesee, F. (1999). Gesture and early bilingual development. *Developmental Psychology*, 35(2), 514–526. doi:10.1037/0012-1649.35.2.514

Oller, D. K., Wieman, L., Doyle, W., & Ross, C. (1976). Infant babbling and speech. *Journal of Child Language*, 3, 1–11. doi:10.1017/S0305000900001276

Owens, R. E. (2016). *Language development: An introduction*. Boston, MA: Pearson.

Özçalışkan, S., Gentner, D., & Goldin-Meadow, S. (2013). Do iconic gestures pave the way for children's early verbs? *Applied Psycholinguistics*, 35(6), 1143–1162. doi:10.1017/S0142716412000720

Özçalışkan, S., & Goldin-Meadow, S. (2005). Gesture is at the cutting edge of early language development. *Cognition*, 96, B101–B113. doi:10.1016/j.cognition.2005.01.001

Özçalışkan, S., & Goldin-Meadow, S. (2011). Is there an iconic gesture spurt at 26 months? In S. Gale, & I. Mika (Eds.), *Integrating gestures: The interdisciplinary nature of gesture*, (pp. 163–174). Amsterdam: John Benjamins.

Patel, R. (2004). The acoustics of contrastive prosody in adults with cerebral palsy. *Journal of Medical Speech-Language Pathology*, 12(4), 189–193.

Patel, R., & Brayton, J. T. (2009). Identifying prosodic contrasts in utterances produced by 4-, 7-, and 11-year-old children. *Journal of Speech Language and Hearing Research*, 52, 790–801. doi:10.1044/1092-4388(2008/07-0137)

Patel, R., & Grigos, M. I. (2006). Acoustic characterization of the question-statement contrast in 4, 7, and 11 year-old children. *Speech Communication*, 48, 1308–1318. doi:10.1016/j.specom.2006.06.007

Port, R. (2003). Meter and speech. *Journal of Phonetics*, 31, 599–611. doi:10.1016/j.wocn.2003.08.001

Renard, R. (1979). *Introduction à la méthode verbo-tonale de correction phonétique*. Bruxelles: Didier, Mons: CIPA.

Rochet-Capellan, A., Laboissierre, R., Galvan, A., & Schwartz, J. (2008). The speech focus position effect on jaw-finger coordination in a pointing task. *Journal of Speech, Language, and HearingResearch*, 51, 1507–1521. doi:10.1044/1092-4388(2008/07-0173)

Rusiewicz, H. L. (2011, September). *Synchronization of speech and gesture: A dynamic systems perspective*. Paper presented at the 2nd Gesture and Speech in Interaction (GESPIN), Bielefeld, Germany.

Rusiewicz, H. L. & Rivera, J. L. (2017). The effect of hand gesture cues within the treatment of /r/ for a college-aged adult with persisting childhood apraxia of speech. *American Journal of Speech-Language Pathology*, 26(4), 1236–1243.

Rusiewicz, H. L., Shaiman, S., Iverson, J. M., & Szuminsky, N. (2013). Effects of prosody and position on the timing ofdeictic gestures. *Journal of Speech, Language, and Hearing Research*, 56(2), 458–470. doi:10.1044/1092-4388(2012/11-0283)

Rusiewicz, H. L., Shaiman, S., Iverson, J. M., & Szuminsky, N. (2014). Effects of perturbation and prosody on the coordination of speech and gesture. *Speech Communication*, 57, 283–300. doi:10.1016/j.specom.2013.06.004

Sauer, E., Levine, S. C., & Goldin-Meadow, S. (2010). Early gesture predicts language delay in childrenwith pre- or perinatal brain lesions. *Child Development*, 81(2), 528–539. doi:10.1111/j.1467-8624.2009.01413.x

Smith, A., & Zelaznik, H. N. (2004). Development of functional synergies for speech motor coordination in childhood and adolescence. *Developmental Psychobiology*, 45(1), 22–33. doi:10.1002/dev.20009

Steve, R. (2010). Babbling and chewing: Jaw kinematics from 8 to 22 months. *Journal of Phonetics*, 38(3), 445–458. doi:10.1016/j.wocn.2010.05.001

Tellier, M. (2008). The effect of gestures on second language memorisation by young children. *Gesture*, 8, 219–235. doi:10.1075/gest.8.2.06tel

Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A new look at infant pointing. *Child Development*, 78(3), 705–722. doi:10.1111/j.1467-8624.2007.01025.x

Tuite, K. (1993). The production of gesture. *Semiotica*, 93, 8–105. doi:10.1515/semi.1993.93.1-2.83

Volterra, V., Caselli, M. C., Capirci, O., & Pizzuto, E. (2005). Gesture and the emergence and development of language. In M. Tomasello, & D. Slobin (Eds.), *Beyond nature-nurture. Essays in honor of Elizabeth Bates* (pp. 3–40). Mahwah, NJ: Lawrence Erlbaum Associates.

Yang, J., Andric, M., & Mathew, M. M. (2015). The neural basis of hand gesture comprehension: Ameta-analysis of functional magnetic resonance imaging studies. *Neuroscience & Bio behavioral Reviews*, 57, 88–104. doi:10.1016/j.neubiorev.2015.08.006

CHAPTER 7

Speech rhythm in development

What is the child acquiring?

Brechtje Post & Elinor Payne
University of Cambridge / University of Oxford

Perception and production studies of speech rhythm development in infants and children paint a complex picture of a universal early perceptual sensitivity to – and production mastery of – cues to rhythm, while the rate of acquisition of rhythmic properties across and within languages appears to be typologically and structurally determined. In this chapter we provide a critical and comprehensive review of the literature that has led to these insights. We then explore how child rhythm development can be accommodated in an integrated approach to speech rhythm in which various structural and performance aspects interact to determine developmental trajectories in rhythm acquisition.

Introduction

Despite the large body of research devoted to it, speech rhythm remains an elusive and controversial phenomenon. Debates over its nature and provenance, and indeed even over its very existence, have surfaced repeatedly in the literature over the last few decades, and continue to this day (e.g. Arvaniti, 2012; Cummins, 2015; Nolan & Jeon, 2014; White, 2014). Attempts to define or pin down speech rhythm have invariably invited some form of dissent from other scholars, and the diverging views on rhythm have also brought different methodological approaches. Nevertheless, for those unfamiliar with the literature, and as a reasonable heuristic for understanding the frame of reference of this body of work, the phenomenon referred to as rhythm in connected speech can, without too much controversy we believe, be understood to refer to the organization and contrast of a sequence of repeated events. An example of such an event could be, for instance, syllable stress, occurring as it does repeatedly and, for example in English, in contrastive alternation with lack of stress, and in specific ways. Scholars may disagree as to the nature of these events and their co-ordination (apparent or otherwise), whether they are

DOI 10.1075/tilar.23.07pos

planned or not, which factors shape the patterning, and whether their rhythmic alternation is perceived or has physical reality.

We briefly review the evolution of the debate on speech rhythm in the first section, highlighting some of the main positions, with a condensed sketch of our own approach. This serves as a backdrop to the principal purpose of this chapter, which is a critical review of research investigating rhythm in developmental speech, be that in perception (Section "Perception"), production (Section "Production"), or bilingual children (Section "Bilingual production"). The scope of this chapter is limited to research on rhythm in a 'holistic' sense, and does not seek to provide an account of the development of individual prosodic properties (e.g. Bhatara, Boll-Avetisyan, Höhle, & Nazzi, this volume), though the research we review assumes that language-specific differences in those properties almost certainly play a role in rhythm development. In the last section we discuss how these previous strands of research can be brought together within our integrated approach to speech rhythm, and we explore implications for future avenues of research.

Background: Rhythm research and methodology

The observation that languages vary in their rhythmic organization – or at least are perceived as doing so – has a long history (see Armstrong, 1932; Lloyd James, 1940, for early discussion), and was traditionally explained as being rooted in languages privileging different phonological units (mora, syllable or foot). Early attempts to identify the acoustic basis of the percept that arises include Abercrombie (1967), Pike (1945), and Ladefoged (1975) and focus on timing, with the claim that languages tend towards isochrony in the particular unit that they privilege. Thus, in what can be thought of as the first wave of speech rhythm research, the rhythm percept was famously described as being classed accordingly as 'stress-timed', 'syllable-timed', or 'mora-timed'. Exhaustive acoustic investigation failed to find conclusive evidence of syllable, mora or foot isochrony in speech production (see Cutler, 1991, for a review); nevertheless, the percept of a difference *of some sort* is strong and supported by subsequent psycholinguistic research, including in infants (Nazzi, Bertoncini, & Mehler, 1998; Nazzi, Jusczyk, & Johnson, 2000; cf. Bhatara, Boll-Avetisyan, Höhle, & Nazzi, this volume).[1]

1. We retain the terms 'stress-timed' and 'syllable-timed' since they are the prevailing terms in the literature for what are, we would argue, the endpoints of a gradiently varying rhythmic continuum (or continua) in perception and production. This does not imply that we believe that languages uniformly show isochronous timing, nor that they can be categorically classified.

In a second wave of speech rhythm research, alternative explanations for the source of perceived rhythmic difference were put forward (Bertinetto, 1981; Dasher & Bolinger, 1982; Dauer, 1983, 1987; Roach, 1982). In particular, it was proposed that the rhythm percept emerges from distinct language-specific sets of phonological and phonetic properties, including – but not limited to – the complexity of syllable structure and the presence versus absence of vowel reduction. Thus, so-called 'stress-timed' languages such as English have a greater range of syllable structures, a propensity for heavier syllables to attract stress, and a tendency towards reduction of unstressed vowels, while in so-called 'syllable-timed' languages such as French, open syllables are far more common and vowel reduction is less evident. So even while isochrony had by this point been rejected, the idea of rhythm classification persisted and the search for possible class correlates was cast more widely.

This hypothesis, which has been referred to as the 'phonology-derived' hypothesis, in effect attempts to capture differences in the degree to which languages use duration to convey prosodic organization and prominence. The hypothesis inspired the formulation of various types of 'rhythm metrics' designed to quantify the claimed greater variability of duration in vocalic and consonantal intervals and the claimed smaller overall amount of vocalic material in 'stress-timed' languages. The interval measures developed by Ramus and colleagues (Ramus & Mehler, 1999; Ramus, Nespor, & Mehler, 1999) were also motivated and validated by the discovery that infants attend to rhythmic differences from birth (Nazzi et al., 1998), indicating that the differences can be captured objectively from the acoustic stream, i.e. before lexical or phonological analysis is available to the infant. From this, Ramus et al. (1999: 270) argue that "the infant primarily perceives speech as a succession of vowels of variable durations and intensities, alternating with periods of unanalysed noise (i.e., consonants)". This insight led Ramus et al. (1999) to develop three measures of utterance rhythm which can be extracted purely on acoustic grounds, on the basis of vocalic and consonantal intervals: (i) the standard deviation of vocalic intervals (ΔV); (ii) the standard deviation of intervocalic (i.e., consonantal) intervals (ΔC); and (iii) the proportion of total utterance duration which comprises vocalic intervals (%V). Ramus, Dupoux, and Mehler (2003) showed similar patterns of adult between-language discrimination to those found for young infants.

Independently, and around the same time, Grabe and colleagues (Grabe & Low, 2002; Low, Grabe, & Nolan, 2000) developed the Pairwise Variability Index (PVI), which, while also based on vocalic and consonantal intervals, has the added factor of characterising the variability in sequential pairs. Rather than taking a global measure of the amount and variability in vocalic material, the PVI attempts to capture sequential differences in vocalic interval duration, and specifically between stressed and unstressed syllables. The motivation for looking at the sequential nature of the contrast is that prosodic prominence depends on syntagmatic contrast: on this view, the critical comparison is with what has gone immediately before and with what lies immediately ahead. The PVI is calculated as the mean of differences between successive intervals, and is normalized (nPVI) for variability of speech rate by dividing by the sum of intervals.

Over the last 15 years or so, the methodology developed in this second wave of speech rhythm research has generated a large body of research that has applied the metrics in different linguistic contexts. It has also attracted close, and often critical, scrutiny of the value and legitimacy of metrics in capturing rhythm: metric scores are noted to vary widely over different types of elicitation, different material within a single language, different speakers and speech contexts, with the result that nothing like robustly predictable scores for languages, let alone for rhythm classes, can be established. The reported inconsistency in the results produced by metrics has led to what may be called a third wave of research, with several scholars (e.g. Arvaniti, 2012; Rathcke & Smith, 2015) questioning or even rejecting the use of interval-based metrics, either as lacking in explanatory power, or indeed, more fundamentally, as simply mistaken in approach. However, some of this criticism of metrics arises from an implicit, but unnecessary, assumption that the metrics are intrinsically tied to the rhythm class hypothesis. As a consequence, since the idea of rhythm classification itself has been discredited, the metrics themselves also appear to have been discredited by association.

Instead, we argue that the metrics should be seen, in substance, as a neutral tool for comparing holistic timing characteristics, many or all of which may contribute to linguistic discrimination by both adults and infants. Their use does not imply, or plead for, the existence of robust, invariable categories in either the speech stream itself or how this is perceived. Nor does it imply that discrimination is made on the basis of durational correlates alone. The rhythm metrics can provide a lens for examining some aspects of the rhythm percept, just as any other acoustic metric used to examine the percept of a given phenomenon in speech (e.g. Voice Onset Time for different language-specific categories of stop consonants). The differing patterns of findings are to be considered within this context, rather than thrown out with the proverbial bathwater just because they fail to deliver a single robust and fixed measure.

Other third wave critics questioned the narrowness of the earlier waves, arguing that the notion of speech rhythm needs to be conceptualized in a much broader framework, either by taking into account other biological as well as non-biological phenomena involving regularly recurring events (e.g. music, Hawkins, Cross, & Ogden, 2013), or questioning that the notion of a rhythm percept is truly meaningful with regard to speech (e.g. Hawkins, 2014; Nolan & Jeon, 2014; White, 2014). Although the development of many of these other types of rhythm has been studied quite extensively (e.g. Phillips-Silver & Trainor, 2005), to our knowledge, none of the third wave insights have been explored in the study of child speech.

Our own approach seeks to reconcile the apparently conflicting waves by incorporating their key insights into an integrated account of speech rhythm (Payne, Post, Astruc, Prieto, & Vanrell, 2010; Payne, Post, Prieto, Vanrell, & Astruc, 2012). We take a holistic view in which the rhythmic 'experience' is an experience of the temporal[2] organization of a sequence of similar events or objects, and of other parameters defining the relationship between these events/objects, i.e. contrastive properties – also in speech. In seeking the source of the speech rhythm percept in production we must then identify *all* potential sources of organization and contrast. These lie not solely in the handful of structural properties identified in the first and second waves of speech rhythm research, but in a much broader set of phonological and prosodic structures (Prieto, Vanrell, Astruc, Payne, & Post, 2012), as well as non-linguistic factors that shape the (variable) implementation of those structures (e.g. speaking style), all of which are underpinned by neurological constraints that shape our experience of rhythm. Thus, on our interpretation, speech rhythm is *multi-systemic*, since it involves the complex interaction of structural properties of linguistic systems; it is *multi-componential*, since it encompasses the linguistic component as well as (neuro)biological and performance components (e.g. adaptation in child-directed speech, Payne et al., 2010); but it is also *multi-parametric*, since the organization and contrasting of rhythmic events in time is signaled by multiple acoustic parameters, including duration, but also fundamental frequency and intensity. We will explore the general implications of such an approach for child rhythm research in the last section of this chapter.

Perception

Although more recent studies (cf. White, Delle Luche, & Floccia, 2016) have since contested certain findings and interpretations, much of the groundwork in the perception of rhythm in language development was laid down by a series of psycholinguistic studies in the late 1990s and early 2000s (see in particular Nazzi et al., 1998, 2000; Nazzi & Ramus, 2003; Ramus, 2002; also Bhatara, Boll-Avetisyan, Höhle, & Nazzi, this volume). Arguably the main purpose of this body of work was not to examine the acquisition of rhythm per se, but rather to investigate the role of rhythm in facilitating language learning overall. As such, it continues a vein

2. Or perhaps spatial, if we extend the notion to the visual domain.

of research into the 'prosodic bootstrapping' hypothesis (see de Carvalho, Dautriche, Millotte, & Christophe, this volume). Rhythm was proposed early on as a possible means by which the infant is introduced to pre-lexical phonology, that is, to the representation of the speech signal that is best suited to her native language. Initial studies examined the influence of rhythmic structure of a language on the process of spoken word recognition and thus of word formation and lexical access, with evidence suggesting that listeners are guided by the assumption that word boundaries typically coincide with the boundaries of the privileged rhythmic units of their native language (as shown by Cutler and colleagues for adults; e.g. Cutler, 1991). Infants were shown to exploit rhythmic cues to segmentation at an earlier developmental stage than other cues (Echols, Crowhurst, & Childers, 1997; Johnson & Jusczyk, 2001; Jusczyk, 1997, 1999; Jusczyk, Houston, & Newsome, 1999), prompting the claim that rhythm is key to building a lexicon and going on to acquire syntax (Morgan, 1986).

Subsequent studies also began to investigate infants' sensitivity to rhythm in a more holistic way (i.e. not focusing on specific word-level properties like stress or word boundaries, but sentence-level recurring rhythmic patterns). In a series of studies on French newborns, Nazzi et al. (1998) establish that very young infants can discriminate between languages on the basis of prosodic information alone (also de Carvalho, Dautriche, Millotte, & Christophe, this volume). The infants discriminated between supposed rhythm classes (e.g. between 'stress-timed' English and Dutch on the one hand and 'syllable-timed' Italian and Spanish), but failed to discriminate within a class. In a further study, Nazzi et al. (2000) showed similar discrimination patterns at 5 months, but additionally found that discrimination between two languages within the same rhythm class was possible provided one of the languages was the infants' native language. Interestingly, Ramus (2002) established that discrimination was facilitated for newborns when they could exploit pitch cues in addition to purely durational rhythmic cues, supporting our multi-parametric view of rhythm development (Payne et al., 2012; cf. Frota & Butler, this volume, for intonation).

There is a running, more or less implicit, assumption of a classificatory basis to this series of studies, reflected in a reworked version of the prosodic bootstrapping hypothesis (see especially Nazzi & Ramus, 2003), whereby the emergence in infancy of speech segmentation procedures – deemed crucial for word acquisition – is specific to the so-called classes of speech rhythm. The idea of neat rhythm classes is compatible with a UG framework, with the setting of parameters determining the grammatical properties of the language and directing the infant on a subsequent guided pathway. However, the claim that speech rhythm falls into neat classes is highly contentious; not only is the cross-study, cross-speaker evidence in speech production weak, recent research by White et al. (2016) has critically undermined the rhythm class hypothesis also in perception. In a study employing similar age profile and methodology to those used by Nazzi et al. (2000) but selecting different pairs of languages, they show that, *contra* the rhythm class hypothesis, 5-month-old English-ambient infants discriminate between intact Spanish and French (i.e. within class), but not between Finnish and French, or between Finnish and Spanish (i.e. also within class). These findings suggest that infants do not discriminate on the basis of rhythm class membership, but on a variety of other parameters (cf. White et al., 2016). One such parameter is likely to be sensitivity to final lengthening, as shown for infants by White, Floccia, Goslin, and Butler (2014) and specific durational cues more generally, as shown for adults in White, Mattys, and Wiget (2012). Interestingly, in a further, as yet unpublished, experiment, they found that, unlike 5-month-olds, 8-month-olds *did* discriminate between Finnish and Spanish, a finding that is worthy of more investigation since it suggests that discriminatory sensitivity for rhythm may itself evolve with development.

However, the discrediting of rhythm *classes* does not negate the possibility of systematic variability in a global rhythmic percept, formed from multiple timing (and potentially other) factors. Thus, the seminal findings on infant sensitivity to global rhythmic properties, and that these properties differ cross-linguistically, remain significant. The possibility that these sensitivities assist infants in their learning about the structures of their language also remains. If one takes the view that rhythm is a holistic prosodic phenomenon that emerges from multiple, complex factors (Payne et al., 2010, 2012), then it follows that an infant exposed to a particular rhythm may be able to extract (and thus reverse engineer) information about the language that contributed to the generation of that particular rhythm in the first place.

Moreover, the significance of the evidence for an early sensitivity to rhythm does not hinge on the nativist angle, nor does it necessarily lead to a nativist conclusion. Ramus, Mehler, Morris, Miller, and Hauser (2000: 349) suggest that, while it may be that humans alone make meaningful use of spoken language, the manner in which human newborns tune into these critical properties of speech relies on general processes of the primate auditory system rather than a "unique constellation of perceptual and neurobiological mechanisms". They present evidence for this from experiments showing the ability in both human newborns and cotton-top tamarin monkeys to discriminate unfamiliar languages. This combined body of evidence fits comfortably with a constructionist account, and in fact, constructionist/emergentist accounts have long stressed the importance of rhythm by suggesting that it may provide a basic grid for the acquisition of language as such (e.g. Vihman, Davis, & DePaolis, 1995).

In sum, perception studies of rhythm development have shown that even neonates are sensitive to the rhythm of their native language, that they use multiple acoustic cues in rhythm perception, and that rhythm may be exploited in the segmentation of speech, hence allowing for the possibility that it plays a crucial role in bootstrapping the acquisition of lexical, syntactic, and other linguistic knowledge.

Production

The pivotal role attributed to rhythm for prosodic bootstrapping through perception finds a strong precursor in much earlier production work on rhythm development (Allen, 1983; Allen & Hawkins, 1978, 1980). Crucially, Allen and Hawkins (1980) claimed that rhythm confers predictability which can be exploited in child speech development. More specifically, they viewed rhythm as the result of "a number of concurrent phonological processes, involving not just the timing of intervals between syllables and stresses […] as well as a host of other phenomena" such as phonological vowel length, phonological voice, pre-final lengthening, and accentuation (Allen & Hawkins, 1980: 229). They argued that these 'phonological processes' give rise to predictability of the temporal-sequential shape of the utterance which is used to facilitate speech comprehension – much as Cutler and others had already claimed for stress (Cutler, 1976; Cutler & Foss, 1977; Huggins, 1972; Shields, McHugh, & Martin, 1974) – in early language acquisition.

Allen and Hawkins arrived at this conceptualisation of rhythm through a series of production studies in which they explored the acquisition of certain phonological phenomena. For instance, they established that vowel reduction and accentuation are difficult for children to master (1978, 1980). In taking this approach, their work pre-echoes the second wave 'phonology-derived' approach, at least in spirit, and it is also akin to the multi-systemic aspect of our approach (Payne et al., 2010, 2012). However, they did not attempt to directly relate the development they observed for individual phenomena to the overall development of speech rhythm in young infants and children, nor to the rhythm class hypothesis that held sway at the time, other than to suggest that child speech appears to be more 'syllable-timed' (Allen & Hawkins, 1980).

After a prolonged gap in child rhythm production research, Konopczynski (1995) took up Allen and Hawkins' notion of a 'phonological rhythm' in the first study to examine the developmental trajectory of rhythm acquisition cross-linguistically, which she operationalised in terms of syllable durations.[3] She

3. Her primary objective was in fact to chart universal tendencies in the development of stress in typologically different systems, but we would argue that her choice of measure does not allow her to relate her findings directly to the development of stress.

confirmed that there is indeed an early universal tendency towards 'syllable-timing', but the infants then proceeded to produce adult-like patterns at different ages depending on rhythm class, with 'syllable-timed' French and Hungarian acquired in the first half of the second year, 'in-between-timed' Spanish early in the third year, and 'stress-timed' English and Portuguese in the course of the third year.

These findings align well with Allen (1983) who concludes that two-year-olds "speak with French prosody a substantial part of the time", and Hallé, de Boysson-Bardies, and Vihman (1991), who showed that Japanese and French children produced different durational patterns, both in babbling and one-word utterances, by about 18 months (see also Frota & Butler, this volume). Konopczynski accounts for the later development of English and Portuguese in terms of syllable structure, the "naturalness of final lengthening" and the predictability of stress, echoing work by Vihman et al. (1995) who pointed out that the English trochaic strong-weak patterns are not easily mastered, and that the different components of rhythmic prominence may develop at different speeds in acquisition (see also Vihman, this volume).

Focusing on four-year old French, English, and German children, Grabe, Watson, and Post (1999) and Grabe, Gut, Post, and Watson (2001) looked at slightly older children, and applied the newly developed rhythm metrics for the first time to child speech data. Using the vocalic normalised Pairwise Variability Index (nPVI–V),[4] Grabe et al. (1999) found that the rhythmic patterns produced by French children did not differ significantly from those produced by their mothers, which suggests that adult-like durational patterning is achieved well before age 4 in French. The English children, by contrast, still showed much more even-timed rhythmic patterning than their mothers. This places the end-point of the acquisition process for English at a considerably later age than suggested by Konopczynski (1995).

Grabe et al. (1999: 1204) accounted for their findings in terms of the "greater complexity" of rhythmic structure in English as compared to French, where 'complexity' refers to properties like the predictability of accent placement, the phonetic parameters involved in cueing accents, and the need to compress or lengthen syllables in the context of a 'stress-timed' language. They concluded that children acquire speech rhythm moving from a structurally less variable 'syllable-timed' to an increasingly variable 'stress-timed' rhythmic pattern (cf. Li & Post, 2014; Ordin & Polyanskaya, 2014, for L2 rhythm acquisition). However, the comparison with German in Grabe et al. (2001) suggested that it is not just the *presence* of the properties that affects the rate of acquisition, but also the *degree* to which they

4. Note that this study predates the advent of other vocalic and consonantal rhythm metrics.

are present. 'Stress-timed' German rhythm was found to be in place at 4, unlike English, even though it shares the properties that have been claimed to contribute to the percept of 'stress-timing' for English. However, in many respects, German has those properties to a lesser extent (e.g. less vowel reduction and less final and accentual lengthening; Delattre, 1966, 1969).

Vihman and colleagues (DePaolis, Vihman, & Kunnari, 2008; Vihman, Nakai, & DePaolis, 2006) examined rhythm development longitudinally and cross-linguistically in young infants (10–18 months, including English, Welsh, French, and Finnish). Perhaps unsurprisingly, they confirmed the earlier findings of Allen (1983), Hallé et al. (1991), and Konopczynski (1995) that French infants are already on target by age 2, unlike English 2-year-olds, but they also found that Finnish infants have better control in their production of prosodic cues (measured in fundamental frequency, intensity and amplitude in DePaolis et al. (2008)). In fact, the Finnish data suggest that rhythmic separation between languages begins as early as the four-word point (around 1;0). They propose that the greater rhythmic variability that is inherent to 'stress-timing' is phonetically more challenging than 'syllable-timing', since the child needs to acquire more phonotactically complex and less frequent sequences and segments (Vihman et al., 2006). This meshes neatly with the complexity hypothesis of Grabe and colleagues (Grabe et al., 1999, 2001).

Payne et al. (2012) shed new light on this hypothesis by showing that early child speech is not necessarily categorically more 'syllable-timed' than adult speech, using interval-based rhythm metrics on data from English, Spanish and Catalan 2-, 4- and 6-year-olds. Although they found that child speech is indeed overall 'more vocalic' (higher %V) and – as has been found in all previous studies – *less* variable in vocalic interval duration (both characteristic of 'syllable-timing'), by contrast, consonantal intervals are much *more* variable, and can still be off target even at age 6. This implies that 'syllable-timing' – or what is loosely associated with this percept – is not an across-the-board intrinsic property of all aspects of early productions. Unexpectedly, they also found that the English children lagged behind the Spanish and Catalan children at all three ages, and had not yet fully mastered adult-like rhythm even at age 6.

Thus, Payne et al. (2012) painted a much more refined picture of emerging language-dependent rhythmic properties which do not develop in a synchronised or cumulative fashion either cross-linguistically, or within one and the same language. They suggested that vocalic interval variability could largely be a skill to be mastered as the relevant phonological knowledge develops, while consonant interval variability must at least in part be an artefact of immature articulatory co-ordination skills (interacting with language-specific phonotactics). The complexity of rhythm development, then, arises from the competing demands that are placed on the child's developing abilities (cf. Vihman et al., 2006). They concluded that a proper understanding of rhythm development requires a closer examination of the multiple phonetic and/or phonological sources that underlie it, requiring an analysis of the various prosodic factors involved (cf. Prieto et al., 2012), as well as segmental and syllable structures and their phonetic implementation.

In sum, children start out with more 'even-timed' rhythm (as quantified in syllabic and vocalic intervals), and produce more vocalic than consonantal material in their utterances overall, but language-specific rhythmic patterning starts to emerge from as early as 12 months. By contrast, consonantal material is more '*un*even' timed, with control being mastered later, and variability diminishing as the child gets older. The rate of acquisition depends on typological factors, with adult-like patterning as early as age 2 in more 'syllable-timed' languages, and off-target production still present at age 6 in 'stress-timed' English.

Bilingual production

Interestingly, there are just as many studies of the development of speech rhythm in simultaneous bilingual children as of monolinguals, but they all used rhythm metrics as their quantification method. In the bilingual literature, there are remarkable commonalities between the monolingual and bilingual groups that were compared in the experiments: early even-timing (in vocalic material) and divergent developmental paths related to rhythm typology. However, rhythmic separation between the two languages of the bilingual child occurred a little later than the emergence of language-specific rhythmic patterns in the monolingual data would suggest: between the ages of 3 and 4 (Bunta & Ingram, 2007; Kehoe, Lleó, & Rakow, 2011; Schmidt & Post, 2015a), and the languages were rhythmically more similar than in monolinguals. Kehoe et al. (2011) account for this 'intermediate' bilingual rhythmic patterning in terms of phonetic compromise.

Further evidence for typological effects was found in the asymmetrical influence of the ambient language on the developmental paths of the two languages in bilinguals (Bunta & Ingram, 2007; Lleó, Rakow, & Kehoe, 2007; Mok, 2011, 2013; Schmidt & Post, 2015a; Whitworth, 2002). For instance, in Mok (2011, 2013), bilingual Cantonese was comparable with monolingual Cantonese, while bilingual English was less variable and simpler in structure than monolingual English. This reflects the vulnerability of phonological features that are more challenging phonetically, according to Mok (2013), which appears to put bilinguals at a relative disadvantage. By contrast, when the effects of ambient language were factored out, the language interaction in bilinguals was found to confer an advantage, which was reflected in a faster mastery of consonantal timing (Schmidt & Post 2015a, cf. Kehoe et al., 2011).

Thus, the bilingual rhythm data confirm that typological factors play a key role in determining the rate of acquisition, and also suggest that rhythmic properties may not develop in a cumulative, or synchronised, fashion (cf. Payne et al., 2012).

The integrated approach to rhythm development

The perception and production studies reviewed in the previous sections have shown that language-specific rhythm emerges very early. This lends plausibility to the hypothesis that rhythm plays an important part in 'bootstrapping' language acquisition, since its early availability means that it can be exploited in the segmentation of speech. The literature also shows consistently that typological factors play a crucial role in the *rate* of rhythm acquisition. Although there is a widely attested early universal tendency towards 'syllable-timing' (at least in terms of the vocalic material produced by infants), children who are acquiring 'stress-timed' languages take much longer to fully master native-like rhythm. This finding has variably been accounted for in terms of the greater structural complexity of 'syllable-timed' languages (Grabe et al., 1999), the degree to which such structural properties occur in the language (Grabe et al., 2001), and the phonetic challenge they pose to the child (Mok, 2013; Vihman et al., 2006). The third main conclusion that can be drawn from the literature is that rhythm development is complex, and does not proceed in a linear, synchronised, or cumulative fashion, since vocalic and consonantal variability develop at different rates within as well as across languages. In Payne et al. (2012), we suggest that multiple different phonetic and phonological sources are likely to underlie the development of vocalic and consonantal variation.

How would our integrated approach (Payne et al., 2010, 2012) in which multiple systemic, componential, and parametric factors function as sources for the rhythm percept (Section "Background: Rhythm research and methodology"), account for the development of speech rhythm? What is it, precisely, that the child is acquiring?

We would contend, firstly, that the child needs to master a considerable number of different phonological and prosodic structural properties such as segmental properties, syllable structure types, metrical structure, lexical tone, pitch accent types, and boundary tones. As is well-known, these properties have their own developmental paths (Bhatara, Boll-Avetisyan, Höhle, & Nazzi, this volume). For instance, a number of studies have shown that the full set of intonational primitives (i.e. pitch accents and boundary tones) is in place by age 2 (e.g. Chen & Fikkert, 2007, for Dutch; and for Catalan, Prieto & Vanrell, 2009, and Prieto et al., 2012), while syllable structure is still not always fully acquired by age 4 (Payne et al., 2012). We predict that these differential rates of acquisition for the individual properties should be reflected in the development of the child's speech rhythm (cf. Frota & Butler, this volume, for intonation development).

It is also well-known that the developmental trajectory for each of the properties is language-specific. For instance, whether falling or rising intonation is acquired first is determined by the language's prosodic typology (e.g. Hallé et al., 1991; Whalen, Levitt, & Wang, 1991). The relative complexity of syllable structures is also reflected in cross-linguistic differences in the number and types of production errors (omission, insertion, or substitution) at different ages, and in the structures that are attempted at earlier ages. Such typological differences pose a greater challenge to integrating segmental and prosodic aspects of utterances (e.g. Vihman, 2014) in the structurally more complex language (cf. Grabe et al., 1999; Grabe et al., 2001; Schmidt & Post, 2015b). We predict that cross-linguistic differences of this type should be reflected in child rhythm as language-specific developmental paths, with more divergent paths for languages that are rhythmically more dissimilar from one another. As is clear from the discussion above, this has also been a common assumption in the child rhythm literature, though the assumption has not been tested.

However, in addition to the structural properties themselves, knowledge of the phonological, morphological, syntactic, pragmatic, discourse and other linguistic structural factors that determine the choice of primitives and their placement in connected speech will need to be acquired by the child. That is, apart from the type of phonological/prosodic language-specific properties mentioned above, the properties of other parts of the language system, as well as the interfaces between them, will determine how, for instance, pitch accent type and accent placement are determined at the sentence level in a specific language (that has them). For instance, Snow (1994) found that children start to control pre-final lengthening after the onset of the multi-word stage (1;5–2;0), but then sometimes show a regression a few months later, which could potentially be due to the child being required to integrate more complex syntactic structure with previous experience of one-word utterances.

Language-specific combinatory restrictions will also constrain how the primitives pertaining to the different parts of the system can be strung together in connected speech (e.g. phonotactics and tonotactics, but also phrase structure rules, devices signaling functions in turn-taking, etc). The child will need to learn how to combine the elements (or how to deconstruct them in perception).

Another implication of our approach is that the child will need to learn the language-specific acoustic/perceptual exponents of the phonological and prosodic properties, and how integrating them in connected speech affects them in terms of their linguistic-phonetic implementation. This is in addition to the many ways in which the durational parameters combine with other phonetic parameters to mark these properties in language-specific ways (the complex "pattern of correspondence" between the "rhythmic structure" of the utterance and its measurable correlates in Allen & Hawkins 1980: 230). Thus, young children reach tonal targets with remarkable precision from a very early age (e.g. by age 2 years in Astruc, Payne, Post, Prieto, & Vanrell, 2013), while the durational marking of heads and edges (accentual and pre-final lengthening) is not necessarily acquired even by age 6.

We tested our prediction that rhythm development as quantified by rhythm metrics can be traced back directly to systemic and componential (including performance) factors (only one parameter – duration – was included in the analyses). The focus was on identifying the likely systemic sources for the common findings in the child rhythm literature (of early even timing in terms of syllable isochrony and less vocalic variability, more *un*even timing in terms of consonantal variability, increasing divergence between languages with a different rhythmic typology from a common developmental point of divergence around 2 years of age, and incomplete acquisition in typologically 'more complex' languages even at age 6).

For monolinguals, the results confirmed that the early bias towards 'syllable-timing' for vocalic duration is primarily rooted in the immature mastery of the complex multi-systemic and multi-componential factors that govern head- and edge-marking in connected speech, both in structural terms, and in the phonetic implementation; but also in a relative lack of durational differentiation between different syllable types. Consonantal variability, by contrast, was found to be directly related to the number of phones and syllables and their complexity (including types of errors within syllables) produced by the children at different ages in the different languages, as predicted by our approach.

For bilinguals, Schmidt and Post (2015b) found a relative advantage in their rate of rhythm development over monolingual peers, especially in the structurally more complex language (English as opposed to Spanish, in this study), which was mirrored in their more adult-like durational marking of heads and edges (i.e. accentual and phrase-final lengthening; Schmidt & Post, 2015b). They suggested that this can perhaps be explained in terms of the bilinguals' dual language input. This exposes them to a greater variety of structures and language-specific phonetic implementations which could result in more developed motor control (Davis, MacNeilage, & Matyear, 2002), and/or in better developed representations for the relevant linguistic structures. This could be due to transfer of structures between the two languages (Paradis, 2001) and the need to consistently differentiate between the two languages, or to an increase in executive function (Bialystok, Luk, Peets, & Yang, 2010).

We conclude that rhythm development can be fruitfully accounted for in an integrated approach in which the emergence of speech rhythm is driven by the closely interrelated development of language-specific and domain-general structural properties and mapping relations, and an increasing knowledge and mastery of linguistic-phonetic and performance-driven implementation. The findings to date suggest a number of future avenues for research which remain to be explored, including the many (neuro)biological and performance components that mould the development of speech rhythm (cf. Payne et al., 2010, 2012).

Acknowledgements

We are very grateful for the many fruitful and inspiring discussions we have had over the years with colleagues about rhythm development which have informed our thinking, in particular with Lluïsa Astruc, Pilar Prieto, and Maria del Mar Vanrell in the context of a series of projects on prosodic development (funded by Batista i Roca grants 2007 PBR 29 and 2009 PBR 00018, P.I. Pilar Prieto; and British Academy grant SG-5177, P.I. Brechtje Post), and with Sarah Hawkins, Francis Nolan, Marilyn Vihman, and Laurence White. This chapter has also benefited greatly from the comments by the editors and two anonymous reviewers.

References

Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh: Edinburgh University Press.

Allen G. D. (1983). Some suprasegmental contours in French two-year-old children's speech. *Phonetica*, 40, 269–292. doi:10.1159/000261696

Allen, G. D., & Hawkins, S. (1978). The development of phonological rhythm. In A. Bell, & J. Hooper (Eds.), *Syllables and segments* (pp. 172–185). Amsterdam: North-Holland Publishing.

Allen, G. D., & Hawkins, S. (1980). Phonological rhythm: Definition and development. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology*. *Volume* 1: *Production* (pp. 227–256). New York, NY: Academic Press.

Armstrong, L. (1932). *The phonetics of French*. London: G. Bell and Sons.

Arvaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics*, 40, 351–373. doi:10.1016/j.wocn.2012.02.003

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. doi:10.1177/0023830912460494

Bertinetto, P. M. (1981). *Strutture prosodiche dell'italiano*. Firenze: Accademia della Crusca.

Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition*, 13(4), 525–531. doi:10.1017/S1366728909990423

Bunta, F., & Ingram, D. (2007). The acquisition of speech rhythm by bilingual Spanish- and English-speaking 4- and 5-year-old children. *Journal of Speech, Language, and Hearing Research*, 50(4), 999–1014. doi:10.1044/1092-4388(2007/070)

Chen, A., & Fikkert, P. (2007). Intonation of early two-word utterances in Dutch. *Proceedings of the 16th International Congress of Phonetic Sciences* – Saarbrücken, 315–320.

Cummins, F. (2015). Rhythm and speech. In M. A. Redford (Ed.), *The handbook of speech production* (pp. 158–177). Chichester: John Wiley and Sons.

Cutler, A. (1976). Phoneme-monitoring reaction time as a function of preceding intonation contour. *Perception and Psychophysics*, 20, 55–60. doi:10.3758/BF03198706

Cutler, A. (1991). Linguistic rhythm and speech segmentation. In J. Sundberg, L. Nord, & R. Carlson (Eds.), *Music, language, speech and brain* (pp. 157–166). London: Macmillan. doi:10.1007/978-1-349-12670-5_14

Cutler, A., & Foss, D. J. (1977). On the role of sentence stress in sentence processing. *Language and Speech*, 21, 1–10

Dasher, R., & Bolinger, D. (1982). On pre-accentual lengthening. *Journal of the International Phonetic Association*, 12, 58–71. doi:10.1017/S0025100300002462

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, 11, 51–62.

Dauer, R. M. (1987). Phonetic and phonological components of language rhythm. *Proceedings of the 11th ICPhS* – Tallinn, 447–450.

Davis, B. L., MacNeilage, P. F., & Matyear, C. L. (2002). Acquisition of serial complexity in speech production: A comparison of phonetic and phonological approaches to first word production. *Phonetica*, 59, 75–109. doi:10.1159/000066065

Delattre, P. (1966). A comparison of syllable length conditioning among languages. *International Review of Applied Linguistics in Language Teaching*, 4, 183–198. doi:10.1515/iral.1966.4.1-4.183

Delattre, P. (1969). An acoustic and articulatory study of vowel reduction in four languages. *International Review of Applied Linguistics in Language Teaching*, 7, 295–325. doi:10.1515/iral.1969.7.4.295

DePaolis, R. A., Vihman, M. M., & Kunnari, S. (2008). Prosody in production at the onset of word use: A cross-linguistic study. *Journal of Phonetics*, 36, 406–426. doi:10.1016/j.wocn.2008.01.003

Echols, C., Crowhurst, M., & Childers, C. (1997). Perception of rhythmic units in speech by infants and adults, *Journal of Memory and Language*, 36, 202–225. doi:10.1006/jmla.1996.2483

Grabe, E., Gut, U., Post, B., & Watson, I. (2001). The acquisition of rhythm in English, French and German. *Proceedings of the Child Language Seminar 1999* – London, 156–162.

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven, & N. Warner (Eds.), *Papers in Laboratory Phonology 7* (pp. 515–546). Cambridge: Cambridge University Press.

Grabe, E., Watson, I., & Post, B. (1999). The acquisition of rhythmic patterns in English and French. *Proceedings of the 14th International Congress of Phonetic Sciences* – San Francisco, 1201–1204.

Hallé, P., de Boysson-Bardies, B., & Vihman, M. (1991). Beginnings of prosodic organisation: Intonation and duration patterns of disyllables produced by Japanese and French infants. *Language and Speech*, 34(4), 299–318.

Hawkins, S. (2014). Situational influences on rhythmicity in speech, music, and their interaction. *Philosophical Transactions of the Royal Society*, 369, 20130398. doi:10.1098/rstb.2013.0398

Hawkins, S., Cross I., & Ogden, R. (2013). Communicative interaction in spontaneous music and speech. In M. Orwin, C. Howes, & R. Kempson (Eds.), *Language, music and interaction*, (pp. 285–329). London: College Publications.

Huggins, A. W. F. (1972). On the perception of temporal phenomena in speech. *Journal of the Acoustical Society of America*, 51, 1279–1290. doi:10.1121/1.1912972

Johnson, E., & Jusczyk, P. (2001). Word segmentation by 8-month-olds: When speech cues count for more than statistics. *Journal of Memory and Language*, 44, 548–567. doi:10.1006/jmla.2000.2755

Jusczyk, P. (1997). *The discovery of spoken language*. Cambridge, MA: The MIT Press.

Jusczyk, P. (1999). How infants begin to extract words from fluent speech. *Trends in Cognitive Science*, 3, 323–328. doi:10.1016/S1364-6613(99)01363-7

Jusczyk, P., Houston, D., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, 39, 159–207. doi:10.1006/cogp.1999.0716

Kehoe, M., Lleó, C., & Rakow, M. (2011). Speech rhythm in the pronunciation of German and Spanish monolingual and German-Spanish bilingual 3-year-olds. *Linguistische Berichte*, 227, 323–351.

Konopczynski, G. (1995). A developmental model of acquisition of rhythmic patterns: Results from a cross-linguistic study. *Proceedings of the Intonational Congress of Phonetic Sciences* – Stockholm, Vol. 4, 22–25.

Ladefoged, P. (1975). *A course in phonetics* (1st. ed.). New York, NY: Harcourt Brace Jovanovich.

Li, A., & Post, B. (2014). L2 acquisition of prosodic properties of speech rhythm: Evidence from L1 Mandarin and German learners of English. *Studies in Second Language Acquisition*, 36(2), 223–255. doi:10.1017/S0272263113000752

Lleó, C., Rakow, M., & Kehoe, M. (2007). Acquiring rhythmically different languages in a bilingual context. *Proceedings of the 16th International Congress of Phonetic Sciences* – Saarbrücken, 1545–1548.

Lloyd James, A. (1940). *Speech signals in telephony*. London: Pitman and Son.

Low, E. L., Grabe, E., & Nolan F. (2000). Quantitative characterisations of speech rhythm: "Syllable-timing" in Singapore English. *Language and Speech*, 43(1), 377–401.

Mok, P. (2011). The acquisition of speech rhythm by three-year-old bilingual and monolingual children: Cantonese and English. *Bilingualism: Language and Cognition*, 14(4), 458–472. doi:10.1017/S1366728910000453

Mok, P. (2013). Speech rhythm of monolingual and bilingual children at 2;06: Cantonese and English. *Bilingualism: Language and Cognition*, 16, 693–703. doi:10.1017/S1366728912000636

Morgan. J. L. (1986). *From simple input to complex grammar*. Cambridge, MA: The MIT Press.

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

Nazzi, T., Jusczyk, P. W., & Johnson, E. K. (2000). Language discrimination by English learning 5-month-olds: Effects of rhythm and familiarity. *Journal of Memory and Language*, 43, 1–19. doi:10.1006/jmla.2000.2698

Nazzi, T., & Ramus, F. (2003). Perception and acquisition of linguistic rhythm by infants. *Speech Communication*, 41, 233–243. doi:10.1016/S0167-6393(02)00106-1

Nolan, F., & Jeon, H.-S. (2014). Speech rhythm: A metaphor? *Philosophical Transactions of the Royal Society B*, 369, 20130396. doi:10.1098/rstb.2013.0396

Ordin, M., & Polyanskaya, L. (2014). Development of timing patterns in first and second languages. *System*, 42, 244–257. doi:10.1016/j.system.2013.12.004
Paradis, J. (2001). Do bilingual two-year-olds have separate phonological systems. *International Journal of Bilingualism*, 5(1), 19–38. doi:10.1177/13670069010050010201
Payne, E., Post, B., Astruc, L., Prieto, P., & Vanrell, M. M. (2010). Rhythmic modification in child directed speech. In M. Russo (Ed.), *Prosodic Universals. Comparative Studies in Rhythmic Modeling and Rhythm Typology* (pp. 147–184). Roma: Aracne Editrice.
Payne, E., Post, B., Prieto, P., Vanrell, M. M., & Astruc, L. (2012). Measuring child rhythm. *Language and Speech*, 55, 202–228. doi:10.1177/0023830911417687
Phillips-Silver, J., & Trainor, L. J. (2005). Feeling the beat: Movement influences infant rhythm perception. *Science*, 308, 1483. doi:10.1126/science.1110922
Pike, J. L. (1945). *The intonation of American English*. Ann Arbor, MI: University of Michigan Press.
Prieto, P., & Vanrell, M. M. (2009). L'adquisició de l'entonació del català. In Devís, E., & Carol, L. (Eds.), *Studi catalani. Suoni e parole* (pp. 67–82). Bologna: Bononia University Press.
Prieto, P., Vanrell, M. M., Astruc, L., Payne, E., & Post, B. (2012). Phonotactic and phrasal properties of speech rhythm. Evidence from Catalan, English, and Spanish. *Speech Communication*, 54(6), 681–702. doi:10.1016/j.specom.2011.12.001
Ramus, F. (2002). Language discrimination by newborns: Teasing apart phonotactic, rhythmic, and intonational cues. *Annual Review of Language Acquisition*, 2, 85–115. doi:10.1075/arla.2.05ram
Ramus, F., Dupoux, E., & Mehler, J. (2003). The psychological reality of rhythm classes: Perceptual studies. *Proceedings of the 15th International Congress of Phonetic Sciences* – Barcelona, 337–342.
Ramus, F., Hauser, M., Miller, C., Morris, D., & Mehler, J. (2000). Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science*, 288(5464), 349–351. doi:10.1126/science.288.5464.349
Ramus, F., & Mehler, J. (1999). Language identification with suprasegmental cues: A study based on speech resynthesis, *Journal of the Acoustical Society of America*, 105(1), 512–521. doi:10.1121/1.424522
Ramus, F., & Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal, *Cognition*, 73, 265–292. doi:10.1016/S0010-0277(99)00058-X
Rathcke, T., & Smith, R. (2015). Speech timing and linguistic rhythm: On the acoustic bases of rhythm typologies. *The Journal of the Acoustical Society of America*, 137, 2834–2845. doi:10.1121/1.4919322
Roach, P. (1982). On the distinction between 'stress-timed' and 'syllable-timed' languages. In D. Crystal (Ed.), *Linguistic controversies* (pp. 73–79). London: Edward Arnold.
Schmidt, A-E., & Post, B. (2015a). Language interaction in the development of speech rhythm in simultaneous bilinguals. In E. Delais-Roussarie, M. Avanzi, & S. Herment (Eds.), *Prosody and language in contact* (pp. 271–291). Heidelberg: Springer.
Schmidt, A-E., & Post, B. (2015b). The development of prosodic features and their contribution to rhythm production in simultaneous bilinguals. *Language and Speech*, 58(1), 24–47. doi:10.1177/0023830914565809
Shields, J. L., McHugh, A., & Martin, J. G. (1974). Reaction time to phoneme targets as a function of rhythmic cues in continuous speech. *Journal of Experimental Psychology*, 102(2), 250–255. doi:10.1037/h0035855

Snow, D. (1994). Phrase-final syllable lengthening and intonation in early child speech. *Journal of Speech and Hearing Research*, 37, 831–840. doi:10.1044/jshr.3704.831

Vihman, M. M. (2014). *Phonological development: The origins of language in the child*. Oxford: Blackwell.

Vihman, M. M., Davis, B. L., & DePaolis, R. A. (1995). Prosodic analysis of babbling and first words: A comparison of English and French. *Proceedings of the XIIIth International Congress of Phonetic Sciences* – Stockholm, Vol. 4, 14–27.

Vihman, M. M., Nakai, S., & DePaolis, R. A. (2006). Getting the rhythm right: A cross-linguistic study of segmental duration in babbling and first words. In L. Goldstein, D. Whalen, & C. Best (Eds.), *Laboratory phonology 8: Phonology and phonetics* (pp. 341–366). Berlin: Mouton de Gruyter.

Whalen, D. H., Levitt, A. G., & Wang, Q. (1991). Intonational differences between the reduplicative babbling of French-and English-learning infants. *Journal of Child Language*, 18(3), 501–516. doi:10.1017/S0305000900011223

White, L. (2014). Communicative function and prosodic form in speech timing. *Speech Communication*, 63–64, 38–54. doi:10.1016/j.specom.2014.04.003

White, L., Delle Luche, C., & Floccia, C. (2016). Five-month-old infants' discrimination of unfamiliar languages does not accord with "rhythm class". *Proceedings of Speech Prosody 2016* – Boston, 567–571.

White, L., Floccia, C., Goslin, J., & Butler, J. (2014). Utterance-final lengthening is predictive of infants' discrimination of English accents. *Language Learning*, 64, 27–44. doi:10.1111/lang.12060

White, L., Mattys, S. L., & Wiget, L. (2012). Language categorization by adults is based on sensitivity to durational cues, not rhythm class. *Journal of Memory and Language*, 66, 665–679. doi:10.1016/j.jml.2011.12.010

Whitworth, N. (2002). Speech rhythm production in three German-English bilingual families. *Leeds Working Papers in Linguistics and Phonetics*, 9, 175–205.

CHAPTER 8

Early development of intonation

Perception and production

Sónia Frota & Joseph Butler
Universidade de Lisboa

This chapter focuses on early development of intonation. Together with a precocious sensitivity to prosody documented in the literature, recent research has shown that infants' early perception of pitch-based categories is already language-specific by 4–5 months, and that their discrimination abilities differ not only according to ambient language but also as a function of pitch properties (e.g., pitch direction, or pitch alignment). On the production side, and focusing on studies within the Autosegmental-Metrical framework, findings suggest that key landmarks in intonational development precede and constrain the acquisition of other aspects of grammar (e.g., word and phrase size, and combinatorial speech). Both from a perception and production point of view, language specific effects emerge very early on in development, underlying cross-linguistic differences.

Introduction

A precocious sensitivity to prosody is by now a well-established fact in early language development. Newborn infants have been shown to be sensitive to the global rhythmic properties that distinguish between languages (Nazzi, Bertoncini, & Mehler, 1998), to word-level rhythm, that is, to different patterns of lexical stress (Sansavini, Bertoncini, & Giovanelli, 1997), to the contrast between function words, which are unstressed, and content words, which bear stress (Shi, Werker, & Mogan, 1999), to the prosodic cues, pitch included, that signal boundaries between words (Christophe, Dupoux, Bertoncini, & Mehler, 1994), and to the differences in pitch contours that result from contrasting lexical pitch accents (Nazzi, Floccia, & Bertoncini, 1998). Furthermore, newborns were found to rely more on vowels than on consonants when recognizing words, unlike adults and toddlers (Benavides-Varela, Hochmann, Macagno, Nespor, & Mehler, 2012). This is consistent with a precocious sensitivity to prosody, given that vowels are the main

DOI 10.1075/tilar.23.08fro

carriers of prosody. Although these studies suggest that newborns' sensitivity to prosody may not depend on language experience, they leave open the possibility that this early sensitivity may evolve differently according to the language-specific prosodic features of the ambient language. In this chapter, we focus on early development of intonation drawing on recent findings for several languages from perception and production in infants and young children.

Studies into the early perception of prosody have mostly examined the acquisition of lexical prosodic contrasts, namely word stress and lexical pitch (on word stress, Bhatara, Boll-Avetisyan, Hohle, & Nazzi, this volume). For lexical pitch, it has been shown that learners of tone languages (e.g., Mandarin), unlike learners of non-tone languages (e.g., English), show stable discrimination of distinct lexical tones throughout the first year of life, and as early as 4–5 months (Mattock & Burnham, 2006; Mattock, Molnar, Polka, & Burnham, 2008; Yeung, Chen, & Werker, 2013). Similarly, learners of pitch accent languages (e.g., Japanese) are able to discriminate native pitch accent contrasts from as early as 4 months (Sato, Sogabe, & Mazuka, 2009). Interestingly, phonetic variability seems to hinder the discrimination of native lexical tones (Shi, 2010, using variable tone contexts), whereas the perception of native pitch accent is not affected (Sato et al., 2009, using segmental variation). Much less is known about how the perception of intonation develops, and whether parallels may be found with lexical pitch, or lexical prosody in general. This topic only very recently has gained researchers' attention. The second section of the present chapter is devoted to early perception of intonation and its developmental path. We review the recent growing body of literature, and discuss the findings as evidence for general perceptual abilities for pitch-based contrasts and/or for the early role played by language experience.

Unlike perception, studies on intonational development from the viewpoint of production have been around for nearly three decades. Most of these studies followed a contour-based approach to the development of intonation, focusing on overall shapes and acoustic measures of pitch over the whole utterance or its terminal contour (e.g., DePaolis, Vihman, & Kunnari, 2008; Esteve-Gibert & Prieto, 2013; Hallé, Boysson-Bardies, & Vihman, 1991; Frota & Vigário, 1995; Snow, 1994, 2002, 2006; see Snow & Balog, 2002, for a review). They have revealed cross-linguistic differences in the type of contours (falling or rising) first produced by the child, consistent with communicative intentions and intonational differences in the ambient languages. However, the contour-based approach is not very informative about the structural properties of intonation, such as the types of nuclear pitch accent and boundary tone, the distribution of pitch events, or the ways in which the tune aligns with the text. These properties are known to be crucial to the description of any intonational system (Gussenhoven, 2007; Ladd, 2008), and it is thus critical to understand how they are acquired. In the Autosegmental-Metrical

(AM) framework, currently the most widely used approach to intonation (e.g., Frota & Prieto, 2015a; Jun, 2014a), these properties are at the core of intonational description and analysis. Studies of early intonation within the AM framework are however rare. In the third section of this chapter, we focus on these studies, review the results reported, and discuss similarities and differences across languages in the path of intonational development and its relation to other aspects of grammatical development.

The AM framework allows for the distinction between a phonological and a phonetic level in intonation, and the comparison across languages at both levels. Within this model, the intonation contour is formed by a string of high and low tonal events that associate with lexically stressed syllables or with prosodic phrase edges. The former are known as pitch accents and the latter as boundary tones, and either can be single-tone (high or low, e.g., H* or L% respectively for a pitch accent and a boundary tone) or some combination of high and low (e.g., H + L* or LH%). Recent work on early child intonation has shown that this model is adequate to investigate the development of intonation in perception and production. Therefore, our review mostly focuses on work that has used the AM approach to understand the developmental path of intonation in infancy and early childhood.

Given the only recent research focus on early development of intonation, it is not surprising that our understanding of how intonation is acquired, based on evidence from both perception and production, and using a full-fledged phonological framework such as the AM model, is still fragmentary. Understanding the developmental path of intonation towards a full adult-like competence will impact on our knowledge of the acquisition of other linguistic categories and structures, as intonation might help the infant and young child to bootstrap aspects of the grammatical organization of the native language (Höhle, 2009; Jusczyk, 1997). It will also contribute towards tracing the development of communication and social interaction, and potentially signal language impairments (Filipe, Branco, Frota, Castro, & Vicente, 2014; Koegel, Koegel, Green-Hopkins, & Barnes, 2010). Both old and new methods and techniques to examine language processing in infants and toddlers can be explored to further study early intonation. These potential avenues for research, and their implications, are discussed in the final section of this chapter.

Early perception of intonation

It has already been established that infants show a preference for rising contours, high pitch and expanded melodies from birth, and that they are sensitive to changes in melodic contours (e.g. Papoušek, Bornstein, Nuzzo, Papoušek, & Symmes, 1990; Trehub, Bull, & Thorpe, 1984), however grammatically relevant intonation contrasts have only recently begun to be examined. Intonation is a prosodic dimension that varies across languages and impacts upon meaning, similarly to lexical stress, lexical tone and lexical pitch accent, but, differently, intonation conveys phrasal meanings, such as sentence type distinctions and pragmatic meaning, rather than word level meanings.

Perception of native intonation

Of the few studies that have investigated early intonation perception, most have focused on perception relating to sentence type distinctions. Best, Levitt, and McRoberts (1991) were the first to address this issue. They asked whether the perception of intonation followed the general developmental change reported for the acquisition of segments and lexical tone, known as perceptual narrowing or perceptual attunement, whereby discrimination of non-native contrasts declines in the second half of the infants' first year of life giving way to discrimination abilities attuned to the native language. To this end, they examined both native and non-native discrimination of exclamations and wh-questions by English-learning infants. They found that at 6–8 months infants were able to discriminate both native and non-native (Spanish) contrasts, however at 10–12 months they failed to discriminate either. Although these findings appear to contradict the view that discrimination abilities are refined through native language exposure, their stimuli contained both word order (inversion in wh-questions) and lexical properties (presence of wh-word), in addition to (highly variable) prosodic cues, which may have impacted on infants' discrimination performance. Similarly, Geffen and Mintz (2011) also investigated discrimination of the statement/yes-no question contrast, using stimuli that included word order cues as well as intonation cues, utilizing a modified version of the Headturn Preference Procedure (HPP). Intonation cues were described as larger mean pitch changes for questions. They found that English-learning infants at 7 months successfully discriminated this contrast, but, as in Best et al.'s (1991) study, it is not clear whether infants attended to intonation cues or the lexical/word order cues also present. A recent study by Geffen and Mintz (2015), again using the modified version of the HPP, found that English-learning infants at 12 months were able to distinguish between declarative and interrogative sentence types when prosodic cues were removed, leaving only the lexical distribution patterns, thus suggesting a sensitivity to word order cues.

Two studies have also investigated sentence type distinctions, using stimuli that differ only in intonation cues. Soderstrom, Ko, and Nevzorova (2011), using a Visual Habituation paradigm, utilized multiword, uninverted yes-no question and statement sentences, therefore neutralizing the word order cue to questions in English. Overall, they found a preference for questions but no direct evidence of discrimination. These findings could possibly indicate a general preference for high/rising pitch (Papoušek et al., 1990; Trehub et al., 1984). However, the large age range used in this study (4.5 to 24 months) makes it difficult to draw conclusions about the developmental trajectory of perceptual abilities from this data. Additionally, the intonation used within each sentence type in their stimuli was very variable (e.g., declarative sentences ending with either flat, falling or bell shaped contours) and may have impacted on infants' perception of the contours. Frota, Butler, and Vigário (2014) used single prosodic word utterances differing in prosodic cues related to the statement/yes-no question contrast in a language, European Portuguese (EP), where this contrast is marked by prosodic cues (differently to English, in EP the yes-no question counterpart of *Ele gosta da Maria* 'He likes Mary' is *Ele gosta da Maria?* with the same word order as the statement, but not *Gosta ele da Maria?*, 'Does he like Mary?'). Using a similar Visual Habituation paradigm, they found that, at both 5–6 and 8–9 months, EP-learning infants were able to discriminate the prosodic contrast, despite segmental variability. In contrast with Soderstrom et al.'s (2011) study, Frota et al. (2014) used intonation contours consistently; a falling pattern for the declarative and a falling-rising pattern for the interrogative. To our knowledge, this was the first study to demonstrate successful infant discrimination of an intonation contrast, suggesting precocious discrimination abilities for intonation in the presence of segmental variability similar to those reported for lexical pitch accent.

Following on from this study, Butler, Vigário, and Frota (2016) investigated a further prosodic contrast in EP that cues the difference between all-new information (broad focus) and the highlighting of a particular word (narrow/contrastive focus). This distinction is primarily marked in declarative utterances by the timing of the pitch fall, which is timed early with respect to the stressed syllable of the nuclear word in broad focus and late in narrow/contrastive focus. Differently to previous studies into intonation contrasts, which have mostly focused on pitch height/direction contrasts, this study addressed a pitch timing distinction. Using the Visual Habituation paradigm, Butler et al. (2016) found that EP learning infants discriminated this contrast at 12 months, but not at 7 months, deviating from previous findings of a precocious ability to perceive pitch distinctions. These findings suggest that there may be different developmental trajectories of the perception of different prosodic contrasts (namely, pitch height/direction and pitch timing), and underlines the importance of the nature of the prosodic cues signaling a given contrast in a given language.

Unlike other studies, Frota et al. (2014) and Butler et al. (2016) provide a description of the contours that characterize the materials used, following the AM model. Given the intonational account of the contrasts, we can describe the pitch height/direction sentence type distinction as manifested by the pitch event at the right edge of the utterance, that is as a boundary tone contrast where L% in statements and LH% in questions follow the falling pitch in the nuclear word, as shown in Figure 1. The focus-related pitch timing contrast, on the other hand, is manifested by the way the peak aligns with the text in the nuclear pitch fall. The temporal coordination of the pitch fall with the stressed syllable yields a peak just before this syllable in the broad focus case (H + L*), whereas in the narrow/contrastive focus case the peak occurs on the stressed syllable (Figure 2). In either case the melody at the right edge of the utterance is low.

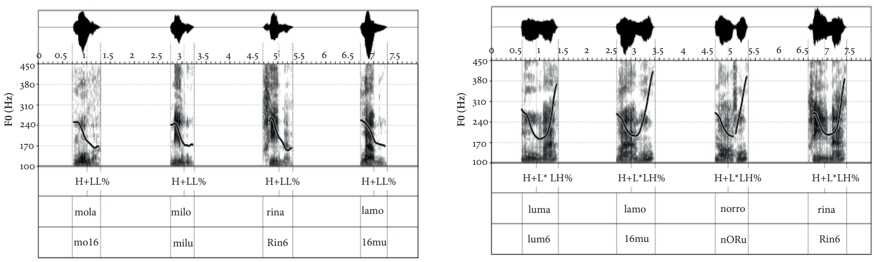


Figure 1. Intonation contours of illustrative statement (left panel) and question (right panel) stimuli, with prosodic labelling and orthographic and phonetic transcription (in SAMPA), from Frota et al. (2014).

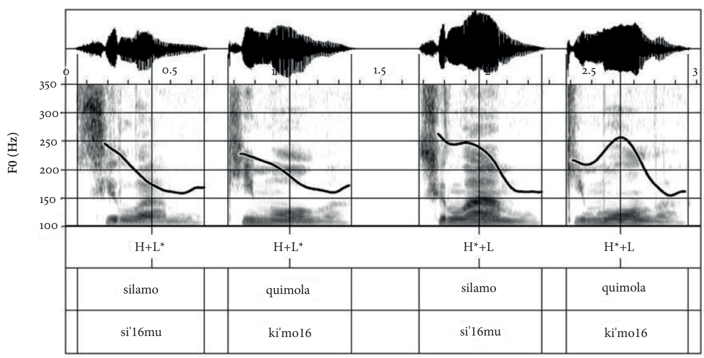


Figure 2. Intonation contours of broad (left two examples) and narrow (right two examples) focus, with prosodic labelling and orthographic and phonetic transcription (in SAMPA), from Butler et al. (2016).

If different prosodic contrasts have different developmental trajectories, as the perception findings suggest, this has potential implications for the acquisition of linguistic categories marked with different prosodic features. The perception/

discrimination of a given contrast is a pre-requisite, which allows the acquisition of the grammar and pragmatics that the contrast conveys in a given language. In the case of EP, an advantage of pitch direction over pitch timing may suggest an advantage for the acquisition of the sentence type distinction over the focus distinction. In a language like Neapolitan Italian, however, the statement/question distinction is cued by pitch timing (the pitch rise is timed later in questions than in statements, D'Imperio, 2002). The potential impact for language acquisition of the presence or absence of early discrimination abilities related to a specific intonation contrast that cues a given linguistic distinction, within and across languages, remains to be investigated.

The precocious discrimination abilities for lexical pitch accent and some types of intonation contrasts reported in the literature are in line with infants' sensitivity to, and ability to utilise, pitch from a very early age. For example, infants are able to use pitch, and other prosodic cues, to infer and express affective and pragmatic meaning (Esteve-Gibert & Prieto, this volume). Infants have also been shown to be able to use differences in pitch level or pitch shape as a cue to segment clauses and as a cue to prosodic phrase boundaries in order to guide word segmentation (de Carvalho, Dautriche, Milotte, & Christophe, this volume; Seidl, 2007; Shukla, White, & Aslin, 2011). German 9-month-old infants have been shown to utilise pitch in order to aid segmentation, demonstrating segmentation only when high pitch aligns with the stressed syllable, and not when high pitch precedes or follows the stressed syllable (Zahner, Schönhuber, & Braun, 2015). Although infants do not necessarily always use pitch cues in an adult-like fashion, they are sensitive to, and able to use linguistically, pitch cues very early in development.

Perception of Non-native Intonation

Recent work has begun to focus on non-native perception of intonation, to attempt to shed further light on developing perceptual abilities for intonation. Sundara, Molnar, and Frota (2015) tested English- and Basque-learning infants' ability to discriminate the statement/yes-no question contrast, using the stimuli prepared by Frota et al. (2014) and the same type of visual habituation procedure. The reason for using the EP stimuli with English- learning infants was twofold. Firstly, using stimuli that differed only in prosodic cues would allow further explanation for previous findings of a failure to discriminate statement vs. question contrast by English-learning infants (see Soderstrom et al., 2011, and Section "Perception of Native Intonation" above). Secondly, the authors wanted to confirm that EP learning infants' ability to discriminate the pitch contrast in this sentence type distinction, in the face of segmental variability, was the result of their language experience. Furthermore, Basque-learning infants were also tested, because in Basque, unlike in English, yes-no questions and statements can be segmentally identical and differ only in their intonation, like in EP. In Basque, statements typically end in a fall (L%), but only questions have a high pitch immediately before the fall (HL% – Elordieta & Hualde, 2014). Sundara et al. (2015) found that English-learning infants at 4 months failed to discriminate the statement/question contrast using EP stimuli with segmental variability, however Basque-learning infants succeeded in this task. Importantly, English-learning infants also failed when the intonation contrast was presented without segmental variability, using the same phonetic sequence with a final fall or rise (i.e., the same bisyllabic word with either H + L* L% or H + L* LH%). The English-learning infants only succeeded when absence of segmental variability was combined with a more stringent habituation criterion. These findings suggest that English-learning infants have a limited ability to differentiate boundary tones, unlike EP and Basque-learning infants, and that intonation perception is already language-specific by 4–5 months.

Frota, Butler, Lu, and Vigário (2016) also investigated non-native intonation perception in EP-learning infants, again using the visual habituation paradigm. In this study, the authors utilised a segmentally varied non-native contrast, i.e., the distinction between Mandarin Chinese Tone 1 + Tone 4 and Tone 1 + Tone 2. This contrast provides overall similar contour shapes to the statement/yes-no question prosodic contrast in EP, although with a different implementation of the falling/rising patterns throughout the segmental string. The similarity to the EP native contours resides in the final pitch height and direction features, with Tone 1 + Tone 4 showing a final fall and Tone 1 + Tone 2 a final rise preceded by a fall, like in EP statements and questions, respectively. The differences from the EP contours arise from the difference in nature of tune versus tone. In EP, the falling pitch of the nuclear pitch accent (H + L*) in statements and questions is already in the first syllable, whereas in Mandarin the first syllable always has high pitch due to Tone 1. Additionally, in EP the contrast between the two contours is manifested by a low (L%) and rising pitch (LH%) in the second syllable; in Mandarin the second syllable bears falling pitch due to Tone 4 or a falling-rising contour due to the transition between Tone 1 and Tone 2. In other words, the melody in EP spreads over the nuclear word, whereas in Mandarin the domain of the contour is the syllable. In contrast with the findings for the native contrast (Frota et al., 2014), EP-learning infants at both 5–6 months and 8–9 months failed to discriminate this non-native contrast. These findings provide further evidence for language specific pitch perception, suggesting that intonation and lexical tone are already differentiated at 5–6 months.

In sum, work so far into early intonation suggests that there are different developmental trajectories for different types of intonation contrasts. The statement/yes-no question distinction characterized by pitch height/direction acoustic cues appears to be more salient, and thus perceived earlier, than the broad/narrow focus contrast that is defined by a pitch timing cue. These findings highlight the importance of the nature of the cues that signal a given contrast, and may be modulated by language experience. Recent work into non-native perception of intonation suggests that infants' perception is indeed influenced by language experience, and that language-specific perception for pitch, and for the tone/intonation distinction, emerges as early as 4–5 months of age. The findings related to the perception of intonation patterns in the presence of segmental variability show an influence of language experience earlier in development than for the perception of vowels, consonants, or even lexical tone.

Further work is required in order to better understand how the developmental trajectories of different intonation contrasts proceed. To date, only two types of intonation contrasts have been studied, and more work is needed to examine these and other intonation contrasts conveying phrasal meaning distinctions in other languages. Additionally, more work is needed to better define the aspects of infants' particular language experience behind precocious sensitivity to pitch differences across languages.

Emerging intonation in production

Previous studies on intonational development from the viewpoint of production have followed a contour-based approach, describing the overall shapes (e.g., falling, rising, level) of the contours first produced by the child (see Snow & Balog, 2002, for a review). Studies of early intonation within the AM framework are recent. These studies address the structural properties of the melodies produced, and ask when and how young children acquire the intonational system of their ambient language. The acquisition of the intonational system involves mastering the phonology and phonetics of intonation: the former relates to the phonological inventory of pitch accents and boundary tones, and their respective meanings in language grammar and use; the latter concerns the realization of tonal patterns or their phonetic implementation, namely the alignment of pitch events with the text as well as their scaling properties. Both the phonology and the phonetics of intonation may vary in language-specific ways. In some languages, for example, alignment and/or scaling features are used phonologically as part of the system of intonation contrasts that contribute to meaning (as in the case of the focus-related pitch timing contrast in EP, described in Section "Perception of native intonation").

In this section, we review the results reported in studies of the acquisition of the intonational system, and discuss similarities and differences across languages in the path of early intonational development and its relation to other aspects of grammatical development.

Acquiring the phonological inventory of tonal events

The use of the same framework of intonational analysis in various studies and for different languages allows for cross-linguistic comparison of early intonation patterns in relation to the adult intonation system. Results from several languages show that the phonological inventory of nuclear pitch accents and boundary tones is acquired early in development.

Chen and Fikkert (2007), using spontaneous speech data of three Dutch-learning children between 16 and 25 months, report that the full set of Dutch nuclear pitch accents and boundary tones is mastered by 24 months, but not before the onset of the two-word stage. Although intonational development seems to be evident relatively early in Dutch, findings for Catalan, European Portuguese and Spanish show an even earlier development of intonation. Studies on these languages, also using longitudinal corpora of spontaneous speech, all reported that children display an adult-like use of the distinct nuclear contours of their native language from the onset of speech, and appropriately use the different tunes for specific pragmatic meanings (Prieto & Vanrell, 2007, and Prieto, Estrella, Thorson, & Vanrell, 2012, for Spanish and Catalan; Frota & Vigário, 2008, and Frota, Cruz, Matos, & Vigário, 2016, for European Portuguese). For example, the same segmental word can be produced as a one-word utterance expressing a request, an order, or a question, depending on the tune used (Figure 3). Early intonational development in these languages was thus found to be unrelated to the onset of combinatorial speech.

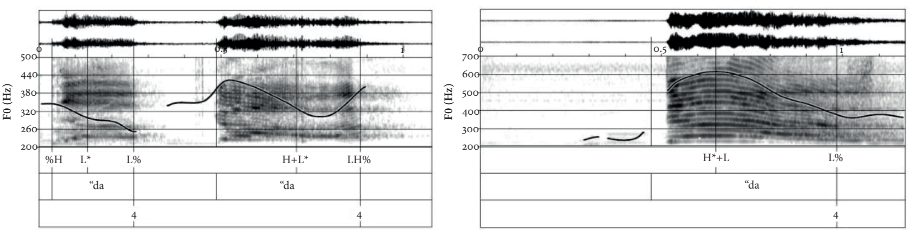


Figure 3. Intonation contours of the utterance *Dá* ('Give') produced by an EP-learning child at 1;05, as a request and a question, in the left panel, and as a command, in the right panel, showing a target-like use of nuclear contours (from Frota & Vigário, 2008).

Yes-no questions have been shown to be among the sentence types that display a high degree of variation in nuclear contours, across languages and language varieties (e.g., Frota & Prieto, 2015b; Savino, 2012). In addition to cross-linguistic variation in the intonation of yes-no questions that convey a similar pragmatic meaning, yes-no questions also display a wealth of pragmatic meanings that may be encoded intonationally in language-specific ways (such as information-seeking, confirmation-seeking, or echo questions). In a study on the development of interrogative intonation in Catalan and Spanish-learning children between 17 and 28 months, Thorson, Borras-Comes, Crespo-Sendra, Vanrell and Prieto (2014) found that children before the two-word stage produce several types of interrogatives, and do so already reflecting the adult inventory patterns. Furthermore, there is a relationship between the type of nuclear pitch configuration produced and the pragmatic meaning of the question that agrees with the adult-like form-meaning relationships. These findings are in line with cross-linguistic evidence on the early production of the language-specific phonological inventory of nuclear pitch accents and boundary tones, and their respective meanings. They also confirm that this early development of the intonational system of the ambient language occurs regardless of the onset of combinatorial speech.

Further steps in the development of the intonational system are related with the use of prenuclear pitch accents, that is, the pitch accents that precede the nuclear contours in two-word and multiword utterances. Again, languages may differ in the types and distribution of pre-nuclear pitch accents (Frota & Prieto, 2015b; Jun, 2014b). In Dutch-learning children, the use of prenuclear pitch accents is not yet adult-like by the age of two, showing a later acquisition than nuclear pitch accents and boundary tones (Chen & Fikkert, 2007). Similar findings are reported for German and Spanish monolinguals, who show adult-like patterns by 36 months (Lleó, 2016; Lleó, Rakow, & Kehoe, 2004).

The placement of pitch accents, including the nuclear pitch accent, is a further aspect that needs to be acquired. In some languages, like English or Dutch, pitch accent placement is strongly governed by semantics and information status, and thus words may show deaccenting (or absence of pitch accent) depending on their semantic properties and on discourse context (Ladd, 2008). That is not the case in European Portuguese, for example, a language characterized by a sparse pitch accent distribution, with very few prenuclear accents, and where nuclear accent placement is final in broad focus utterances (Frota, 2014). The use of pitch accent placement by Dutch-learning two-year olds is not yet target-like, with children typically pitch-accenting both words in two-word utterances (Chen & Fikkert, 2007). In European Portuguese, a language with different pitch accent placement rules from Dutch, a tendency to pitch accent every word was also found before 24 months (Frota, Cruz, Matos, & Vigário, 2016). This tendency was interpreted as a stage in the development of prosodic phrasing, where each word forms a phrase, as also shown by other prosodic cues such as pitch reset and pause distribution. Words become integrated within the same prosodic phrase just before 24 months, and in the case of the EP-learning children studied, well before the onset of combinatorial speech. The findings from Dutch and European Portuguese may suggest a common cross-linguistic pattern in the development of prosodic phrasing, which constrains pitch accent distribution independently of the language-specific rules that govern accent placement.

By and large, the findings available point to a precocious development of the intonational system of the ambient language (albeit with differences across languages): the inventory of nuclear pitch accents and boundary tones is adult-like close to the beginning of speech, together with the form-meaning relationships of the distinct tunes, and further developments in the use of prenuclear accents and pitch accent distribution evolve quickly and seem to be in place by the end of the second year of life. Notably, there is convergent cross-linguistic evidence that emerging intonation and its early development are independent of the onset of combinatorial speech. In Section "Early perception of intonation", it was shown that infants are sensitive to aspects of the intonation of their ambient language from 4–5 months of age, and that they are able to linguistically use pitch cues from 6 months, for example to guide word segmentation. Moreover, infants' perception of a given intonation contrast is a pre-requisite for the acquisition of the grammar and pragmatics it conveys. Thus, findings on infants' perception revealed abilities of prosodic analysis of the speech signal that might help the infant to acquire the lexicon and to bootstrap aspects of the grammatical organization of the native language. It is therefore not surprising that key aspects of the intonational system develop so early in child speech. Like for perception, the findings from production are compatible with the hypothesis that prosody precedes and might promote (further) grammatical development.

Acquiring the phonetics of intonation

The early mastering of crucial aspects of the phonology of intonation does not necessarily entail that their phonetics is also acquired. Indeed, results from different languages support the view that the phonological patterns are in place before the phonetics of tonal alignment and tonal scaling become target-like. Differences across languages in the acquisition of the phonetics of intonation have also been reported, suggesting that the discrepancies between adult and early child language are not simply due to neuromotor constraints on early speech production (DePaolis et al., 2008).

Chen and Fikkert (2007) reported differences in tonal alignment of falling pitch accents (namely, H*L) between early child Dutch (from 16 to 25 months) and adult Dutch. Similarly, findings from European Portuguese show a tendency towards late peak alignment in falling nuclear accents in the beginnings of child speech, which is especially evident in the case of the H + L* pitch accent (Frota, Cruz, Matos, &

Vigário, 2016). In EP, tonal alignment approached the adult-like pattern around 21 months. The fine-grained realization of alignment patterns in this language is important given the phonological use of tonal alignment in the H + L*/H* + L pitch accent contrast which underlies the distinction between broad focus and narrow focus in statements (see Section "Perception of native intonation"). Children have the ability to produce the difference in association and alignment between the two bitonal accents as early as 21 months, and thus well before the onset of combinatorial speech. Unlike for Dutch and EP, Spanish- and Catalan-learning children have been reported to master tune-text alignment from the beginning of speech (Prieto et al., 2012). Differently from previous studies that were based on spontaneous child speech, Astruc, Payne, Post, Vanrell, and Prieto (2013) designed a study to examine the phonetic detail of the realization of tonal targets between 24 months and six years of age in three different languages – Spanish, Catalan and English, using materials elicited by means of a controlled naming task. Their results show precise realization of tonal targets from an early age, together with differences across languages: Spanish-learning children already displayed adult-like alignment, especially of high targets, by 24 months, and earlier than Catalan- and English-learning children whose alignment patterns became increasingly target-like as they grow older. The differences across languages that have emerged in these studies strongly suggest an effect of the ambient language on the acquisition of tonal alignment patterns.

Like tonal alignment, tonal scaling was also measured in early child speech with different results across languages. The scaling of nuclear contours was not initially adult-like in Catalan, Spanish, and European Portuguese children, with the low boundary tone of statements (L%) first produced as a kind of mid tone (Frota, Cruz, Matos, & Vigário, 2016; Prieto et al., 2012). Results from EP show a rapid development of scaling: by 18 months, the scaling of L% is already target-like, with other boundary tones (such as !H%) becoming target-like in their scaling between 20 and 22 months. Thus, a fine control of scaling seems to begin to evolve very early, even before alignment. The development of scaling, however, seems to succeed the mastering of alignment in Spanish and Catalan (Lleó et al., 2004; Prieto et al., 2012). Spanish-learning children do not show target-like scaling of nuclear and prenuclear accents in statements and yes-no questions before 36 months (Lleó & Rakow, 2011), whereas alignment is close to the adult-target much earlier. Nevertheless, Spanish, Catalan and English 24 month-olds have been shown to be able to control pitch scaling across words with different stress patterns, thus adjusting to the space available for the tonal movement to unfold (Astruc et al., 2013). Thus, we need to conclude that overall the phonetics of intonation is acquired early, and detailed prosodic information precedes fine-grained segmental information in acquisition, given that the phonetics is already close to the adult model when children still use truncation, simplification, or metathesis to deal with challenging segmental and word structures.

The language-specific properties that may influence the acquisition of the phonetics of intonation have yet to be examined. For example, in EP, the intonational grammar of scaling is simpler than the grammar of alignment, as scaling differences are not explored and encoded in the intonation system, unlike alignment differences (Frota, 2014). By contrast, in Spanish and Catalan the intonational grammar of scaling is more complex than in EP, with a variety of nuclear pitch accents that contrast precisely in the scaling of the peak (H + L* and ¡H + L*, L + H* and L + ¡H*), as well more complex scaling contrasts in boundary tones (e.g., LH% and L!H% – Hualde & Prieto, 2015; Prieto, 2014). This seems to suggest that phonological complexity impacts on the acquisition of tonal implementation, with more complex tonal dimensions being acquired later than simpler ones. This suggestion is in line with findings about the prosodic marking of focus in child speech in languages like Dutch and Swedish. In Dutch, prosodic focus marking involves the use of accent placement, deaccenting, choice of pitch accent type, and the particular phonetics of focal accents (Chen, 2011). In Swedish, it involves the interaction between lexical pitch accents and phrasal intonation, namely a prominence-marking high tone (Bruce, 2007). In both languages, using semi-spontaneous productions elicited by means of a picture-matching game, the phonology of prosodic focus was found to be about right by four-five years of age, with the phonetics of focus marking developing later than the phonology (Chen, 2011; Chen, this volume; Romøren & Chen, 2015). These studies of focus marking in child language provide further support for results on the development of intonation in the first years of life, showing that the phonological patterns are in place before the phonetics matches the adult language.

Conclusion

This chapter focuses on early development of intonation, drawing on recent findings from perception and production, for several languages. Most of the studies reviewed have used the AM framework of intonational analysis, allowing for cross-linguistic comparison of early intonation patterns in relation to the adult intonation system.

The recent growing body of literature on infants' perception of intonation has shown that discrimination abilities differ as a function of the types of intonation contrasts (manifested as pitch direction, pitch height, or pitch alignment distinctions), and according to the ambient language. The first finding revealed the importance of the cues that signal a given contrast, and the potential impact for language acquisition of the presence or absence of early discrimination abilities related to a specific intonation contrast that cues a given linguistic distinction (e.g., sentence types, information structure, phrasing). The second finding highlighted that, alongside a well-documented precocious sensitivity to prosody, there is an effect of language experience in the perception of intonation patterns earlier in development than for the perception of vowels, consonants, or even lexical tone. These precocious perception abilities in the first year of life, and their attunement to the ambient language, set the stage for emerging intonation in production throughout the second year.

On the production side, cross-linguistic evidence points to an early development of the intonational system of the ambient language (albeit with differences across languages in relation to the developmental path): the phonological inventory of nuclear contours is adult-like close to the onset of speech, together with the form-meaning relationships of the distinct tunes, and further developments in the use of prenuclear accents and pitch accent distribution evolve quickly and seem to be in place by the end of the second year of life. Although the acquisition of the phonetics of tonal alignment and tonal scaling lags behind that of phonological patterns, several aspects of the phonetics of intonation are also acquired by 24 months, preceding fine-grained segmental information. Furthermore, there is convergent cross-linguistic evidence that emerging intonation, and its early development, are independent from the onset of combinatorial speech.

Findings from early perception and production of intonation suggest a link between the two, that supports precocious development. Infants' perception revealed abilities of prosodic analysis of the speech signal that might help the infant to acquire the lexicon and to bootstrap aspects of the grammar of the native language. It is therefore not surprising that key aspects of the intonational system develop early in child speech. As for perception, the findings from production are compatible with the hypothesis that prosody precedes and might promote (further) grammatical development. There are aspects of intonational development, however, that are acquired later than the first years of life, as in the case of focus development (Section "Acquiring the phonetics of intonation"; Chen, this volume; Ito, this volume), or the relationship between speaker belief states and prosody (Armstrong & Hübscher, this volume). Even if intonational development naturally proceeds with age, the perception and production abilities acquired early are crucial to foster further language development (see also, Bhatara et al., this volume; de Carvalho et al., this volume).

The overall finding of differences across languages in the developmental trajectories of early intonational acquisition suggests that perception is not simply guided by (domain-general) abilities, and early production is not only driven by neuromotor constraints, which would entail similar paths. However, more work is needed to better define the language-specific properties that may influence infants' sensitivity to pitch differences across languages, as well as differences in the acquisition of the phonology and phonetics of intonation revealed in production.

Despite recent advances, the current understanding of how intonation is acquired is still fragmentary. The AM framework of intonation analysis has offered an useful tool to investigate the development of intonation in perception and production, but only a few languages have been examined so far. The fact that, in the languages covered, only a handful of intonation contrasts, and their realizations, have been investigated is a further limitation. It is also important to note that all studies to date that have examined infants' perceptual abilities are discrimination studies. The lack of identification studies in this area, although probably motivated by the young ages observed, is a particularly relevant issue that needs to be acknowledged and addressed in future research. Clearly, more data and, in particular, more controlled experiments are needed. These could take advantage of 'old' behavioral methods in combination with newer techniques, such as event-related potentials (ERPs), near infrared spectroscophy (NIRS), or eye-tracking, all of which may be used to assess early stages in development and hold the promise to offer new insights in this exciting field of research.

Understanding the developmental path of intonation towards a full adult-like competence will not only impact on our knowledge of intonation itself, but contribute to our understanding of the acquisition of other linguistic categories and structures, given the potentially bootstrapping role of intonational cues (Höhle, 2009). Furthermore, as a lack of sensitivity to intonation, together with its atypical production, may indicate low social skills and developmental disorders (Koegel et al., 2010; Peppé, this volume), it will also contribute towards tracing the development of communication and social interaction, and signal language impairments.

Funding

This research was supported by grants PTDC/CLE-LIN/108722/2008, EXCL/MHC-LIN/0688/2012, PTDC/MHCLIN/3901/2014, and IF/01614/2012 from the Foundation for Science and Technology (Portugal) and Fundo Social Europeu – Programa Operacional Potencial Humano.

Acknowledgements

The authors would like to thank the reviewers and the editors of this volume for comments on previous versions of this chapter. The authors also gratefully thank Marina Vigário, Marisa Cruz, Cátia Severino, Susana Correia, Nuno Matos and Simão Cortês for help at different stages of the present work. Special thanks are due to the children, families and nurseries that have collaborated in our studies.

References

Astruc, L., Payne, E., Post, B., Vanrell, M. M., & Prieto, P. (2013). Tonal targets in early child English, Spanish and Catalan. *Language and Speech*, 56(2), 229–253. doi: 10.1177/0023830912460494

Benavides-Varela, S., Hochmann, J. R., Macagno, F., Nespor, M., & Mehler, J. (2012). Newborn's brain activity signals the origin of word memories. *Proceedings of the National Academy of Sciences*, 109(44), 17908–17913. doi: 10.1073/pnas.1205413109

Best, C. T., Levitt, A., & McRoberts, G. W. (1991). Examination of language-specific influences in infants' discrimination of prosodic categories. *Proceedings of the XIIth International Congress of Phonetic Sciences 4* – Aix-en Provence, 162–165.

Bruce, G. (2007). Components of a prosodic typology of Swedish intonation. In T. Riad, & C. Gussenhoven (Eds.), *Tones and tunes: Typological studies in word and sentence prosody* (Vol. 1, pp. 114–145). Berlin: Mouton de Gruyter. doi: 10.1515/9783110207569.113

Butler, J., Vigário, M., & Frota, S. (2016). Infants' perception of the intonation of broad and narrow focus. *Language Learning and Development*, 12(1), 1–13. doi: 10.1080/15475441.2015.1020376

Chen, A. (2011). Tuning information packaging: Intonational realization of topic and focus in child Dutch. *Journal of Child Language*, 38, 1055–1083. doi: 10.1017/S0305000910000541

Chen, A., & Fikkert, P. (2007). Intonation of early two-word utterances in Dutch. *Proceedings of the XVIth International Congress of Phonetic Sciences* – Saarbrücken, 315–320.

Christophe, A., Dupoux, E., Bertoncini, J., & Mehler, J. (1994). Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. *Journal of the Acoustical Society of America*, 95, 1570–1580. doi: 10.1121/1.408544

D'Imperio, (2002). Italian intonation: An overview and some questions. *Probus*, 14(1), 37–69.

DePaolis, R. A., Vihman, M. M., & Kunnari, S. (2008). Prosody in production at the onset of word use: A cross-linguistic study. *Journal of Phonetics*, 36, 406–422. doi: 10.1016/j.wocn.2008.01.003

Elordieta, G., & Hualde, J. I. (2014). Intonation in Basque. In S.-A. Jun (Ed.), *Prosodic typology II: The phonology of intonation and phrasing* (pp. 405–463). Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199567300.003.0014

Esteve-Gibert, N., & Prieto, P. (2013). Prosody signals the emergence of intentional communication in the first year of life: Evidence from Catalan-babbling infants. *Journal of Child Language*, 40(5), 919–944. doi: 10.1017/S0305000912000359

Filipe, M., Branco, P., Frota, S., Castro, L., & Vicente, S. (2014). Atypical prosody in Asperger syndrome: Perceptual and acoustic measurements. *Journal of Autism and Development Disorders*, 44(8), 1972–1981. doi: 10.1007/s10803-014-2073-2

Frota, S. (2014). The intonational phonology of European Portuguese. In S.-A. Jun (Ed.), *Prosodic typology II: The phonology of intonation and phrasing* (pp. 6–42). Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199567300.003.0002

Frota, S., Cruz, M., Matos, N., & Vigário, M. (2016). Early prosodic development: Emerging intonationand phrasing in European Portuguese. In M. E. Armstrong, N. Henriksen, & M. M. Vanrell (Eds.), *Intonational grammar in Ibero-Romance: Approaches acrosslinguistic subfields* (pp. 277–294). Amsterdam: John Benjamins.

Frota, S., & Prieto, P. (2015a). *Intonation in Romance*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199685332.001.0001

Frota, S., & Prieto, P. (2015b). Intonation in Romance: Systemic similarities and differences. In S. Frota & P. Prieto (Eds.), *Intonation in Romance* (pp. 392–418). Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199685332.003.0011

Frota, S., & Vigário, M. (2008). The intonation of one-word and first two-word utterances in European Portuguese. Paper presented at the Third Conference on Tone and Intonation (TIE 3), Lisbon, Portugal.

Frota, S., & Vigário, M. (1995). The intonation of one European Portuguese infant: A first approach. In I. Hub Faria, & M. João Freitas (Eds.), *Studies on the acquisition of Portuguese* (pp. 17–34). Lisboa: APL/Colibri.

Frota, S., Butler, J., & Vigário, M. (2014). Infants' perception of intonation: Is it a statement or a question? *Infancy*, 19(2), 194–213. doi: 10.1111/infa.12037

Frota, S., Butler, J., Lu, S., & Vigário, M. (2016). Infants' perception of native and non-native pitch contrasts. *Proceedings of the Speech Prosody* – Boston, 692–696. doi: 10.21437/SpeechProsody.2016-142

Geffen, S., & Mintz, T. (2015). Can you believe it? 12-month-olds use word order to distinguish between declaratives and polar interrogatives. *Language Learning and Development*, 11(3), 270–284. doi: 10.1080/15475441.2014.951595

Geffen, S., & Mintz, T. (2011). Seven-month-olds discrimination of statements and questions. *Proceedings of the 36th Boston University Conference on Language Development*, Supplement. Retrieved from http://www.bu.edu/bucld/proceedings/supplement/vol36/.

Gussenhoven, C. (2007). The phonology of intonation. In P. de Lacy (Ed.), *The Cambridge handbook of phonology* (pp. 253–280). Cambridge: Cambridge University Press.

Hallé, P., de Boysson-Bardies, B., & Vihman, M. M. (1991). Beginnings of prosodic organization: Intonation and duration patterns of disyllables produced by Japanese and French infants. *Language and Speech*, 34, 299–318.

Höhle, B. (2009). Bootstrapping mechanisms in first language acquisition. *Linguistics*, 47(2), 359–382. doi: 10.1515/LING.2009.013

Hualde, J. I., & Prieto, P. (2015). Intonational variation in Spanish: European and American varieties. In S. Frota, & P. Prieto (Eds.), *Intonation in Romance* (pp. 350–391). Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199685332.003.0010

Jun, S-A. (Ed.) (2014a). *Prosodic typology II: The phonology of intonation and phrasing*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199567300.001.0001

Jun, S-A. (2014b). Prosodic typology: By prominence type, word prosody, and macro-rhythm. In S-A. Jun (Ed.), *Prosodic typology II. The phonology of intonation and phrasing* (pp. 520–540). Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199567300.003.0017

Jusczyk, P. W. (1997). *The discovery of spoken language*. Cambridge, MA: The MIT Press.

Koegel, L. K., Koegel, R. L., Green-Hopkins, I., & Barnes, C. C. (2010). Brief report: Question-asking and collateral language acquisition in children with autism. *Journal of Autism and Developmental Disorders*, 40(4), 509–515. doi: 10.1007/s10803-009-0896-z

Ladd, D. R. (2008). *Intonational phonology* (2nd ed.). Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511808814

Lleó, C. (2016). A preliminary study of wh-questions in German and Spanish child language. In M. E. Armstrong, N. Henriksen & M. M. Vanrell (Eds.), *Intonational grammar in Ibero-Romance: Approaches across linguistic subfields* (pp. 325–350). Amsterdam: John Benjamins.

Lleó, C., & Rakow, M. (2011). Intonation targets of yes-no questions by Spanish and German monolingual and bilingual 2;0- and 3;0-year-olds. In T. Kupisch, & E. Rinke (Eds.), *The development of grammar: Language acquisition and diachronic change* (pp. 213–234). Amsterdam: John Benjamins. doi: 10.1075/hsm.11.14lle

Lleó, C., Rakow, M., & Kehoe, M. (2004). Acquisition of language-specific pitch accent by Spanish and German monolingual and bilingual children. In T. L. Face (Ed.), *Laboratory approaches to Spanish phonology* (pp. 3–27). Berlin: Mouton de Gruyter.
Mattock, K., & Burnham, D. (2006). Chinese and English infants' tone perception: Evidence for perceptual reorganization. *Infancy*, 10(3), 241–265. doi: 10.1207/s15327078in1003_3
Mattock, K., Molnar, M., Polka, L., & Burnham, D. (2008). The developmental course of lexical tone perception in the first year of life. *Cognition*, 106(3), 1367–1381. doi: 10.1016/j.cognition.2007.07.002
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 & Development*, 21(4), 779–784. doi: 10.1016/S0163-6383(98)90044-3
Papoušek, M., Bornstein, M. H., Nuzzo, C., Papoušek, H., & Symmes, D. (1990). Infant responses to prototypical melodic contours in parental speech. *Infant Behavior and Development*, 13, 539–545. doi: 10.1016/0163-6383(90)90022-Z
Prieto, P. (2014). The intonational phonology of Catalan. In S-A. Jun (ed.), *Prosodic typology II: The phonology of intonation and phrasing* (pp. 43–80). Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199567300.003.0003
Prieto, P., & Vanrell, M. M. (2007). Early intonational development in Catalan. *Proceedings of the XVIth International Congress of Phonetic Sciences* – Saarbrücken, 309–314.
Prieto, P., Estrella, A., Thorson, J., & Vanrell, M. M. (2012). Is prosodic development correlated with grammatical development? Evidence from Catalan, English, and Spanish. *Journal of Child Language*, 39(2), 221–257. doi: 10.1017/S030500091100002X
Romøren, A. S. H., & Chen, A. (2015). The acquisition of prosodic focus marking in Central Swedish: Sorting out lexical and post-lexical tones. *Proceedings of the 18th International Congress of Phonetic Sciences* – Glasgow, 0268.
Sansavini, A., Bertoncini, J., & Giovanelli, G. (1997). Newborns discriminate the rhythm of multisyllabic stressed words. *Developmental Psychology*, 33(1), 3–11. doi: 10.1037/0012-1649.33.1.3
Sato, Y., Sogabe, Y., & Mazuka, R. (2009). Development of hemispheric specialization for lexical pitch-accent in Japanese infants. *Journal of Cognitive Neuroscience*, 22(11), 2503–2513. doi: 10.1162/jocn.2009.21377
Savino, M. (2012). The intonation of polar questions in Italian: Where is the rise? *Journal of the International Phonetic Association*, 42(1), 23–48. doi: 10.1017/S002510031100048X
Seidl, A. (2007). Infants' use and weighting of prosodic cues in clause segmentation. *Journal of Memory and Language*, 57, 24–48. doi: 10.1016/j.jml.2006.10.004
Shi, R. (2010). Contextual variability and infants' perception of tonal categories. *Chinese Journal of Phonetics*, 2, 1–9.
Shi, R., Werker, J. F., & Morgan, J. L. (1999). Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition*, 72(2), B11–B21. doi: 10.1016/S0010-0277(99)00047-5
Shukla, M., White, K. S., & Aslin, R. N. (2011). Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. *Proceedings of the National Academy of Sciences of the United States of America*, 108(15), 6038–6043. doi: 10.1073/pnas.1017617108
Snow, D. (2006). Regression and reorganisation of intonation between 6 and 23 months. *Child Development*, 77, 281–296. doi: 10.1111/j.1467-8624.2006.00870.x

Snow, D. (2002). Intonation in the monosyllabic utterances of 1-year-olds. *Infant Behavior and Development*, 24, 393–407. doi: 10.1016/S0163-6383(02)00084-X

Snow, D. (1994). Phrase-final syllable lengthening and intonation in early child speech. *Journal of Speech and Hearing Research*, 37, 831–840. doi: 10.1044/jshr.3704.831

Snow, D., & Balog, H. L. (2002). Do children produce the melody before the words? A review of developmental intonation research. *Lingua*, 112, 1025–1058. doi: 10.1016/S0024-3841(02)00060-8

Soderstrom, M., Ko, E., & Nevzorova, U. (2011). It's a question? Infants attend differently to yes/no questions and declaratives. *Infant Behavior and Development*, 34, 107–110. doi: 10.1016/j.infbeh.2010.10.003

Sundara, M., Molnar, M., & Frota, S. (2015). The perception of boundary tones in infancy. *Proceedings of the 18th International Congress of Phonetic Sciences* – Glasgow, 0361.

Thorson, J., Borras-Comes, J., Crespo-Sendra, V., Vanrell, M. M., & Prieto, P. (2014). The acquisition of melodic form and meaning in yes-no interrogatives by Catalan and Spanish speaking children. *Probus*, 26(1), 59–82.

Trehub, S. E., Bull, D., & Thorpe, L. (1984). Infants' perception of melodies: The role of melodic contour. *Child Development*, 55, 821–830.

Yeung, H. H., Chen, K. H., & Werker, J. F. (2013). When does native language input affect phonetic perception? The precocious case of lexical tone. *Journal of Memory and Language*, 68(2), 123–139. doi: 10.1016/j.jml.2012.09.004

Zahner, K., Schönhuber, M., & Braun, B. (2015). The limits of metrical segmentation: intonation modulates infants' extraction of embedded trochees. *Journal of Child Language*, 18, 1–27. doi: 10.1017/S0305000915000744

CHAPTER 9

Prosodic phonology in acquisition

A focus on children's early word productions

Margaret Kehoe
University of Geneva

This chapter investigates how the theory of prosodic phonology has been applied to child language data, focusing on children's early words. First, we consider early accounts of prosodic structure development in which the tools of prosodic phonology were used to explain the shape of children's word productions. We then go on to consider later accounts in which factors apart from prosodic structure, such as the frequency of input forms, perceptual bias, and segmental factors, have been integrated into recent models. Referring to findings on a wide range of languages (e.g., Spanish, Catalan, Portuguese, Hebrew, Greek, and Japanese) we examine support for the minimal word, a bimoraic constraint on word production, and compare prosodic structure development across different groups of languages.

Introduction

The theory of prosodic phonology assumes that an abstract organizational structure divides speech into prosodic units of various sizes. These units are arranged in a hierarchy that extends from moras and syllables at the bottom to intonational phrases at the top (Nespor & Vogel, 1996). Many studies in phonological acquisition have explored the relationship between these linguistic structural units and their manifestation in child speech. Some studies have asked questions about how the hierarchy is acquired – whether children have access to all units at the beginning or whether they acquire units in a bottom-up versus top-down fashion. Other studies have examined children's development of prosodic structure over time. The aim of this chapter is to investigate the ways in which the theory of prosodic phonology has been applied to the field of phonological acquisition, focusing on children's early word productions (see also Demuth, this volume). Thus, we put the emphasis on prosodic structure approaches involving the foot and the prosodic word. For an alternate prosodic account of early word production according to a usage-based

DOI 10.1075/tilar.23.09keh

approach, see Vihman (this volume) and for information on the role of prosody in speech segmentation and word learning see Teixidó, François, Bosch, and Männel (this volume) and Thorson (this volume). We start by providing a background to the theory and then proceed to examine its application to child language.

Theory of prosodic phonology

The theory of prosodic phonology developed out of a series of studies published in the mid-seventies through to the nineties of the last century (Clements & Keyser, 1983; Halle & Vergnaud, 1980; Liberman & Prince, 1977; Nespor & Vogel, 1986; Selkirk, 1980), the aim of which was to account for important sound patterns by reference to structure above the phonological segment. As such, the approach was allied to the general movement of non-linear, metrical, and auto-segmental phonology. Two of the earliest proponents of this approach, Liberman and Prince (1977), recognized the usefulness of a hierarchy of phonological constituents to capture prominence relationships among words and syllables in a sentence. Selkirk (1980) went on to enrich Liberman and Prince's (1977) conception of suprasegmental structure by positing the existence of a set of prosodic categories, that is, "isolable subunits of prosodic structure" (Selkirk, 1980: 565), which allowed a straightforward solution to the problems of English stress and rhythm. Probably the most standard resource on Prosodic Phonology is Nespor and Vogel's (1986) book *Prosodic Phonology*, which provides further evidence for a "theory that organizes a given string of language into a series of hierarchically arranged constituents that in turn form the contexts within which phonological rules apply" (Nespor & Vogel, 1986: 6). Nespor and Vogel (1986) outline the general principles of the theory as well as offer internal and external evidence for each of the constituents of the prosodic hierarchy. According to the authors, these constituents allow a better characterization of phonological rules which govern segments (e.g., American English flapping, British English r-insertion) as well as rules which govern intonational, rhythmic and phrasing patterns, all of which have proven difficult to account for by syntactic structure alone.

Different versions of the prosodic hierarchy have been proposed over the years. An adaptation of Nespor and Vogel's (1986) version of the prosodic hierarchy is shown in (1).

(1) Utterance (U)
Intonation Phrase (IP)
Phonological Phrase (PhP)
Prosodic Word (ω)
Foot (F, ϕ)
Syllable (σ)
Mora (μ)

The higher level constituents, the intonational and phonological phrase, have an important role in formalizing intonation and boundary effects. These units are integral to the syntax-phonology interface, although it is generally acknowledged that there is no clear relationship between prosodic and syntactic constituents. Minimality constraints, phonotactic generalizations, and the application of phonological processes support the domain of the prosodic word (Hall & Kleinhenz, 1999; Peperkamp, 1997) whereas rhythm generalizations support the domain of the foot (Hayes, 1995). The foot is also the unit of stress assignment. There are two main types of feet: trochaic and iambic. A **trochee** or trochaic foot has the head syllable on the left. An **iamb** or iambic foot has the head syllable on the right. The syllable is the locus of phonotactic generalizations relating to the Sonority Sequencing Principle (Blevins, 1995). Finally, phenomena such as compensatory lengthening and closed syllable shortening support the constituent of a mora which is a unit of syllable weight (Ota, 2001).

Prosodic phonology in acquisition

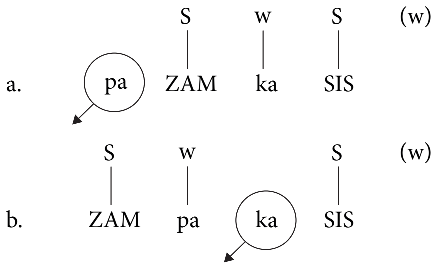
Many authors have appealed to the prosodic units, the mora, the syllable, the foot, and the prosodic word, to explain a number of phenomena in child phonology. In particular, one set of findings has revolved around the central idea that there is a shape constraint in early acquisition such that children's word productions conform to a consistent size and rhythmic pattern (Demuth & Fee, 1995; Fikkert, 1994; Pater, 1997). Depending upon the approach, this constraint has taken the form of a bimoraic minimum or maximum, referred to as the minimal word constraint (Demuth & Fee, 1995; Ota, 2001). It has taken the form of a universal constraint on production (and perception), the trochaic bias (Allen & Hawkins, 1980), or a language-specific constraint implicated in theories of stress development such as a trochaic foot template (Fikkert, 1994). Early accounts of prosodic development which appealed to these constraints include Gerken's (1991, 1994) Strong (Weak) (S(W)) production template, Fikkert's (1994) account of stress acquisition in Dutch, and Demuth and Fee's (1995) account of prosodic development based on the units of the prosodic hierarchy.

Early accounts of prosodic structure development

Gerken's S(W) production template

Drawing on Allen and Hawkin's (1980) notion of a trochaic template, Gerken (1991, 1994) accounted for children's unstressed syllable deletions in multisyllabic words on the basis of a S(W) production template. Gerken (1991, 1994) proposed that children align a trochaic S(W) metrical constraint at the beginning of an intended word. Weak syllables that do not fit the template are omitted. For example, Gerken showed that when children were asked to imitate a four-syllable word with the metrical pattern WSWS such as paZAMkaSIS, children omitted the first unstressed syllable more frequently than the second; in ZAMpakaSIS, children deleted the second weak syllable more frequently than the first, presumably because in both situations, the weak syllable did not slot into the trochaic template (see (2)).

(2) Gerken's (1991, 1994) production-based metrical constraint



Fikkert's (1994) model of stress acquisition

One of the first authors to consider stress acquisition in terms of the development of prosodic structure was Fikkert (1994). She noted that several phenomena provide insight into the acquisition of stress. The most obvious one was stress shift but others include the deletion and addition of syllables. Fikkert (1994) framed syllable deletion in terms of the linguistic phenomena of truncation, which involves "mapping of a melody to a specified template" (Fikkert, 1994: 199). Thus, she linked syllable deletion to foot structure in a similar way to that of Allen and Hawkin's (1980) and Gerken (1991, 1994).

Focusing on the phenomena of stress errors, truncation, and syllable epenthesis, she documented four stages of stress development in Dutch. Her interpretation of results appealed to theoretical devices within the theory of prosodic morphology (McCarthy & Prince, 1986, 1990), such as prosodic circumscription and template mapping, and was formulated within the principles-and-parameters view of language development (Dresher & Kaye, 1990). Transition from one developmental stage to the next involved the setting of metrical parameters or the extension of the child's prosodic template. Stages of stress development for the word konijn "rabbit" a two syllable word with final stress, an exceptional form in Dutch, are illustrated in (3).

(3) Stages of stress acquisition in two-syllable words with final stress (e.g., konijn) according to Fikkert (1994)

Stage 1	Prosodic circumscription and mapping to trochaic template			
	konijn /koːˈnɛin/	[ŋɛi], [lɛin]	'rabbit'	Tom 1;5.14
Stage 2	Circumscription of an additional syllable and mapping to a trochaic template			
	konijn /koːˈnɛin/	[ˈkuːlɑ]	'rabbit'	Tom 1;7.23
Stage 3	Template expands to two feet.			
	konijn /koːˈnɛin/	[ˈkɔˈlɛi], [ˈkɔˈnɛin]	'rabbit'	Tom 1;7.23
Stage 4	Acquisition of the main stress rule			
	konijn /koːˈnɛin/	[koːˈlɛin]	'rabbit'	Tom 1;8.6

At the first stage, the child extracts or circumscribes the rightmost stressed syllable of the target form and maps it onto a trochaic template, resulting in a monosyllabic output. At the second stage, the child extracts the leftmost syllable of the target, adds it to his/her production, and maps this onto the trochaic template. This results in stress shift because the unstressed syllable of the target form receives the primary stress. At the third stage, the template expands from one to two feet, leading to forms with level stress. At the final stage, the child sets stress parameters such as the main stress rule, thus allowing target-like forms (forms with one primary stress) to emerge.

Demuth and Fee's (1995) prosodic hierarchy account

Following closely behind Fikkert's (1994) model of stress acquisition, Demuth and Fee (1995; see also Demuth, 1996a, 1996b) developed a prosodic structure account of children's early word productions. On the basis of English and Dutch data, they proposed that children's earliest words take the form of linguistically motivated units of the prosodic hierarchy: Children first have access to syllables (σ), then, to feet (F) and then later again, to prosodic words (PrWd) (as shown in (4)). In between the development of feet and prosodic words, moraic units (μ) become important.

(4) Stages of Prosodic Representations (adapted from Demuth and Fee, 1995: 20)

Stage 1	>	Stage 2a	>	Stage 2b,c	>	Stage 3
σ		F		μ		PrWd

The most well-known aspect of Demuth and Fee's (1995) account is the minimal word stage (see Section *Minimal Word Constraint*), whereby they claim that children pass through a stage in which words are minimally and maximally bimoraic. Children first realize minimal words as disyllabic forms, but later as monosyllabic forms, when coda consonants and vowel length contrasts emerge. Following the minimal word stage, words progress from a single foot to two-feet and then later include unfooted syllables, indicative of children's abilities to realize phonological (i.e., prosodic) words. During the stress-feet stage, children initially place equal stress on both stressed syllables but gradually acquire the ability to produce one primary stress per prosodic word. Demuth and Fee's (1995) stages of development are shown in (5).

(5) Demuth and Fee's (1995) stages in the development of prosodic structure

Stage 1. Core syllables – CV
No vowel length distinctions

Stage 2. Minimal words
a. Core syllables – (C)VCV
b. Closed syllables – (C)VC
c. Vowel length distinctions – (C)VV

Stage 3. Stress Feet
a. One Stress-Foot
b. Two Stress-Feet – each with primary stress
c. Feet – one primary stress per word

Stage 4. Phonological Words
Extrametrical syllables permitted

Demuth and Fee (1995) consider two reasons as to why children display restricted word shapes in early phonological development. The first is that children do not have access to the full prosodic hierarchy. Children's forms are first made up of a syllable and then later of more articulated prosodic structures. This approach is akin to maturational accounts, which assume that linguistic structure develops over time (Borer & Wexler, 1987). The second is the possibility that children have access to all units but only display evidence of some of them due to high ranking constraints which yield unmarked prosodic structure. This latter explanation is consistent with optimality theoretical accounts of phonological development. While much of recent work in phonological acquisition seems to support the second view, some researchers uphold the position that children begin acquisition without the full set of prosodic constituents (Goad, 2016; Goad & Brannen, 2003). Goad (2016), for example, argues that a constraint-based account cannot derive a stage in which CV or core syllables are optimal, suggesting that, at this point of time, foot structure has not been projected.

Later accounts of prosodic structure development

Since the publication of these earlier accounts, an accumulation of data has accrued on prosodic development in a diverse range of languages. Whereas the original studies focused on prosodic development in Dutch and English, there are now studies on the development of prosodic structure in German (Grimm, 2007), Greek (Tzakosta, 2004), Spanish (Lleó, 2002, 2006; Prieto, 2006), Brazilian Portuguese (Santos, 2005, 2006), European Portuguese (Correia, 2009; Vigário, Freitas, & Frota, 2006), French (Braud & Wauquier-Gravelines, 2004, April; Goad & Buckley, 2006; Paradis, Petitclerc, & Genesee, 1997; Rose, 2000; Vihman, DePaolis, & Davis, 1998), Japanese (Ota, 2001), Hebrew (Adam & Bat-El, 2008a, 2009; Ben-David, 2012), and Finnish (Saaristo-Helin, Kunnari, & Savinainen-Makkonen, 2011). The study of different languages has led to observations that cannot be easily incorporated into earlier models of prosodic development. In addition, earlier approaches placed most of the emphasis on the theoretical tools of prosodic phonology, whereas later approaches have recognized the need to integrate factors, apart from prosodic structure, into their models. These factors include frequency of input forms, perceptual bias, and segmental factors. In the following sections, we summarize the findings of these later studies in terms of four main themes: factors influencing word shape apart from prosodic structure; the minimal word constraint; prosodic structure development (i.e., following the minimal word period); and top-down versus bottom-up development.

Factors influencing word shape apart from prosodic structure

Frequency of input forms

Frequency of input forms has been identified as an important factor influencing word shapes. Children produce monosyllables for a longer period of time in languages which contain high percentages of monosyllables such as English and Catalan (Prieto, 2006; Roark & Demuth, 2000). They produce words of three or more syllables early in languages that are characterized by high percentages of multisyllabic words, such as Finnish and Spanish (Lleó, 2006; Savinainen-Makkonen, 2000). They produce words with initial unstressed syllables early in Spanish most likely due to the high frequency of WS and WSW words in the target language (Lleó & Demuth, 1999).

Vigário et al. (2006) conducted analyses of the statistical properties of adult input and child output and found that adult speech provided a close match to the prosodic word shapes in child speech. That is, the proportions of monosyllabic CV and tri-syllabic forms in child speech were directly correlated to those in the adult input. In sum, any account of children's prosodic word shapes needs to integrate both grammar (i.e., prosodic structure) and input frequency (Vigário et al., 2006).

Perceptual bias or acoustic prominence

Before prosodic structure accounts were proposed, investigators accounted for children's syllable deletion or truncation patterns on the basis of perceptual salience or acoustic prominence (Blasdell & Jensen, 1970; Echols & Newport, 1992). Authors observed that the syllables most likely to be produced in multisyllabic words were the ones that received acoustic prominence, because of their position in the word and the application of stress. Echols and Newport's (1992) study of English-speaking children's multisyllabic productions demonstrated that stressed and word-final unstressed syllables were more frequently and accurately produced than non-final unstressed syllables. Similar observations were reported later by Kehoe and Stoel-Gammon (1997b), Pater (1997), and Snow (1998). In SWW words (e.g., animal, elephant), a metrical template account (Gerken, 1991, 1994) would predict that the first weak syllable is preserved but the findings of these authors revealed that it was the second or word-final weak syllable that was most often preserved in English.

Optimality theoretical (OT) approaches to truncation have offered a way of reconciling prosodic structure with perceptual bias (Pater & Paradis, 1996; Pater, 1997). Pater and Paradis (1996) dissect truncation into two main components: (a) the size restriction and (b) content preservation. Essentially, prosodic structure is reflected by the size restriction and perceptual bias by content preservation. To achieve the pattern of stressed and word-final syllable preservation, Pater and Paradis (1996) employ two faithfulness constraints: StressFaith and RightAnchor. These constraints grant special status to the heads and edges of domains and are grounded in the functional importance of word edges and acoustic salience.

Segmental factors

Another factor that may influence syllable deletion and, consequently, prosodic shape is the segmental content of the target word. Kehoe and Stoel-Gammon (1997b) observed that weak syllables with obstruent onsets were preserved more frequently than weak syllables with sonorant onsets in English-speaking children's productions of real and nonsense words. Taelman (2004), as well as observing this effect in Dutch-speaking children, documented other segmental factors that accounted for a considerable percentage of children's truncations. These included: (a) vowel type (e.g., full vowel quality vs. schwa), (b) syllable structure (e.g., the presence or absence of a coda consonant), and (c) the occurrence of two reduplicating onset consonants (e.g., papagai /ˌpapəˈgaj/). Savinainen-Makkonen (2000) reported that the presence of a stop consonant influenced whether syllables were retained in Finnish-speaking children's productions. Ota (2001) observed consistent syllable deletion by Japanese-speaking children when target syllables contained devoiced vowels or had flap onsets. Finally, Adam and Bat-El (2008b) report that some Hebrew-speaking children preferentially select syllables with /a/ at the early stages of development (see also Tzakosta, 2004, for Greek). In sum, studies on different languages indicate that segmental factors influence children's syllable deletion or truncation patterns. This factor needs to be integrated into models of prosodic word shape.

Minimal word constraint

Following Demuth and Fee's (1995) proposal of a minimal word period in early child speech, numerous studies have sought support for this period in the productions of children speaking different languages. Studies by Kehoe and Stoel-Gammon (2001), Demuth, Culbertson, and Alter (2006), and more recently by Miles, Yuen, Cox, and Demuth (2016) show that English-speaking children produce codas more frequently after short than long vowels, thus, offering further support for a bimoraic maximum in English. Compelling evidence for a minimal word period comes from Japanese as well (Ota, 2001). Although Japanese may contain monomoraic lexical items, Japanese children frequently lengthen the vowels in these words, as shown in (6).

(6) Lengthening of monomoraic targets by Japanese-speaking children (adapted from Ota, 2001: 86)

/me/	[meː]	Hiromi	1;9.11	"eye"
/ki/	[kiː]	Hiromi	1;9.28	"tree"

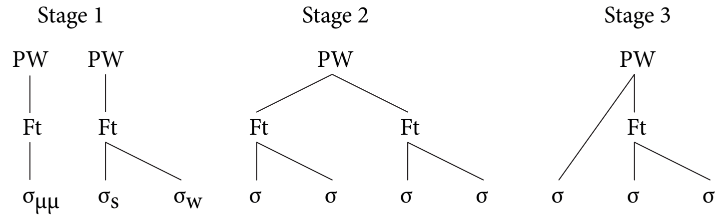
An analysis of pitch accent is consistent with the fact that these output forms conform to bimoraic structure. In Japanese, a HL pitch accent is realized on accented bimoraic words but not on monomoraic words. Ota (2001) observed that when monomoraic targets underwent lengthening, they displayed the HL contour typical of target bimoraic forms.

In contrast, limited support for a minimal word period has been reported in languages such as French, Catalan, Portuguese, and Hebrew (Ben-David, 2012; Demuth & Johnson, 2003; Prieto, 2006; Prieto & Bosch-Baliarda, 2006; Vigário et al., 2006). Demuth and Johnson (2003) document high numbers of sub-minimal forms in a French-speaking child's early words, consistent with the prosodic structure of French which may contain monomoraic lexical items. Prieto and Bosch-Balliarda (2006) also found that Catalan-speaking children produced sub-minimal forms and did not augment them by lengthening the vowel (see also Ben-David's, 2012, study of Hebrew-speaking children). Similarly, Vigário et al. (2006) found proportions of sub-minimal and supra-minimal forms in the outputs of Portuguese-speaking children in comparable proportions to those found in the ambient language. Thus, it seems that the minimal word constraint is not universal but may be active in certain languages such as English and Japanese, which are quantity sensitive or moraic (Prieto, 2006).

Prosodic structure development

Accounts by Fikkert (1994) and Demuth and Fee (1995) viewed development as involving the expansion of prosodic structure from one stress-foot, to two stress-feet and, finally, to a stage in which extrametrical or unfooted syllables could be permitted. The stress foot was a trochaic one consistent with the language-specific stress patterns of Dutch and English. We summarize these stages in (7).

(7) Development of prosodic structure in English and Dutch according to prosodic structure accounts by Demuth and Fee (1995) and Fikkert (1994)



We now consider more recent studies on prosodic structure development in Dutch, English, and German and in other languages such as Portuguese, Spanish, Catalan, French, and Hebrew. The stages depicted in (7) through to (10) are assumed to be descriptions of the phonological representations available to the child at a given stage of development.

Findings on Germanic languages
Subsequent studies in Dutch and English were not able to replicate all aspects of Fikkert's (1994) and Demuth and Fee's (1995) models (Kehoe, 1999/2000; Kehoe & Stoel-Gammon, 1997a, 1997b; Taelman, 2004; Wijnen, Krikhaar, & Den Os, 1994). Taelman (2004), employing a dense data-base of a single Dutch-speaking child (Maarten database) as well as Fikkert's own database of Dutch-speaking children (CLPF database), did not find evidence for a S(w) (i.e., one foot) template that gave way to a S(w)S(w) (i.e., two feet) template. Instead, she found evidence for a monosyllabic S template that gave way to a variety of prosodic shapes, including wS and sSw. In other words, truncations did not always lead to optimal prosodic shapes.

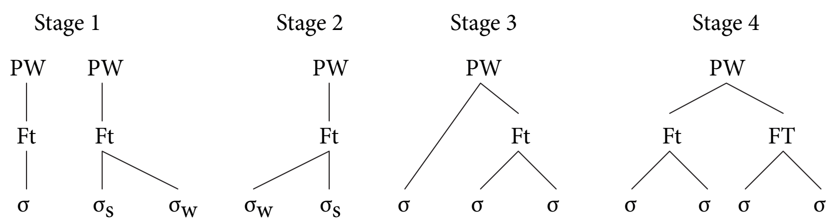
Kehoe (1999/2000) also queried the frequency of shape restrictions in the prosodic development of English. In her analysis of the multisyllabic productions of 39 children, aged 1;10 to 2;10, she found shape uniformity to be less pervasive than what has generally been claimed for English. The majority of two-year-olds in her study (10 out of 12 children) produced output forms greater than one foot, and only one child (aged 2;4 years) produced output forms consistent with two feet. Instead, there was a strong correlation between the prosodic shape of the target word and the child's output. Kehoe articulated an account of the later stages of prosodic acquisition in terms of expanding prosodic (and segmental) faithfulness between input and output.

Grimm's (2007) study of prosodic development in German showed stronger support for Fikkert's (1994) (and Demuth & Fee's, 1995) stages of development, in that children truncated words to a single foot and produced outputs consistent with two feet. One of the four German children, however, displayed an intermediate stage, in which she produced two syllable forms with final stress, forms that would not be consistent with Fikkert's (1994) model but would be with Kehoe's (1999/2000) and Taelman's (2004) accounts.

Findings on "Romance and other" languages
Recent findings in Romance languages and other languages such as Hebrew, Greek, and Japanese suggest clear differences in prosodic development from those in Germanic languages. One difference is that multisyllabic output forms are more frequent (Lleó & Demuth, 1999; Vigário et al., 2006). Another difference is that unstressed syllables appear earlier in word-initial position. These forms may stem from the correct production of iambic target forms or from the addition of a syllable in word-initial position. We divide these studies into those that document an initial stage of trochaic forms and those that find evidence for both trochaic and iambic forms early on.

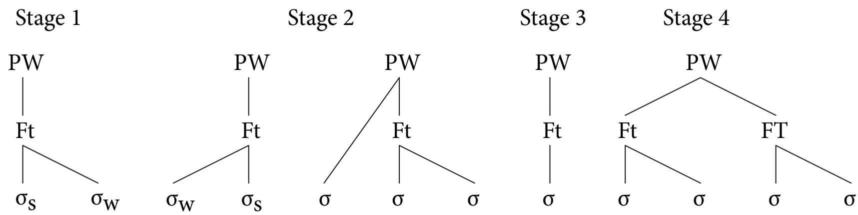
Trochaic forms first. Catalan and Hebrew are languages in which monosyllables and trochaic forms appear before iambic forms. For example, Prieto (2006) and Ben-David (2012) indicate that there is a phase in which WS and WSW targets are produced as S and SW forms respectively. This is somewhat unexpected in Hebrew since it is a language with word-final stress leading some authors to argue that the findings in Hebrew acquisition are consistent with a universal trochaic bias (Adam & Bat-El, 2009). Following the production of trochaic forms, Hebrew- and Catalan-speaking children produce WS targets as WS but reduce WSW targets to SW. Later, they are able to produce WSW forms in which the initial syllable is unfooted. According to Prieto (2006), the results in Catalan can be explained by two maximality constraints. At the first stage, prosodic words are maximally moraic trochees (two moras or a bisyllabic trochac form). At the next stage, they are maximally bisyllabic (either trochaic or iambic). Stages of prosodic structure development in Hebrew and Catalan are shown in (8).

(8) Development of prosodic structure in Catalan and Hebrew



Spanish is similar to Catalan and Hebrew in that bisyllabic trochaic forms are present early on. Lleó (2006), however, reports that word-initial unstressed syllables appear quickly after this, resulting in either WS or WSW forms; only later do monosyllabic forms appear. The late appearance of monosyllables is most likely due to the low frequency of monosyllables in Spanish. Prosodic structure development in Spanish according to Lleó (2006) is shown in (9). In contrast to Lleó (2006), Prieto (2006) observes that WSW targets are truncated for a longer period than WS targets in Spanish, suggesting two substages for stage 2: a. (σˈσ) b. σ(ˈσσ).

(9) Development of prosodic structure in Spanish



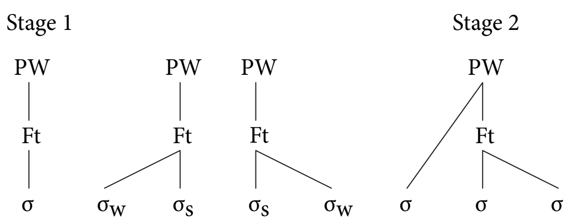
Iambic and trochaic forms early on. Studies in Portuguese (Correira, 2009; Santos, 2005, 2006; Vigário et al., 2006) find no clear word shape at the beginning or find a predominance of iambic forms. Santos (2005, 2006), reporting on two Brazilian Portuguese children's word productions, argues for an iambic bias in Brazilian Portuguese: iambs are realized correctly before trochees; trochees undergo truncation to monosyllables but iambs do not, and WSW words are truncated to both WS and SW forms. Correia's (2009) study of prosodic development in European Portuguese corroborates Santos' (2006) findings, although primarily for the early stages. She observed an initial stage (Stage 1) characterized by monosyllables or by WS forms in which the monosyllable was reduplicated or in which a syllable was epenthesized at the left edge of the word. Following this, there was a period in which both trochaic and iambic forms were equally present. Finally, a definitive trochaic pattern emerged which is consistent with the fact that Portuguese is a trochaic language.

Tzakosta (2004) also reports that Greek-speaking children do not display any strong metrical preferences at the beginning. For example, a WSW target form can be produced either as an iambic or as a trochaic output. Similarly, stress errors create both trochaic and iambic outputs. One other interesting observation of Tzakosta's (2004) study was that she was not able to delineate clear stages of acquisition due to a great deal of within- and between-subject variability of her participants, leading her to argue that children may adopt multiple grammars at any one time. Thus, the enterprise started by Fikkert (1994) and Demuth and Fee (1995) in which stage-like development of prosodic structure was proposed may be a useful generalization but may not represent the reality of variable child data.

Japanese is not a stress language like Portuguese and Greek; however Ota's (2001) analysis of prosodic structure development in Japanese shows similarities with Portuguese and Greek. At the same time as children augment monomoraic target forms to bimoraic, they also produce disyllabic forms. They rarely truncate target disyllabic forms regardless of whether the accent is on the initial or final syllable, whereas they frequently reduce target forms of three-syllables or longer. Thus, the output forms are maximally disyllabic, although they may be iambic- or trochaic-like in form.

A simplified illustration of prosodic structure development is given in (10) for Portuguese, Greek and Japanese. At the earliest stage, monosyllables and disyllabic feet appear at the same time. We do not show moraic structure for monosyllables as they may be monomoraic or bimoraic in Portuguese but bimoraic only in Japanese. In addition, we show the feet as being iambic in form, although analyses in Portuguese suggest that the initial unstressed syllable may be attached to higher prosodic domains (Correira, 2009; Santos, 2006). Due to differences between languages, we do not dwell on the later stages. Correira's (2009) findings on European Portuguese and Tzakosta's (2004) findings on Greek indicate that WSW forms emerge before SWW and WWS forms. Thus, we show this structure at stage 2 in (10).

(10) Development of prosodic structure in Portuguese, Greek, and Japanese



Another language in which iambic forms are present from the beginning is French. French, is not a lexical stress language. Stress is assigned to the last syllable of a phonological phrase (Dell, 1984) but an accent, which has been interpreted as secondary stress or as intonation, may also be assigned to the initial syllable. This accent is cued by pitch, whereas the final accent is cued by duration.

Studies, which have focused on prosodic development in French, reveal that French-speaking children produce monosyllables and iambic forms at the beginning (Demuth & Johnson, 2003; Goad & Buckley, 2006; Rose, 2000; Vihman et al., 1998). Beyond the two-syllable stage, Goad and Buckley (2006) interpret children's productions as showing evidence of either unfooted syllables or two-feet structure, the latter suggesting that children have developed awareness of the word-initial accent. Examples of one foot, unfooted and two feet outputs are given in (11a,b,c) respectively. In essence, Goad and Buckley (2006) outline a developmental path, in which productions develop from one to two stress feet not unlike what has been posited for English and Dutch, however, the prosodic structure is in keeping with the language-specific one of French.

(11) Three-syllable word patterns in French (adapted from Goad & Buckley, 2006: 134)

a. σ σ σ > (σ ˈσ)
papillon /papiˈjɔ̃/ [bəˈpɔː] Clara 1;4 "butterfly"

b. σ σ σ > σ (σ ˈσ)
apricot /abʁiˈko/ [pupæˈkoː] Clara 1;7-1;10 "apricot"

c. σ σ σ > (ˌσ) (σ ˈσ)
Clementine /klemɑ̃tsin/ [ˌætæ̃ˈʃjɔ̃] Clara 1;7-1;10 "Clementine"

Braud and Wauquier-Graveline's (2004, April) study on truncation in French provides support for a three-syllable template, in which the first and the final syllables are anchor points, as shown in (12). These two syllables are respectively filled by a functional element (e.g., a determiner) and by the final accented syllable of the lexical item. The internal syllables, represented as $(\sigma)_n$ are more fragile and subject to omission. This representation is similar to the one posited by Goad and Buckley (2006) in (11c), though, in Braud and Wauquier-Graveline's (2004, April) account, no foot structure is assumed. Indeed, they contest linguistic approaches that have tried to enforce metrical devices from other languages onto French. Thus, they offer an alternate account of prosodic structure development in French.

(12) [ˌσ (σ)$_n$ ˈσ]

Prosodic structure development: Final word

The above groupings of languages do not conform to any obvious typological categories so we may wonder whether there is any basis – linguistic or otherwise – to them. In keeping with our previous discussion on factors influencing word shape, the frequency of word shapes in the input may underlie some of the consistencies in the above groupings. Romance languages are characterized by a higher frequency of multisyllabic words than Germanic languages and by word shapes in which stress (particularly primary stress) is removed from word-initial position (Lleó, 2002; Lleó & Demuth, 1999; Prieto, 2006; Vigário et al., 2006). These characteristics are also true of Greek and Japanese (Ota, 2006; Tzakosta, 2004). Hebrew does not contain a high percentage of multisyllabic words in child-directed speech, but disyllabic words are predominantly iambic, and trisyllabic words have stress on the penult or final syllable (Segal, Nir-Sagiv, Kishon-Rabin, & Ravid, 2009). In contrast, most words in English start with stressed or strong syllables (Cutler & Carter, 1987) and disyllabic iambic forms are infrequent (Roark & Demuth, 2000), as they are in German as well (Bartels, Darcy, & Höhle, 2009). These input tendencies may explain the earlier production of longer forms and the earlier appearance of unstressed syllables in word-initial position in "Romance and other" languages. The nature of unstressed syllables, whether they are reduced or unreduced, may play a role in some of the differences as well. The high percentages of schwa syllables in English and to some extent in Dutch may underlie the slow appearance of unstressed syllables in the Germanic languages (Kehoe, 1999/2000; Taelman, 2004). Whether the distinction between reduced versus unreduced can explain all of the above findings remains tentative, however, since Catalan, Portuguese, and Hebrew also contain reduced or schwa syllables and unstressed syllables emerge early in these languages.

As to why WS forms appear particularly early in Portuguese and Greek (also Japanese), we cannot exclude the possibility that methodological factors underlie some of the apparent trends. Baia and Santos (2011) observed different prosodic tendencies in Brazilian Portuguese children's productions depending upon whether naturalistic or experimental data were utilized. Trochaic outputs were more prevalent in the experimental study whereas iambic outputs were more prevalent in the naturalistic data. These differences reflected whether clitic forms, or nouns and verbs versus nouns only were included. Vihman et al. (1998) also reported a high presence of iambic forms in English children's naturalistic productions when the data included both word and phrasal patterns. Thus, closer attention to the nature of the data (naturalistic or experimental) and whether phrasal productions (or clitic forms) and different word classes (nouns vs. verbs) were included would be needed in order to confirm the reality of the above prosodic representations and the apparent differences across languages.

Top-down versus bottom-up development

Scanning across the different accounts of prosodic development, one can observe that two main approaches have been articulated: top-down versus bottom-up. Approaches based on Germanic languages (e.g., Demuth & Fee, 1995; Fikkert, 1994) posit that children start off producing syllables and then produce feet and prosodic words in accordance with the units of the prosodic hierarchy (Syllable > Feet > Prosodic Word). This is referred to as a bottom-up model. Alternatively, certain authors who have analyzed Romance languages suggest that children's initial productions reflect a top-down model. According to Santos (2005), children start off producing intonational or phonological phrases and, only later, syllables or feet. As was mentioned in the preceding review, Portuguese-speaking children rarely produce monosyllables at the beginning. They produce multisyllabic utterances or, if they produce monosyllables, they reduplicate them or add filler syllables (Correira, 2009; Santos, 2006). Spanish-speaking children produce three-syllable prosodic words containing initial unfooted syllables before they produce monosyllables (Lleó, 2006). These findings question the notion that acquiring prosodic structure necessarily proceeds in a bottom-up direction.

Conclusion

This chapter summarized approaches to prosodic development, which stem from the theoretical tenets of prosodic phonology, focusing on children's early word productions. When one compares earlier and later approaches, one observes that earlier approaches were theoretically united in claiming the importance of prosodic structure, whereas latter approaches have been more eclectic in their orientation, integrating frequency, perceptual bias, and segmental aspects into their accounts. These later studies have shown that prosodic structure on its own cannot account for children's early speech patterns; however, they have not shown how it works together with other factors to provide a better explanatory account. Our attempt to summarize findings across studies has led to the random grouping of languages which defy any linguistic typological links (e.g., Catalan and Hebrew; Portuguese and Greek), and more work is needed to explain why children speaking certain languages (e.g., Catalan and Hebrew) produce trochaic before iambic forms whereas children speaking other languages (e.g., Portuguese and Greek) produce trochaic and iambic forms at around the same time. The goal of future research is to integrate the findings from these varied studies into a unified model, which acknowledges the importance of prosodic structure when working in tandem with other linguistic factors.

Acknowledgements

I would like to thank Pilar Prieto, Núria Esteve-Gibert, as well as two anonymous reviewers for their helpful comments on earlier versions of this chapter.

References

Adam, G., & Bat-El, O. (2008a). The trochaic bias is universal: Evidence from Hebrew. *Proceedings of the Conference on Generative Approaches to Language Acquisition* – Barcelona, 12–24.

Adam, G., & Bat-El, O. (2008b). Segmental effects on syllable selection: Evidence from Hebrew. *Proceedings of the Conference on Generative Approaches to Language Acquisition* – Barcelona, 1–11.

Adam, G., & Bat-El, O. (2009). When do universal preferences emerge in language development? The acquisition of Hebrew stress. *Brill's Annual of Afroasiatic Languages and Linguistics*, 1, 255–282.

Allen, G., & Hawkins, S. (1980). Phonological rhythm: Definition and development. In G. Yeni-Komshian, J. Kavanagh, & C. Ferguson (Eds.), *Child phonology: Production* (Vol. 1, pp. 227–256). New York, NY: Academic Press.

Baia, M., & Santos, R. (2011). The discrepancy between naturalistic and experimental studies: The case of the initial prosodic template in Brazilian Portuguese. *Selected Proceedings of the 5th Conference on Laboratory Approaches to Romance Phonology* – Provo, 127–135.

Bartels, S., Darcy, I., & Höhle, B. (2009). Schwa syllables facilitate word segmentation by 9-month-old German-learning infants. *Proceedings of the 33rd Annual Boston University Conference on Language Development* – Boston, 73–84.

Ben-David, A. (2012). The development of prosodic structure: Evidence from typical longitudinal data. *Brill's Annual of Afroasiatic Languages and Linguistics*, 4, 55–79.

Blasdell, R., & Jensen, P. (1970). Stress and word position as determinants of imitation in first-language learners. *Journal of Speech and Hearing Research*, 13, 193–202. doi:10.1044/jshr.1301.193

Blevins, J. (1995). The syllable in phonological theory. In J. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 206–244). Oxford: Blackwell.

Borer, H., & Wexler, K. (1987). The maturation of syntax. In T. Roeper, & E. Williams (Eds.), *Parameter setting* (pp. 123–172). Dordrecht: Reidel. doi:10.1007/978-94-009-3727-7_6

Braud, V., & Wauquier-Gravelines, S. (2004, April). *Approche gabaritique des phénomènes de troncations en français*. Paper presented at the Journées d'Etudes sur la Parole, Fez, Morocco.

Clements, G. N., & Keyser, S. (1983). *CV phonology: A generative theory of the syllable*. Cambridge, MA: The MIT Press.

Correia, S. (2009). *The Acquisition of primary word stress in European Portuguese*. Unpublished PhD dissertation. University of Lisbon.

Cutler, A., & Carter, D. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language*, 2, 133–142. doi:10.1016/0885-2308(87)90004-0

Dell, F. (1984). L'accentuation dans les phrases en français. In F. Dell, D. Hirst, & J-R. Vergnaud (Eds.), *Forme sonore du langage* (pp. 65–122). Paris: Hermann.

Demuth, K. (1996a). Alignment, stress, and parsing in early phonological words. *Proceedings of the UBC International Conference on Phonological Acquisition* – Vancouver, 113–125.

Demuth, K. (1996b). The prosodic structure of early words. In J. Morgan, & K. Demuth (Eds.), *From signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 171–184). Hillsdale, NJ: Lawrence Erlbaum Associates.

Demuth, K., & Fee, E. J. (1995). *Minimal Words in Early Phonological Development*. Unpublished manuscript. Brown University and Dalhousie University.

Demuth, K., & Johnson, M. (2003). Truncation to subminimal words in early French. *The Canadian Journal of Linguistics*, 48, 211–241. doi:10.1353/cjl.2004.0025

Demuth, K., Culbertson, J., & Alter, J. (2006). Word minimality, epenthesis, and coda licensing in the early acquisition of English. *Language and Speech*, 49, 137–174. doi:10.1177/00238309060490020201

Dresher, B., & Kaye, J. (1990). A computational learning model for metrical theory. *Cognition*, 34, 137–195. doi:10.1016/0010-0277(90)90042-I

Echols, C., & Newport, E. (1992). The role of stress and position in determining first words. *Language Acquisition*, 2, 189–220. doi:10.1207/s15327817la0203_1

Fikert, P. (1994). *On the acquisition of prosodic structure*. Dodrecht: ICG Printing.

Gerken, L. (1991). The metrical basis for children's subjectless sentences. *Journal of Memory and Language*, 30, 431–451. doi:10.1016/0749-596X(91)90015-C

Gerken, L. (1994). A metrical template account of children's weak syllable omissions from multisyllabic words. *Journal of Child Language*, 21, 565–584. doi:10.1017/S0305000900009466

Goad, H. (2016). Phonological processes in child speech. In J. Lidz, W. Snyder, & J. Pater (Eds.), *The Oxford handbook of developmental linguistics* (pp. 43–67). Oxford: Oxford University Press.

Goad, H., & Brannen, K. (2003). Phonetic evidence for phonological structure in syllabification. In J. Van de Weijer., V. Van Heuven, & H. Van de Hulst (Eds.), *The phonological spectrum* (Vol. 2, pp. 3–30). Amsterdam: John Benjamins. doi:10.1075/cilt.234.04goa

Goad, H., & Buckley, M. (2006). Prosodic structure in child French: Evidence for the foot. *Catalan Journal of Linguistics*, 5, 109–142.

Grimm, A. (2007). *The development of early prosodic word structure in child German: Simplex words and compounds*. Unpublished PhD dissertation. University of Potsdam.

Hall, A. T., & Kleinhenz, U. (Eds.) (1999). *Studies on the phonological word*. Amsterdam: John Benjamins. doi:10.1075/cilt.174

Halle, M., & Vergnaud, J. (1987). *An essay on stress* (Current Studies in Linguistics 15). Cambridge, MA: The MIT Press.

Hayes, B. (1995). *Metrical stress theory. Principles and case studies*. Chicago, IL: The University of Chicago Press.

Kehoe, M. (1999/2000).Truncation without shape constraints: The latter stages of prosodic acquisition. *Language Acquisition*, 8, 23–67. doi:10.1207/S15327817LA81_2

Kehoe, M., & Stoel-Gammon, C. (1997a). The acquisition of prosodic structure: An investigation of current accounts of children's prosodic development. *Language*, 73(1), 113–144. doi:10.2307/416597

Kehoe, M., & Stoel-Gammon, C. (1997b). Truncation patterns in English-speaking children's word productions. *Journal of Speech, Language, and Hearing Research*, 40, 526–541. doi:10.1044/jslhr.4003.526

Kehoe, M., & Stoel-Gammon, C. (2001). Development of syllable structure in English-speaking children with particular reference to rhymes. *Journal of Child Language*, 28, 393–432. doi:10.1017/S030500090100469X

Liberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, 8, 249–336.

Lleó, C. (2002). The role of markedness in the acquisition of complex prosodic structures in German-Spanish bilinguals. *International Journal of Bilingualism*, 6, 291–313. doi:10.1177/13670069020060030501

Lleó, C. (2006). The acquisition of prosodic word structures in Spanish by monolingual and Spanish-German bilingual children. *Language and Speech*, 49, 205–229. doi:10.1177/00238309060490020401

Lleó, C., & Demuth, K. (1999). Prosodic constraints on the emergence of grammatical morphemes: Cross-linguistic evidence from Germanic and Romance languages. *Proceedings of the 23rd Annual Boston University Conference on Language Development* – Boston, 407–418.

McCarthy, J., & Prince, A. (1986). *Prosodic morphology*. Unpublished manuscript. University of Massachusetts and Brandeis University.

McCarthy, J., & Prince, A. (1990). Foot and word in prosodic morphology: The Arabic broken plural. *Natural Language and Linguistic Theory*, 8, 209–283. doi:10.1007/BF00208524

Miles, K., Yuen, I., Cox, F., & Demuth, K. (2016). The prosodic licensing of coda consonants in early speech: Interactions with vowel length. *Journal of Child Language*, 43, 265–283.

Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris.

Ota, M. (2001). Phonological theory and the development of prosodic structure. Evidence from child Japanese. *Annual Review of Language Acquisition*, 1, 65–118. doi:10.1075/arla.1.03ota

Ota, M. (2006). Input frequency and word truncation in child Japanese: Structural and lexical effects. *Language and Speech*, 49, 261–295. doi:10.1177/00238309060490020601

Paradis, J., Petitclerc, S., & Genesee, F. (1997). Word truncation in French-speaking two year olds. In E. Hughes, M. Hughes, & A. Greenhill (Eds.), *Proceedings of the 21st Annual Boston University Conference on Language Development* (pp. 441–452). Somerville, MA: Cascadilla Press.

Pater, J. (1997). Minimal violation and phonological development. *Language Acquisition*, 6, 201–253. doi:10.1207/s15327817la0603_2

Pater, J., & Paradis, J. (1996). Truncation without templates in child phonology. In A. Stringfellow, D. Cahana-Amitay, E. Hughes, & A. Zukowski (Eds.), *Proceedings of the 20th Annual Boston University Conference on Language Development* (Vol. 2, pp. 540–551). Somerville, MA: Cascadilla Press.

Peperkamp, S. (1997). *Prosodic words*. Den Haag: HAG.

Prieto, P. (2006). The relevance of metrical information in early prosodic word acquisition: A comparison of Catalan and Spanish. *Language and Speech*, 49, 231–259. doi:10.1177/00238309060490020501

Prieto, P., & Bosch-Baliarda, M. (2006). The development of codas in Catalan. *Catalan Journal of Linguistics*, 5, 237–272.

Roark, B., & Demuth, K. (2000). Prosodic constraints and the learner's environment: A corpus study. In S. C. Howell, S. A. Fish, & T. Keith-Lucas (Eds.), *Proceedings of the 24th Annual Boston University Conference on Language Development* (pp. 597–608). Somerville, MA: Cascadilla Press.

Rose, Y. (2000). *Headedness and prosodic licensing in the L1 acquisition of phonology*. Unpublished PhD dissertation. McGill University.

Saaristo-Helin, K., Kunnari S., & Savinainen-Makkonen, T. (2011). Phonological development in children learning Finnish: A review. *First Language*, 31, 342–363. doi:10.1177/0142723710396793

Santos, R. (2005). Strategies for word stress in Brazilian Portuguese. In M. Tzakosta, C. Levelt, & J. van de Weijer (Eds.), *Developmental paths in phonological acquisition*. Special issue of Leiden Papers in Linguistics, 2(1), 71–91.

Santos, R. (2006). Word stress acquisition: a comparison between Brazilian Portuguese and Dutch. In D. Bamman, T. Magnitskaia, & C. Zaller (Eds.), *Proceedings of the 30th Annual Boston University Conference on Language Development* (pp. 525–536). Somerville, MA: Cascadilla Press.

Savinainen-Makkonen, T. (2000). Learning long words: A typological perspective. *Language and Speech*, 43, 205–225. doi:10.1177/00238309000430020401

Segal, O., Nir-Sagiv, B., Kishon-Rabin, L., & Ravid, D. (2009). Prosodic patterns in Hebrew child-directed speech. *Journal of Child Language*, 36, 629–656. doi:10.1017/S030500090800915X

Selkirk, E. (1980). The role of prosodic categories in English word stress. *Linguistic Inquiry*, 11, 563–605.

Snow, D. (1998). A prominence account of syllable reduction in early speech development: The child's prosodic phonology of tiger and giraffe. *Journal of Speech, Language, and Hearing Research*, 41, 1171–1184. doi:10.1044/jslhr.4105.1171

Taelman, H. (2004). *Syllable omissions and additions in Dutch child language: An inquiry into the function of rhythm and the link with innate grammar*. Unpublished PhD dissertation. University of Antwerp.

Tzakosta, M. (2004). *Multiple parallel grammars in the acquisition of stress in Greek L1*. Unpublished PhD dissertation. Utrecht University.

Vigário, M., Freitas, M. J., & Frota, S. (2006). Grammar and frequency effects in the acquisition of prosodic words. *Language and Speech*, 49, 175–203. doi: 10.1177/00238309060490020301

Vihman, M., DePaolis, R., & Davis, B. (1998). Is there a "trochaic bias" in early word learning? Evidence from infant production in English and French. *Child Development*, 69, 935–949. doi: 10.2307/1132354

Wijnen, F., Krikhaar, E., & Den Os, E. (1994). The (non)realization of unstressed elements in children's utterances: Evidence for a rhythmic constraint. *Journal of Child Language*, 21, 59–83. doi: 10.1017/S0305000900008679

CHAPTER 10

The development of prosodic structure

A usage-based approach

Marilyn May Vihman
University of York

Scholarly views are divided as to the source of children's knowledge of prosodic structure. Within the framework of a usage-based approach, this chapter compares prosodic structures in children learning four languages at the end of the single-word period in order to identify sources of both similarities and differences between children, within and across language groups. The similarities can generally be traced back to common constraints on the neurophysiology of infant vocal production, while the differences between language groups reflect ambient language accentual patterning and dominant word shapes. Individual differences within and across groups additionally relate to differing child mappings of input forms to familiar production patterns.

Introduction

Scholarly views are divided as to the source of children's knowledge of prosodic structure. Reflecting one prominent view, Selkirk (1996: 201) refers to 'innate knowledge of … universal constraints'. However, since the arguments for infant foreknowledge of linguistic principles remain more logical than empirical, others find it implausible and unnecessary to invoke innate knowledge. Instead, a usage-based approach (Bybee, 2001, 2010; Munson, Edwards, & Beckman, 2012; Pierrehumbert, 2003) to the acquisition of prosodic structure appeals to learning based on both initial perceptual biases of possible evolutionary origin (Kuhl, 1987) and infants' experience, along with neurophysiological maturation, of vocal production practice (Vihman & Croft, 2007; Vihman, 2014).

Kehoe (2013, this volume) has described the theory of Prosodic Phonology and various attempts to test it as an account of development in different languages (see also Demuth, this volume; Vihman, 2014, Chapter 9). This chapter proposes an alternative account of prosodic development, supported by quantitative

DOI 10.1075/tilar.23.10vih

analysis of early child word structures in four languages. The claim is that instead of positing innate knowledge, it is possible to trace the origins of children's initial representations of prosodic structure to three types of learning that increasingly interact over the course of the first years of life: (i) experience of salient elements of the input speech stream, (ii) constraints imposed by the neurophysiology of vocal production, and (iii) memory processes that relate heard patterns to existing knowledge based on production. This approach is in accord with usage-based accounts of adult phonology (and sound change) that argue that 'inherent biases in production, perception, and learning operate continually to influence the range of variants that arise … The synchronic properties of a given phonological system arise in part through these processes' (Hall, Hume, Jaeger, & Wedel, 2016, p. 57). Furthermore, as will be evident in what follows, the analyses proposed here draw on much the same principles as are outlined by these authors with reference to adults: '(i) linguistic categories maintain some record of experienced variation rather than being fully abstract…and (ii) experiencing a particular category variant influences future production and perception behavior not only for that category, but also for similar categories' (p. 57f.).

As regards perception, ample experimental evidence has demonstrated infant sensitivity to the rhythms of the ambient language from birth, initially gained from 'listening' in the womb (e.g., Mehler et al., 1988; see Post & Payne, this volume). Over the course of the first year of life, infant attraction to what is prosodically salient in a given language further shapes the child's emergent representations of speech. For example, Fernald (1991) described the effects on language development of prosodic modulation in infant-directed speech; Jusczyk (1993, 1997) modelled advances over the first year in the grouping of prominent acoustic features such as syllables with higher or more dynamic pitch pattern or longer duration.

As regards production, it is widely acknowledged that the first adult-like syllables are simple sequences of 'mandibular oscillation' (jaw opening and closing: Davis & MacNeilage, 2000), or < CV >, < CVCV >, < VCV >. These vocalizations, which Oller (1980) termed 'canonical babbling', are reliably first observed from about the middle of the first year. Little if any cross-linguistic variation has been reported in the phonetics of these first adult-like syllables, which appear to be largely maturationally based, given typical auditory and vocal capacities (Oller & Eilers, 1988).

How might adult-like vocal production practice serve to orient the child to the prosodic structures of input speech? Although babbling sometimes involves long sequences of repeated or variegated syllables, it more often takes the form of one- and two-syllable strings – and the first words in most of the languages that have been analysed for early lexical and phonological learning are similarly constrained to one or two syllables. Accordingly, it has been proposed that infants respond differently to speech once they have begun to gain stable control of a small number of adult speech sounds (Vihman, 1993, 1996; for experimental support, see DePaolis, Keren-Portnoy, & Vihman, 2016; DePaolis, Vihman, & Keren-Portnoy, 2011; DePaolis, Vihman, & Nakai, 2013; Majorano, Vihman, & DePaolis, 2014). In the prelinguistic period, then, experience with vocal production can be taken to create a sample of known patterns to which novel word forms can be mapped. This in turn leads to more robust representations of such forms as they are fleetingly but repeatedly heard in ambient speech (see studies of 'the production effect in memory' in both adults and older children, e.g., MacLeod, Gopie, Hourihan, Neary, & Ozubko, 2010; Icht & Mama, 2015; Zamuner, Morin-Lessard, Strahm, & Page, 2016).

Thus, in this approach, accentual prominence initially attracts and shapes infant attention to speech (Teixidó, François, Bosch, & Männel, this volume) and vocal production further supports retention of certain repeatedly heard patterns. In addition, the relatively infrequent but readily recognized words produced in isolation in the everyday routines of infant life have been found to play a disproportionate role in the choice of first words to produce (Brent & Siskind, 2001; Vihman, in press). All of these factors are critical elements in the construction of phonological memory, or the ability to retain novel phonological strings (Keren-Portnoy, Vihman, DePaolis, Whitaker, & Williams, 2010), a key element in word learning (Gathercole & Baddeley, 1989). It is thus likely that the gradual process of construction of memory for speech, based on vocal practice, is one reason for the delay between first adult-like vocal production, almost always observed in the middle of the first year, and first word use, typically identified only toward the end of the first year at the earliest.

The term 'prosodic structure' can be used in a number of different ways. It may refer to the accentual pattern of words (trochaic, or strong-weak, vs. iambic, or weak-strong stress on disyllables, for example; see Bhatara, Boll-Avetisyan, Höhle, & Nazzi, this volume), the overall rhythmic pattern of a language (e.g., Mehler et al., 1988; Mehler, Dupoux, Nazzi, & Dehaene-Lambertz, 1996; Ramus, Nespor, & Mehler, 1999) or, more specifically, the units of the prosodic hierarchy proposed by Selkirk (Selkirk, 1978, 1980; see also Nespor & Vogel, 1986). 'Prosodic structure' is also used to refer to units below the phonological or prosodic word level, the lowest level in the prosodic hierarchy (Fikkert, 1994; Kehoe, 2013). This chapter will focus on word-shapes consisting of syllable sequences of different lengths and organization, such as the universally occurring <CV>, disyllabic <CVCV>, onsetless <VCV> or the final-consonant-bearing structures <CVC>, <CVCVC>. In this usage, 'prosodic structure' is similar to the skeleton of autosegmental phonology (Goldsmith, 1990).

The prosodic structures that will be the focus of interest here are based on early word forms observable in data from children learning one of four languages.

These structures can be taken to reflect infant experience with the most frequent accentual patterns in the input, their own production of adult-like vocalizations and their growing ability to map input word forms onto their existing vocal patterns. The analyses reported here will make it possible to identify similarities and differences in the types of early structures produced. After the findings have been presented and illustrated, the likely sources of children's early prosodic structures will be more broadly considered, including aspects of input speech, constraints on vocal production and the interaction of the two.

A developmental perspective on early prosodic structures

The 'accuracy' of children's first words, first noted by Ferguson and Farwell (1975), is largely due to children's (unconscious) 'selection' of the word targets that are the most accessible, given their limited production skills. As illustrated with the first few words identified for 48 children learning 10 languages (Menn & Vihman, 2011, App. 1), target words are seldom more than one or two syllables in length and word length is usually matched by the child forms. The target words are limited to a single supraglottal consonant type per word about half the time and to stops, nasals, glottals and glides somewhat more often; this also supports relatively accurate production, as these are the consonants that are well practiced in babbling. Finally, although word-final consonants are generally omitted and heterogeneous clusters reduced, diphthongs are often matched. This is the starting point of word production: Children are sensitive to target forms for which they may have a rough vocal match in repertoire and their first identifiable words are similarly constrained cross-linguistically.

Children's production skills advance only quite slowly; what changes with the onset of word production is child familiarity, through on-going experience with production and the acoustic output that accompanies it, with a growing number of phonological forms related to often-heard input forms. For the purposes of understanding the formation of preferred prosodic structures a key point is that the timing of this development is highly variable, differing from one child to the next, but that persistent production of words of a particular prosodic structure is typically seen only after a period of early word use. An initial data sample, leading to a few well-developed word-production routines, needs to develop first. This accounts for the typical lack of systematicity, or relatively rare re-use of particular word shapes, syllables or features, in the first few recorded words (Menn & Vihman, 2011).

Individual phonological templates have been defined as idiosyncratic patterns or prosodic structures used by a given child, sometimes based on 'overselecting' of a particular structure compared to other children learning the same language but more often based on adaptation, or modification of target word forms to fit the particular structure (Vihman, 2014). Vihman and Velleman (2000) introduced the terms 'select' and 'adapt' to characterize words that differ in this way in their relationship to the child's template; the terms are applied here to prosodic structures that are not further segmentally specified.

Data analysis

To gain an idea of cross-linguistic similarity or differences, comparably complete data are taken for analysis from one recording session for each of 44 children learning one of four languages: American or British English, French, Italian and Finnish. The data derive from the end of the single-word period, when children can be observed to be producing a small set of preferred prosodic structures, some of which may come to serve as templates, or more fully specified attractors for new word forms. Actual word forms produced by children in each group are amply illustrated, to demonstrate how children pursue individual developmental paths, each of them in their own way mapping more or less well-suited adult target forms to the prosodic structures that the child has mastered.

All of the data analysed were recorded while the children were in interaction with one or more family members in the presence of a familiar investigator, for 20–30 minutes of unstructured play. Table 1 lists the number of children in each group, their age range (year; month) and the mean number of different target words produced in the session analysed. Imitations as well as spontaneous word forms and variant prosodic shapes that fit distinct structures were included in the analysis to assure sufficient material, at as early an age as possible, to identify favoured prosodic structures and allow for cross-linguistic comparative analysis.

Table 1. Language groups, ages and mean number of words attempted ('target words') Data sources: Boysson-Bardies & Vihman, 1991; Kunnari, 2000; Veneziano & Sinclair, 2000; Vihman, 1993, 1996, 2010, 2014; Vihman, Keren-Portnoy, Khattab, & Wauquier, 2011, July; Vihman & Kunnari, 2006; Vihman & Majorano, 2017; Vihman & Velleman, 1989; Vihman, Velleman, & McCune, 1994b; Wauquier & Yamaguchi, 2013.

Language group	Number of children	Age-range	Mean target words
English	US 6, UK 4	US 1;4–1;5, UK 1;5–1;11	US 35, UK 41
Finnish	6	1;3–1;11	30
French	13	1;2–1;10	35
Italian	15	(1;3, 1;9)	41

For all but the Italian data the first recording in which about 25 different word types were produced spontaneously was analysed. The Italian data are drawn from a study of 30 children recorded twice, at 1 year and 3 months, and at 1 year and 9 months (Vihman & Majorano, 2017); only the 15 children who produced, at either age, between 30 and 60 words in the recorded sessions are included here, in order to provide a comparable level of prosodic variants and structural diversity to that seen in the other groups.

For each child, more than a single variant word shape (token) of any word type was included in the prosodic structure analysis only if those variants fall into distinct structures (see Vihman, 2016). To focus in on the main prosodic structures used in each group, only those that accounted for 20% or more of a child's variants are listed. None of the structures are specified as to segmental content for any slot, distinguishing the 'generic' prosodic structures from the more specified 'templates' described elsewhere. (For example, a melodic template may be specified to include constraints on place of articulation – e.g., C1 is labial, C2 is coronal or velar, Macken, 1979, or a particular segment may be specified, like medial yod, in Priestly, 1977, or medial [l], in Vihman, 1993.)

English

Table 2 provides an overview of the structures most used by the 10 children acquiring English.

Table 2. Prosodic structures in English.

Structure	Number of children with over 20% use of structure	Mean proportion of all variants (and range)
CV	6	.31 (.23–.44)
CV1V2	2	.24 (.21–.26)
CV and CVV combined*	3	.31 (.26–.35)
CVC, CVʔ	5	.34 (.23–.44)
C_1VC_1V	3	.30 (.28–.31)
$C_1VC_2V(C)$	3	.25 (.22–.26)
C_1VC_2VC	1	.41

* 'Combined' figures refer to children whose use reaches 20% only when two structures are combined that, for other children, reach that level independently; cf. also Table 14, Finnish.

The hallmark of the English profile is monosyllabic production, often with a coda. Note that just one child makes criterial use (i.e., at least 20% of the structures produced by that child) of the more challenging structure, disyllabic variants with coda (< CVCVC >).

English: Selected and adapted variants

By the end of the single word point, when the first word combinations can be observed, most children are beginning to fit more difficult target words into their favored structures. Tables 3 and 4 provide examples to illustrate the effect of a child's preferred production patterns on their phonological advances, drawing on the children with the highest proportion of a given structure. Selected words are defined as accurate within the limits of the child's resources: coda omission is disregarded, for example, if the child is not yet producing codas. Adapted words reflect changes to fit the structure, such as truncation (Emily, *apple, button*) or metathesis (Timmy, *sun,* Sean, *Ian*). Note that some children preferentially select words of a particular structure but do not show adaptation (e.g., Geoff, for < CVC >).

Table 3. < CV(V) > Note that vowel length, which is not contrastive in English, is disregarded in allocating forms to this structure. (im. 'imitation', *N 'number of tokens of same form', if more than one)

	Select		**Adapt**
Emily (.53: 24/45)			
beads	pi (*2)	*all gone* (im.)	kɑ̃:
box	p^{h}ɑ:	*apple*	bɪ: (*2, im.)
more	mɔə	*button*	bɑ:
Timmy (.40: 14/43)			
ball	bæ::	*balloon*	bɛ::
car	kɑ, gɑ	*sun*	næ
Sean (.40: 14/43)			
blocks	bɑ (2)	*bug*	βɑɪ
blue	bwe:, bʊ^{h},	*dog*	dɑʊ, taʊ̥
mama	mɑ, mɑ: (3), mɔ:	*Ian* (im.)	nĩ
quack	ʔɑ::	*mouse*	mɑʊ̥, mɑ:ɨ

Table 4. < CVC >

	Select
Geoff (.44: 17/45)	
cars	k^{h}ɑ:ʃ
dog	gɒːg
pig	p^{h}iːk^{h}
sheep	p^{h}iːp^{h}
toys	t^{h}ɔʊːsː

French

Table 5 provides an overview of the most-used structures in French. Overall, the children produce a higher proportion of disyllabic than monosyllabic structures, the reverse of the English distribution. Furthermore, one of the two most-often used French structures is the onsetless disyllable (VCV), which is not found in English.

Table 5. Prosodic structures in French

Structure	Number of children with over 20%use of structure	Mean proportion of all variants (and range)
CV	10	.32 (.22–.54)
CVC	1	.26
VCV	8	.37 (.21–.64)
C_1VC_1V (including Reduplication)	9	.31 (.22–.42)
C_1VC_2V	2	.31 (.23–.38)

French: Selected and adapted variants

Tables 6 to 9 illustrate French structures. Table 6 again shows adaptations that include not only truncation but also metathesis (Camille, *musique,* Julien, *pique*). Whereas the English < CV(V) > adaptations are generally based on the initial (stressed) syllable of the target word, in French both the initial (Camille, *couvercle,* Julien *canard*) and final syllables (Camille, *attends, chercher*; Julien *araignée, encore, lapin*) provide a basis for the child forms.

Table 6. < CV >

	Select		Adapt
Camille (.54: 25/46)			
chat 'cat'	ʃa	*attends* 'wait!'	tã
main 'hand'	mɛ̃	*chercher* 'look for'	ʃe
plus '[no] more'	py	*couvercle* 'lid'	ku
		musique 'music'	ki
Julien (.50: 20/40)			
cou 'neck'	ku	*araignée* 'spider'	ne (*3)
plein 'full'	pɛ̃	*encore* 'again, more'	kɔ̃
thé 'tea'	te (*3), ke (*3)	*lapin* 'rabbit'	pɛ̃
trois 'three'	ka (*4)	*pique* 'stings'	ki (*8), ke, ti

Table 6. (*Continued*)

	Select		Adapt
Basile (.36: 21/59)			
bottes 'boots'	bo (*2)	*bisou* 'kiss'	bu
ça 'that'	ʒa (*10)	*livre* 'book'	lə (*3)
Adeline (.34: 13/38)			
lave 'wash'	la (*4)	*au revoir* 'goodbye'	bwa
moi 'me'	ma	*lave* 'wash'	kɔ

As can be seen in Table 7, Beryl, who makes especially heavy use of the < VCV > structure, also shows instances of metathesis (*crapeau*, perhaps also *nuage*).

Table 7. < VCV >

	Select		Adapt
Beryl (.64: 46/72)			
encore 'again, more'	ãkɔ	*l'éléphant* 'the elephant'	afɔ
étoile 'star'	ata	*crapeau* 'toad'	ako (*2)
hibou 'owl'	abu	*micro* 'microphone'	a χo (*13)
un coeur 'a heart'	akɔ	*nɥaʒ* 'cloud'	aça
Charles (.44: 14/32)			
allo 'hello'	al:ɔ	*c'est bon* 'it's good'	habõ
attends! 'wait'	atæ	*garcon* 'boy'	haʒœ (*3)
au revoir 'byebye'	avwa	*va pas* 'doesn't go'	apa (*3)
Adeline (.42: 16/38)			
au revoir 'goodbye'	ova, opa	*bisou* 'kiss'	atu
trois 'three'	ka	*chapeau* 'hat'	apo (*4)
brosse 'brush'	ə gʁɔ	*voiture* 'car'	atu (*2)
Vincent (.41: 19/46)			
allô 'hello (telephone)'	alɔ (*2)	*bravo*	avo (*4), awo
essaye 'try!'	eçɛ	*caché* 'hidden'	açe
ici 'here'	isi	*marche pas* 'doesn't work'	afpa (*6)
Gaël (.37: 18/49)			
écrit 'written'	eχri	*c'est ça* 'that's it'	esa (*11)
ici 'here'	isi	*chercher* 'to look for'	aʃe
ouvrir 'to open'	oχri. uvχri (*2)	grelot 'rattle, bell'	eχo

French children who use harmony generally have a large number of reduplicated forms as well (Table 8).

Table 8. Consonant harmony (and reduplication)

	Select		Adapt
Carole (.42: 16/38)			
Babar	baba (x3)	*chapeau* 'hat'	papo: (x3)
bébé 'baby'	be:be	*fromage* 'cheese'	məmæ
Mimi	mimi (2)	*lapin* 'rabbit'	pa:pa
Laurent (.41: 13/32)			
bébé 'baby'	bəbə	*chapeau* 'hat'	bobo (x5)
coucou 'peek-a-boo'	kuku	*chausettes/chaussures* 'socks/ shoes'	ʃoʃy
dodo 'sleep (BT)'	dodɔ	*gateau* 'cake'	tato (x3)

One French child makes use of a < C_1VC_2V > structure in over a third of his variants, with medial [l] in all 12 forms (see Vihman, 1993, who shows Laurent's attraction to medial [l] already in his first word forms).

Table 9. < C_1VC_2V >

	Select		Adapt
Laurent (.38: 12/32)			
ballon 'big ball'	palɔ̃	*(la) brosse* 'brush'	bəla
de l'eau 'some water'	dəlo	*(le) canard* 'duck'	kɔɭa
voilà 'here you are'	wal:a	*(le) chapeau* 'hat'	bolo (x2)

It is evident here that word forms with more than a single supraglottal consonant are especially challenging for the children and that different children deal differently with the difficulty. Certain aspects of French structure appear to affect these pre-grammatical word productions. First, phrase-final lengthening may draw attention away from the word-initial consonant (Vihman, Nakai, DePaolis, & Hallé, 2004), resulting in some children making (unconscious or implicit) recourse to < VCV >, which removes the problem. Second, the fact that French nouns are typically labelled with an article (*le chien, la brosse, un coeur, de l'eau*) may help account for the strong preference for disyllables over monosyllables in the French children's production (Wauquier & Yamaguchi, 2013). More specifically, the definite articles and pronouns in /lV…/ (*le* 'the, m.', *la* 'the, f.', *les* 'the, pl.', *lui* 'to him', *leur* 'their') may have shaped Laurent's attraction to medial [l], with the routine use of a fixed sequence < CVlV > providing another solution to the memory and planning issues posed by within-word consonant change.

Italian

Table 10 provides an overview of the most-used structures in Italian. The minimal < CV(V) > structure occurs only rarely. Italian content words are seldom monosyllabic; two of the three monosyllables produced by several children each are deictics (*qua* 'there', *qui* 'here').

Table 10. Prosodic structures in Italian. (C_o indicates optional consonant: i.e., forms with and without medial clusters are included)

Structure	Number of children with over 20% use of structure	Mean proportion of all variants (and range)
VCC_oV	7	.27 (.21–.36)
C_1VC_1V (including Reduplication)	8	.27 (.22–.36)
CVCCV	14	.29 (.22–.38)
C_1VC_2V(CV)	6	.26 (.21–.29)
Longer words	C.L.	.25 (13/51)

Another Italian characteristic is the wider distribution of variants in differing structures: No structures account, on average, for as much as one-third of the variants produced. This may be due to the fact that so many children often produce the geminate < CVCCV > structure that the analysis set it apart from the corresponding medial-singleton structure, < C_1VC_2V > . This latter structure, often with a within-word change of place of articulation, occurs rarely in English and French child forms, mainly as part of a more specified template. No such templatic specialization is seen in these cases in Italian.

Italian: Selected and adapted variants

As in French, many Italian children make use of the onsetless disyllabic pattern, omitting word-initial consonants that are well within their repertoire in other positions (Table 11).

Use of consonant harmony (Table 12) is seen in about half of the Italian children, with some forms showing full reduplication, usually in accurate replication

Table 11. < VCV > (m. 'masculine')

	Select		Adapt
G.A. (.36: 12/33)			
aqua 'water'	ak:a	*grazie* 'thank you'	at:ʃe (6)
ecco 'here it is'	ɛk:o (2)	*latte* 'milk'	at:e
occhio 'eye'	ok:o	*quello* 'that, m.'	ɛl:o (2)
M.G. (.30: 14/46)			
apri 'open (it)!'	api (5)	*basta* 'enough, stop'	at:a (3)
ecco 'here it is'	ɛk:o (12)	*grazie* 'thank you'	at:e (5)
uva 'grape'	uva (5)	*questo* 'this, m.'	et:o (2)

of the target. G.C.'s harmony structure includes about equal numbers of simple harmony and reduplication, with far more selected than adapted variants and none suggesting adaptation to create a reduplicative pattern.

Table 12. Consonant harmony (C1VC1V)

	Select		Adapt
G.C. (.33: 11/33)			
mamma 'mama'	mam:a (3)	*latte* 'milk'	tat:e (2)
nanna 'to sleep (BT)'	nan:a		

Table 13 illustrates the large number of Italian children's variants with non-harmonic two and three-syllable words with medial geminate. These words all involve change of place and/or manner of articulation. Note that adaptation here involves truncation – and many children, like A.P., simply 'select' these words, producing them accurately.

Table 13. Non-harmonic multisyllabic words with medial geminates

	Select		Adapt
A.B. (.38: 15/40)			
bolle 'bubble'	bol:e	*aspetta* 'wait!'	pet:a (2)
grazie 'thank you'	gat:ʃe (5)	*bistecca* 'steak'	tek:a (2)
scarpe 'shoe'	kap:e	*piselli* 'peas'	pɛl:i
A.F. (.36: 16/45)			
bello 'beautiful, nice, m.'	bɛl:o	*padella* 'pan'	pɛl:a

Table 13. (*Continued*)

	Select		Adapt
grosso 'fat, m.'	gɔs:o	*bottiglia* 'bottle'	bot:ja
metti 'put'	met:i	*coperchio* 'lid, cover'	pek:o
A.D.1 (.35: 15/43)			
bimba 'baby, f.'	bimba (3)	*capretta* 'little goat'	pet:a
forte 'strong'	fot:e (3)	*Emilio*	miljo
freddo 'cold'	fed:o		
A.P. (.33: 11/34)			
basta	bat:a		
brucia	buˈtʃa (3)		

Finnish

Table 14 provides an overview of the most-used structures of the six Finnish children.

Table 14. Prosodic structures in Finnish.

Structure	Number of children with over 20% use of structure	Mean proportion of all variants (and range)
CVV	1	.32
VCV	4	.38 (.29–.50)
Reduplication	1	.36
Consonant harmony	4	.29 (.21–.38)
Reduplication and consonant harmony combined	2	.23 (.22–.24)
C_1VC_2V	1	.25 (7/28)

Finnish, like Italian, has contrast between medial singletons and geminates, which are influential in shaping children's early word patterns in languages that have them (see, for example, Khattab & Al-Tamimi, 2013, on Arabic; Vihman & Croft, 2007, on Hindi). However, the Finnish children's most-used prosodic structures are quite unlike those of the Italians.

First, monosyllabic content words are more common in the Finnish core vocabulary than in Italian (making up 13% of the target word types attempted by any of the six children). An important structural difference is the presence of long vowels in Finnish and the 'minimal word' constraint, such that < CV > does not occur in content words (see also Demuth, this volume).

Finnish has many long words, mostly due to the agglutinative morphology (several suffixes, each with separate meaning, can be strung together at the end of a noun, for example; see Savinainen-Makkonen, 2000a), but the basic uninflected content words are seldom longer than two syllables. Few long words were attempted (Eliisa *paperi* 'paper' [pape], Ilari *uppista* 'upsy-daisy' [up:i], Matti, *apelsiini* 'orange' [apetæ] are among the exceptions) and even fewer produced. And although word-medial geminates are common in Finnish as well as in Italian, Ilari, the one child who used non-harmonic < CVC(C)V >, targeted and accurately produced more words with singletons than with geminates in this structure.

Finnish: Selected and adapted variants

Tables 15–17 illustrate the prosodic structures used for over 33% of any child's variant forms in Finnish. Here it can be seen that Finnish structure, like Italian, presents relatively few challenges to the child. For < VCV > initial consonant omission, unknown as a systematic process in English (Savinainen-Makkonen, 2000b), is the main adaptation.

Table 15. < VCV > sP: partitive singular

	Select		Adapt
Atte (.50: 16/32)			
äiti 'mother'	æit:i (*7)	*kala* 'fish'	ala
isi 'father'	içi (*6), aiçi	*loppu* 'end, all done'	op:u (*4)
ukko 'old man'	uk:o (*3), auk:o	*pallo*	al:o (*2)
Matti (.39: 19/49)			
ankka 'duck'	ak:a	*jalka* 'foot'	ak:a
anna 'give'	æn:æ	*katso* 'look'	ato (*7)
omppu 'apple'	ɔp:u (*3)	*vettä* 'water, sP'	et:æ

Reduplication is more common in Finnish than in the other languages, but it nevertheless accounts for over 20% of the variants of only one child, Mira (Table 16); the closely related harmony forms are shown in Table 17.

Table 16. Reduplication (s3 'third person singular')

	Select		Adapt
Mira (.36: 12/33)			
mummu 'grandma'	mum:u	jalka 'foot'	lala
pappa 'grandpa'	pap:a	kukka 'flower'	kak^{h}a
paapaa 'sleeping' (BT)	pa:pa:	nalle 'teddybear'	ʎaʎ:a

Table 17. Consonant harmony

	Select		Adapt
Eliisa (.38: 13/34)			
kukka 'flower'	kuk:a	*nuke* 'doll'	kuk:e
pupu 'bunny'	pupo	*pallo* 'ball'	pap:u
tyttö 'girl'	tyt:ö	*vettä* 'water, sP'	tit:æ
Mira (.33: 11/33)			
nenä 'nose'	nenæ	*lintu* 'bird'	tito
tonttu 'elf, brownie'	təto (*2)	*setä* 'uncle	tetæ
tyttö 'girl'	tyt:o (*2)	*tyyny* 'pillow'	ny:ny

Overview of findings

An overall picture of the distribution of prosodic structures in the four languages is given in Table 18, with numbers of children making criterial use of each structure in each language group. (Recall that 'criterial use' means reaching at least 20% of a child's structures.) With the exception of Italian, where some children are already beginning to accurately produce words of more than two syllables at this lexical level, one- and two-syllable forms account for all of the children's often-used prosodic structures.

Table 18. Number of children producing over 20% of all variants in each structure. Shaded cells highlight use by half or more of the children in the group.

	English	French	Italian	Finnish
Total number of children	**10**	**13**	**15**	**6**
CV and/or CVV	9	10	0	1
CVC	5	1	0	0
VCV	0	8	7	4
Reduplication and/or Consonant Harmony	3	9	8	6
CVCCV	0	0	14	0
C_1VC_2V	4	2	7	1
CVCVC	1	0	0	0
longer words	0	0	1	0

Much of the 'adaptation' seen in the data analysed here relates to the challenge of producing (representing, planning and articulating) two different supraglottal consonants in a single word form. Table 18 shows that both English and Italian groups include more than one child who produces disyllables without harmony to criterion level. In addition, one French child, Laurent, produces 38% such forms, with a more specified structure – < CVlV > – accounting for all 12 such words. Such 'overuse' of a structure has been characterized as a phonological template (Vihman, 1993).

On closer examination, phonological templates can also account for at least two of the cases of high use of the < C_1VC_2V > structure in English. Molly over-selects < CVC > target words (Vihman & Velleman, 1989), but adds a final schwa following her coda nasals – which effectively creates open disyllables (e.g., *bang* [pɑn:ɑ], *down* [t'ɑn:ə] and even, with metathesis, *Nicky* [ɪn:i]). Alice shows a marked preference for palatal patterns, consistently producing disyllables with a medial palatal stop or nasal or yod (e.g., *flower* [p'ɑ:ji, p'aɖi], *lady* [lɛji], *shiny* [tɑ:ji, daji]: Vihman et al., 1994). The one UK English child with over 20% < C_1VC_2V > disyllables has a glottal or glide as one of the two consonants in all but one word in this structure.

In contrast with the English and French groups, disyllables (or longer words) with differing supraglottal consonants are quite common in Italian, even after < CVCCV >, the most used Italian structure, has been set aside. In general, the Italian children show great facility in producing differing consonants within a single word form, whether in disyllables or longer words and whether the word includes one or more geminates or not. This is evident from their lack of heavy concentration on a single structure. Although this might be supposed to relate to the fact that, on average, these Italian children are a little older than the other groups, the structures used here are quite similar to those reported for four Italian children in an earlier study of 11 children followed longitudinally to age two (Keren-Portnoy, Majorano, & Vihman, 2009).

Prosodic structures: Production and representation

Both similarities and differences are apparent in the prosodic structures that the children in different language groups use the most. The similarities can be broadly summarized as (i) a constraint on length, with words of more than two syllables occurring as a dominant structure only in Italian, in a single child; (ii) a tendency to produce open syllables, with only the English-speaking children showing a substantial number of word forms with codas, almost exclusively in monosyllables; and (iii) a tendency to produce words with only a single place or manner of articulation found in many word forms in all four languages. This results in the well-known preference for consonant harmony and reduplication as well as in use of the < VCV > structure, seen in three of the four languages.

The open syllables are a direct reflection of the canonical syllables that constitute the earliest adult-like babbling forms; similarly, both the constraint on word length and the relatively low level of within-word consonant variegation have been identified, through earlier studies of individual children seen longitudinally, as reflecting continuity between babble and word use, both in concurrent analyses of the course of development in words and babble (Vihman, Macken, Miller, Simmons, & Miller, 1985) and in comparisons of words with earlier babble (e.g., Keren-Portnoy et al., 2009; Stoel-Gammon & Cooper, 1984).

It is less straightforward to account for the differences across the language groups. They must in some sense relate to the differences in the adult languages, especially accentual and rhythmic differences. The absence of the < VCV > pattern in English could be ascribed to the preponderance of trochaic content words (Cutler & Carter, 1987), for example, but Finnish is exclusively trochaic and Italian is dominantly so. For both Finnish and Italian, the common presence of medial geminates appears to draw infant attention away from the onset consonant (Savinainen-Makkonen, 2000b; Vihman & Croft, 2007; see Vihman & Majorano, 2017, for experimental evidence from Italian). For French, on the other hand, phrase-final lengthening is a likely contributor to the heavy use of the < VCV > structure, as noted earlier; the structure is also seen in languages more conventionally defined as iambic than French, such as Hebrew (Keren-Portnoy & Segal, 2016).

Three aspects of English structure appear to be particularly important in shaping children's word forms, which differ so strikingly here from those of children acquiring the other three languages. First, monosyllables dominate the input in English (Vihman et al., 1994). Second, most English words have codas. The fact that disyllabic structures with codas are rare even in English is most likely due to the complexity or inherent difficulty of this structure. And, finally, diphthongs are widely (and accurately) produced in early words in English (Kehoe & Stoel-Gammon, 2001) and German (Kehoe & Lleó, 2003). Diphthongs do not appear to constitute a difficulty for production; in the languages in which they occur, children often make use of them from their very first words. On the other hand, the only consonant clusters that consistently occur early in the languages in which they obtain are geminates, which appear to be not only salient but easily produced, given children's slow articulation (Vihman & Velleman, 2000).

Conclusion

Prosodic Phonology (Nespor & Vogel, 1986) has been seen as providing a unified framework for characterizing prosodic development cross-linguistically (Kehoe, this volume). Alternatively, or perhaps complementarily, metrics have been proposed as a way of creating a gradient of rhythmic types that might account for differences in prosodic development (Grabe & Low, 2002; Ramus et al., 1999; Payne, Vanrell, Prieto, Astruc, & Post, 2012; Vihman, Nakai, & DePaolis, 2006; see Post & Payne, this volume). However, our emphasis here on individual differences within broadly similar typological effects by language group leads us to suggest that any one way of characterizing stages in prosodic development overall or the several languages in themselves is unlikely to suffice to account for the different ways in which the ambient language shapes the prosodic structures of individual children. That is because the individual differences between children reflect the many possible, sometimes idiosyncratic ways in which children respond to their experience of listening to and beginning to produce speech.

In short, the claim here is that innate knowledge of universal principles of prosodic structure need not be invoked to account for the structures found to predominate in the word forms analyzed here. Production constraints, a concomitant of the slow pace of advances in the neurophysiological control of vocal gestures, play an important role, as does exposure to frequently occurring structures in input speech. But perhaps the most crucial factor, demonstrated by the substantial role of adaptation in the children's use of prosodic structures, is the individual child's experience with vocal practice and the effect of that practice on memory for word forms. The adaptation of less congenial adult forms to the structures in which a particular child had begun to 'specialize' can be understood to reflect the child's filtering of what they hear through the patterns they have practiced the most. This filter manifests the effect of phonological memory, constructed through practice and serving as an essential framework for mapping new words to production routines. This suggests that the filtering of what children hear through what they are easily able to produce – with notable differences from one child to the next – is a critical aspect of prosodic development.

References

Boysson-Bardies, B.de, & Vihman, M. M. (1991). Adaptation to language: Evidence from babbling and first words in four languages. *Language*, 67, 297–319. doi: 10.1353/lan.1991.0045

Brent, M. R., & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, 81, B33–B44. doi: 10.1016/S0010-0277(01)00122-6

Bybee, J. (2001). *Phonology and language use*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511612886

Bybee, J. (2010). *Language, usage and cognition*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511750526

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language*, 2, 133–142. doi: 10.1016/0885-2308(87)90004-0

Davis, B. L., & MacNeilage, P. F. (2000). An embodiment perspective on the acquisition of speech perception. *Phonetica*, 57, 229–241. doi: 10.1159/000028476

DePaolis, R. A., Keren-Portnoy, T., & Vihman, M. M. (2016). Making sense of infant familiarity and novelty responses to words at lexical onset. *Frontiers in Psychology*, 7, 715.
DePaolis, R., Vihman, M. M., & Keren-Portnoy, T. (2011). Do production patterns influence the processing of speech in prelinguistic infants? *Infant Behavior and Development*, 34, 590–601. doi: 10.1016/j.infbeh.2011.06.005
DePaolis, R. A., Vihman, M. M., & Nakai, S. (2013). The influence of babbling patterns on the processing of speech. *Infant Behavior and Development*, 36, 642–649. doi: 10.1016/j.infbeh.2013.06.007
Ferguson, C. A., & Farwell, C. B. (1975). Words and sounds in early language acquisition. *Language*, 51, 419–439. doi: 10.2307/412864
Fernald, A. (1991). Prosody in speech to children: Prelinguistic and linguistic functions. *Annals of Child Development*, 8, 43–80.
Fikkert, P. (1994). *On the acquisition of prosodic structure*. Amsterdam: Holland Institute of Generative Linguistics.
Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language*, 28, 200–213. doi: 10.1016/0749-596X(89)90044-2
Goldsmith, J. A. (1990). *Autosegmental and metrical phonology*. Oxford: Blackwell.
Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven, & N. Warner (Eds.), *Papers in laboratory phonology VII* (pp. 515–546). Cambridge: Cambridge University Press. doi: 10.1515/9783110197105.515
Hall, K. C., Hume, E., Jaeger, T. F., & Wedel, A. (2016). *The message shapes phonology*. Paper presented at the symposium on The role of predictability in shaping human language sound patterns, Sydney, Australia.
Icht, M., & Mama, Y. (2015). The production effect in memory: A prominent mnemonic in children. *Journal of Child Language*, 42, 1102–1124. doi: 10.1017/S0305000914000713
Jusczyk, P. W. (1993). From general to language-specific capacities: The WRAPSA model of how speech perception develops. *Journal of Phonetics*, 21, 3–28.
Jusczyk, P. W. (1997). *The discovery of spoken language*. Cambridge, MA: The MIT Press.
Kehoe, M. (2013). *The development of prosody and prosodic structure*. New York, NY: Nova Science Publishers.
Kehoe, M., & Stoel-Gammon, C. (2001). Development of syllable structure in English-speaking children with particular reference to rhymes. *Journal of Child Language*, 28, 393–432. doi: 10.1017/S030500090100469X
Kehoe, M., & Lleó, C. (2003). The acquisition of nuclei: A longitudinal analysis of phonological vowel length in three German-speaking children. *Journal of Child Language*, 30, 527–556. doi: 10.1017/S030500090300566X
Keren-Portnoy, T., Majorano, M., & Vihman, M. M. (2009). From phonetics to phonology: The emergence of first words in Italian. *Journal of Child Language*, 36, 235–267. doi: 10.1017/S0305000908008933
Keren-Portnoy, T., & Segal, O. (2016). Phonological development in Israeli-Hebrew-learning infants and toddlers: Perception and perception. In R. A. Berman (Ed.), *Acquisition and development of Hebrew* (pp. 69–94). Amsterdam: John Benjamins.
Keren-Portnoy, T., Vihman, M. M., DePaolis, R., Whitaker, C., & Williams, N. A. (2010). The role of vocal practice in constructing phonological working memory. *Journal of Speech, Language, and Hearing Research*, 53, 1280–1293. doi: 10.1044/1092-4388(2009/09-0003)

Khattab, G., & Al-Tamimi, J. (2013). Early phonological patterns in Lebanese Arabic. In M. M. Vihman, & T. Keren-Portnoy (Eds.), *The emergence of phonology: Whole word approaches, cross-linguistic evidence* (pp. 374–414). Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511980503.018

Kuhl, P. K. (1987). Perception of speech and sound in early infancy. In P. Salapatek, & L. Cohen (Eds.), *Handbook of infant perception* (Vol. 2, pp. 275–381). New York, NY: Academic Press.

Kunnari, S. (2000). *Characteristics of early lexical and phonological development in children acquiring Finnish*. Unpublished PhD dissertation. University of Oulu.

Macken, M. A. (1979). Developmental reorganization of phonology: A hierarchy of basic units of acquisition. *Lingua*, 49, 11–49. doi: 10.1016/0024-3841(79)90073-1

MacLeod, C., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, K. D. (2010). The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 671–685.

Majorano, M., Vihman, M. M., & DePaolis, R. A. (2014). The relationship between infants' production experience and their processing of speech. *Language Learning and Development*, 10, 179–204. doi: 10.1080/15475441.2013.829740

Mehler, J., Dupoux, E., Nazzi, T., & Dehaene-Lambertz, G. (1996). Coping with linguistic diversity: The infant's viewpoint. In J. Morgan, & K. Demuth (Eds.), *From signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 101–116). Mahwah, NJ: Lawrence Erlbaum Associates.

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. doi: 10.1016/0010-0277(88)90035-2

Menn, L., & Vihman, M. M. (2011). Features in child phonology: Inherent, emergent, or artefacts of analysis? In N. Clements, & R. Ridouane (Eds.), *Where do phonological features come from?* (pp. 261–301). Amsterdam: John Benjamins. doi: 10.1075/lfab.6.10men

Munson, B., Edwards, J., & Beckman, M. E. (2012). Phonological representation in language acquisition: Climbing the ladder of abstraction. In A. C. Cohn, C. Fougeron, & M. K. Huffman (Eds.), *Oxford handbook of laboratory phonology* (pp. 288–309). Oxford: Oxford University Press.

Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris.

Oller, D. K. (1980). The emergence of the sounds of speech in infancy. In G. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology*, Vol. I: Production (pp. 93–112). New York, NY: Academic Press.

Oller, D. K., & Eilers, R. E. (1988). The role of audition in infant babbling. *Child Development*, 59, 441–449. doi: 10.2307/1130323

Payne, E., Post, B., Astruc, L., Prieto, P., & Vanrell, M. M. (2012). Measuring child rhythm. *Language and Speech*, 55, 203–229. doi: 10.1177/0023830911417687

Pierrehumbert, J. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech*, 46, 115–154. doi: 10.1177/00238309030460020501

Priestly, T. M. S. (1977). One idiosyncratic strategy in the aquisition of phonology. *Journal of Child Language*, 4, 45–66. doi: 10.1017/S0305000900000477

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73, 265–292. doi: 10.1016/S0010-0277(99)00058-X

Savinainen-Makkonen, T. (2000a). Learning long words: A typological perspective. *Language and Speech*, 43, 205–225. doi: 10.1177/00238309000430020401

Savinainen-Makkonen, T. (2000b). Word initial consonant omissions: A developmental process in children learning Finnish. *First Language*, 20, 161–185. doi: 10.1177/014272370002005903
Selkirk, E. (1978). On prosodic structure and its relation to syntactic structure. In T. Fretheim (Ed.), *Nordic prosody II* (pp. 111–140). Trondheim: Tapir.
Selkirk, E. (1980). The role of prosodic categories in English word stress. *Linguistic Inquiry*, 11, 563–605.
Selkirk, E. (1996). The prosodic structure of function words. In J. Morgan, & K. Demuth (Eds.), *From signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 187–213). Mahwah, NJ: Lawrence Erlbaum Associates.
Stoel-Gammon, C., & Cooper, J. A. (1984). Patterns of early lexical and phonological development. *Journal of Child Language*, 11, 247–271. doi: 10.1017/S0305000900005766
Veneziano, E., & Sinclair, H. (2000). The changing status of 'filler syllables' on the way to grammatical morphemes. *Journal of Child Language*, 27, 461–500. doi: 10.1017/S030500090000427X
Vihman, M. M. (1993). Variable paths to early word production. *Journal of Phonetics*, 21, 61–82.
Vihman, M. M. (1996). *Phonological development: The origins of language in the child*. Oxford: Basil Blackwell.
Vihman, M. M. (2010). Phonological templates in early words: A cross-linguistic study. In C. Fougeron, B. Kühnert, M. D'Imperio, & N. Vallée (Eds.), *Laboratory phonology 10* (pp. 261–284). Berlin: Mouton de Gruyter.
Vihman, M. M. (2014). *Phonological development: The first two years* (2nd ed.). Malden, MA: Wiley–Blackwell.
Vihman, M. M. (2016). Prosodic structures and templates in bilingual phonological development. *Bilingualism: Language and Cognition*, 19, 69–88. doi: 10.1017/S1366728914000790
Vihman, M. M. (in press). First word learning. In S-A. Rueschmeyer, & G. Gaskell (Eds.), *Oxford handbook of psycholinguistics*. Oxford: Oxford University Press.
Vihman, M. M., & Croft, W. (2007). Phonological development: Toward a 'radical' templatic phonology. *Linguistics*, 45, 683–725. doi: 10.1515/LING.2007.021
Vihman, M. M., Kay, E., Boysson-Bardies, B., de, Durand, C., & Sundberg, U. (1994). External sources of individual differences? A cross-linguistic analysis of the phonetics of mothers' speech to one-year-old children. *Developmental Psychology*, 30, 652–663. doi: 10.1037/0012-1649.30.5.651
Vihman, M. M., & Keren-Portnoy, T. (Eds.) (2013). *The emergence of phonology: Whole word approaches, cross-linguistic evidence*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511980503
Vihman, M. M., Keren-Portnoy, T., Khattab, G., & Wauquier, S. (2011, July). *Psychological significance and typology of templates in early word production*. Paper presented at International Association for the Study of Child Language, Montreal, Canada.
Vihman, M. M., & Kunnari, S. (2006). The sources of phonological knowledge: A cross-linguistic perspective. *Recherches Linguistiques de Vincennes*, 35, 133–164. doi: 10.4000/rlv.1467
Vihman, M. M., Macken, M. A., Miller, R., Simmons, H., & Miller, J. (1985). From babbling to speech: A reassessment of the continuity issue. *Language*, 61, 397–445. doi: 10.2307/414151
Vihman, M. M., & Majorano, M. (2017). The role of geminates in infants' early words and word-form recognition. *Journal of Child Language*, 44, 158–184.
Vihman, M. M., Nakai, S., & DePaolis, R. A. (2006). Getting the rhythm right: A cross-linguistic study of segmental duration in babbling and first words. In L. Goldstein, D. Whalen, & C. Best (Eds.), *Laboratory Phonology 8* (pp. 341–366). Berlin: Mouton de Gruyter.

Vihman, M. M., Nakai, S., DePaolis, R. A., & Hallé, P. (2004). The role of accentual pattern in early lexical representation. *Journal of Memory and Language*, 50, 336–353. doi: 10.1016/j.jml.2003.11.004

Vihman, M. M., & Velleman, S. (1989). Phonological reorganization: A case study. *Language and Speech*, 32, 149–170.

Vihman, M. M., & Velleman, S. L. (2000). Phonetics and the origins of phonology. In N. Burton-Roberts, P. Carr & G. Docherty (Eds.), *Phonological knowledge: Its nature and status* (pp. 305–339). Oxford: Oxford University Press.

Vihman, M. M., Velleman, S. L. & McCune, L. (1994). How abstract is child phonology? Towards an integration of linguistic and psychological approaches. In M. Yavas (Ed.), *First and second language phonology* (pp. 9–44). San Diego, CA: Singular.

Wauquier, S., & Yamaguchi, N. (2013). Templates in French. In M. M. Vihman & T. Keren-Portnoy (Eds.), *The emergence of phonology: Whole-word approaches, cross-linguistic evidence* (pp. 317–342). Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511980503.015

Zamuner, T. S., Morin-Lessard, E., Strahm, S., & Page, M. P. A. (2016). Spoken word recognition of novel words, either produced or only heard during learning. *Journal of Memory and Language*, 89, 55–67. doi: 10.1016/j.jml.2003.11.004

CHAPTER 11

Understanding the development of prosodic words

The role of the lexicon

Katherine Demuth
Macquarie University

Children's early speech productions are not entirely adult-like, with syllables and morphemes often missing from early utterances. However, these patterns of development also appear to be influenced by the language being learned. This chapter explores the role of the lexicon as a driving force in understanding both lower Prosodic Word (PW) and higher Phonological Phrase and Intonational Phrase (PP, IP) aspects of children's early speech. Using evidence from cross-linguistic studies, it shows how the prosodic shape and structure of the ambient lexicon influences the course of PW development and the implications this has for the acquisition of grammatical morphemes such as determiners. It concludes by exploring the implications of these findings for a developmental model of speech planning and production.

Introduction

Why do young children produce forms like 'nana' for target words like 'banana'? There is (at least some) agreement that such child forms are the result of early prosodic constraints on the lexicon. The goal of this chapter is to explore what we know about the organization of children's early prosodic words (PWs), how this plays a role in understanding the prosodic organization of children's early utterances, the theoretical and technical developments that made these findings possible, and the many issues that remain for further research. It will focus primarily on issues of syllable structure only as this relates to prosodic constituency, i.e. coda consonants, important for moraic structure and the creation of Feet. (The reader is referred to the following chapters in this volume for further information on prosody as it relates to early word segmentation and word learning (Thorson,

DOI 10.1075/tilar.23.11.dem

this volume; Teixidó, François, Bosch, & Männel, this volume) as well as the early production of words (Kehoe, this volume; Vihman, this volume).

Prosodic structure

In the following discussion it will become clear that children do not simply repeat what they hear, but rather construct their output using their own phonological and prosodic grammar. Although limitations on motor planning may play a role in the types of articulations that can be produced before the age of one (e.g., MacNeilage, 1980), where repeated sequences of CV tend to appear, it is less clear how this restricts the shape of children's early words. Even at 11 months, Lleó (2001) reports that children learning German sometimes produce CVC structures in their babbling forms (cf. Kern & Davis, 2009, for review). This suggests that, even at the babbling stage, the prosodic (phonotactic) structure of the ambient language begins to influence the structure that children's early word-forms will take. This theme of language-specific attunement in prosodic shape of the early lexicon (e.g., Feet) will reverberate throughout this chapter, suggesting that the frequency of ambient language lexical word shapes (in terms of position of stress and the number of syllables) significantly influences the early shape of children's PWs and utterances. Thus, although there are possible 'universal' aspects of early prosodic development (starting simple and small, with evidence of prosodically 'unmarked' early forms), these are rapidly re-shaped by language-specific aspects of the lexicon that the child most frequently hears and must eventually produce. These issues are discussed in more detail below.

Much less is known about how and when higher-level prosodic structures, such as Phonological Phrases (PPs) and or larger Intonational Phrases (IPs) are acquired (though see Teixidó, François, Bosch & Männel, this volume). This involves the lexicon, but also grammatical morphemes, including prosodic clitics such as articles, determiners, auxiliary verbs, pronouns, object clitics, etc. Much morphosyntactic research has suggested that such forms are acquired late due to the later development of syntactic movement operations (e.g., Wexler, 1994) and/or lack of understanding of the semantic contexts in which such forms are used (Hyams, 2007). However, there is also abundant crosslinguistic research now showing that some of the variable (inconsistent) within-speaker use of such forms is due to the prosodic contexts in which they appear (cf. Demuth, 2014, for review). The lexicon again plays an important role in determining the nature of these different prosodic contexts, and how grammatical morphemes come to be prosodically incorporated into both PWs, as well as higher-level prosodic structures. These issues are also briefly explored below, particularly with respect to the variable acquisition of articles and determiners.

Of course, none of the recent advancements in this research would have been possible without several developments, both theoretical and technical. First, research by phonologists such as Selkirk (1984) and Nespor and Vogel (1986) has laid the groundwork for exploring not only the prosodic structure of words, but also higher levels of prosodic structure. This can be captured in terms of the Prosodic Hierarchy, as illustrated in (1).

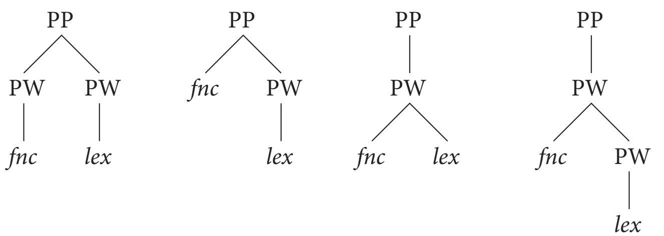
(1) Prosodic hierarchy

Utt	(Phonological Utterance)	I saw the bananas on the floor
IP	(Intonational Phrase)	I saw the bananas
PP	(Phonological Phrase)	the bananas
PW	(Phonological Word)	bananas
Ft	(Foot)	nanas
σ	(Syllable)	nas
μ	(Mora)	na

Further developments of the Prosodic Hierarchy by Selkirk (1996) provided a principled means for exploring within-language and crosslinguistic aspects of prosodic clitic incorporation. This resulted in a more detailed understanding of how function words can be incorporated into higher-level prosodic units, depending on the content of the function word, and how it is prosodified in a given language. This is shown in (2).

(2) Prosodic Words Prosodic Clitics

a. free clitic b. internal clitic c. affixal clitic



These theoretical developments provided a much-needed framework for understanding not only the nature of children's early PW shapes, but also their early incorporation of grammatical morphemes and production of higher-level PPs and IPs.

The second development has been that of constraint-based theories of phonology, such as Optimality Theory (OT) (Prince & Smolensky, 2004). This approach was quickly adopted for better understanding the nature of children's early word productions as unmarked forms (e.g., Spencer, 1986; Demuth, 1995; Gnanadesikan, 2004), where segments and syllables are omitted to adhere to optimal unmarked outputs. In the case of English and Dutch, these were disyllabic, trochaic feet. This constraint-based approach to early word structure has provided a much-needed framework for understanding how and why children produce the early PW forms they do. This will become clear in the following sections in both the discussion of the early forms of the lexicon, as well as the emergence of prosodic clitics.

The third development that has made much of this research possible has been technological developments that have facilitated data collection and analysis, making crosslinguistic audio/video corpora more widely available, much of it now posted in the public domain (e.g. CHILDES website: MacWhinney, 2000). This and PhonBank tools (Rose & MacWhinney, 2014) have greatly enhanced comparative crosslinguistic acquisition research, enabling us to test hypotheses about the role of language-specific prosodic structure on the course of PW development (cf., for example, the papers in Demuth, 2006). The fourth major development has been increased access to acoustic/phonetic analysis tools such as Praat (Boersma & Weenink, 2005). Having both corpora and experimental data with good quality acoustic recordings, and being able to listen to these to code for acoustic cues to feature contrasts (e.g. Stevens, 2002; Song, Demuth, & Shattuck-Hufnagel, 2012), as well as prosodic phenomena such as pitch, intonation, phrase-final lengthening, etc. (e.g. Snow, 1994; Frota, Cruz, Matos, & Vigário, 2016; Thorson, this volume; Teixidó, François, Bosch, & Männel, this volume), has enabled the growth of extremely important, more acoustically-grounded approaches to early prosodic development.

Given these converging areas of advance, it is now possible to explore the nature of children's prosodic development at all levels of the Prosodic Hierarchy, providing a more coherent picture of how and when children begin to produce more adult-like prosodic representations. Given better theory, data and tools for analysis, we can then begin to address issues of why children produce the early PW forms and utterance shapes they do, and better understand how this develops over time. Ultimately what is needed is a full developmental model of lexical access and higher-level speech planning processes. Although such models exist for adults (e.g. Levelt, Roelofs, & Meyer, 1999), many of the methods used to understand the nature of these processes are impractical for studies of child speech. It is therefore hoped that the review below will provide the first steps toward moving to such a goal, providing a better understanding of how and when children begin to produce PWs and utterances in a more adult-like way. We turn now to the growth of the lexicon, with a focus on early PW structure.

Markedness and the growth of prosodic words

Early PWs and feet

There has been much debate about how and why children's early PWs take the form that they do, often truncated, with syllables and segments missing. Given careful within-language analysis, and comparing this across languages, it has been proposed that the shape of children's early PWs is much influenced by markedness constraints (cf. Spencer, 1986; Demuth, 1995; Gnanadesikan, 2004). Thus, children's first words often take the form of CV or CVCV PWs with canonical CV syllables, and often contain reduplicative ($C_1V_1C_1V_1$) forms such as 'dada'. Consonant harmony is rare in adult phonologies of the world (Hansson, 2001; Rose & Walker, 2004). However, children's reduplicative forms appear to be the first step in producing actual PWs. These early PWs gradually become more differentiated, keeping the same vowels while beginning to vary the consonants ($C_1V_1C_2V_1$), or keeping the same consonants while varying the vowels ($C_1V_1C_1V_2$). This then evolves into more differentiated $C_1V_1C_2V_2$ forms. Examples of such forms come from a French-speaking child, who went through all these stages (though only a few stages are are captured in the corpus for a given word) (e.g., *chapeau* /ʃaˈpo/ > [poˈpo] > [taˈpo] …. > [ʃaˈpo] 'hat' (cf. Demuth & Johnson, 2003).

This gradual PW development may be due to increasing articulatory capabilities. Levelt (1995) notes that children's early words tend to exhibit strong coarticulation effects, with labial consonants typically co-occurring with rounded vowels (as in the French example above). As with 'consonant harmony', the strong co-articulatory effects found in early CV sequences minimizes the number of places of articulation the child has to target. Young children have small mouths and a big tongue (Kent, 1976), a challenge on fine-grained tongue movements. As the child's mouth gets bigger and articulatory control of the tongue more refined, the number of potential places of articulation within a PW increases. The reduplicated forms thus slowly give way to more target-like PWs, even though they may continue to take the form of disyllabic feet.

Thus, children's early words often remain 'unmarked' prosodic forms until around 1;6 years, with simple syllable and PW structure. This also means that longer target words are truncated (*banana* > *nana*), and CVC target words are produced either as CV, CV:, or epenthesized to CVCV (e.g., *dog* > *da* ~ *da:* ~ *daga*), at least in languages like English where 80% of words in child-directed speech (CDS) are monosyllabic (Roark & Demuth, 2000). Early disyllabic PWs typically exhibit stress on the first syllable of the word, as is common in English. In closely related Dutch it is clear that this has the status of a constraint, with examples where the stress is shifted to the first syllable of the output form, even when the target word has final stress (e.g., *balon* /ba'lon/ > ['bo:mi] 'ballon') (Fikkert, 1994; Demuth, 1995, 1996a).

In contrast, for languages where the CDS lexicon has many longer words (e.g., 28% of the Spanish CDS has words of 3 syllables or more (Roark & Demuth, 2000), children pass through this stage of PW truncation to a disyllabic foot much more quickly, and begin to produce trisyllabic words, even truncating 4 syllable words to 3 (e.g., *eskalera* > *kalera* 'stairs'; Gennari & Demuth, 1997). In a language like French, where the lexicon is evenly split between disyllabic and monosyllabic lexical forms, and the final syllable in the PP (or PW in citation form) is the most prominent (due to phrase-final 'stress'), children produce 'iambic' disyllabic feet in their early productions (e.g., *chapeau* > [po'po) 'hat') (Demuth & Johnson, 2003; Demuth & Tremblay, 2008). Thus, given the predominant prosodic patterns of the ambient lexicon (i.e., head direction of feet, number of syllables per word), children's early PW shapes begin to diverge crosslinguistically around 1;6–1;8 years, more closely approximating the most common target PW forms.

This is further seen in Catalan, where children omit the initial weak syllable in WS words for much longer than found in closely related Spanish, despite the fact that disyllabic iambic PWs are much more frequent in Catalan (cf. Prieto, 2006). Prieto (2006) suggests that this is due to the much higher frequency CVC structures in Catalan compared to Spanish, boosting the status of monosyllabic moraic trochees early in development.

Thus, around the age of 1;6 years, children's early reduplicative PW and disyllabic feet give way to better matching the more predominant PW shapes of the target language. There are several ways to understand the nature of this developmental change. From a constraint-based perspective, children are becoming more 'faithful' to the input, beginning to map more of the target lexical (segmental and syllabic/prosodic) information into their early PWs. At the speech planning and grammatical development level, children are also beginning to use larger prosodic structures, mapping early function words (such as determiners) into PWs and higher-level PPs and IPs. These issues are explored further below after a consideration of how coda consonants are acquired.

Syllable structure and coda consonants

We have mentioned above that children's early PWs take the form of unmarked prosodic structures, and this appears to hold at the level of the syllable as well. Thus, CV syllable structures are thought to predominate early on, with more complex syllable structures emerging around 1;6 years, at least in Germanic languages. This general pattern again seems to be driven by early sensitivity to language-specific syllable and PW structure. For example, Lleó (2001) reports earlier use of coda consonants in German-Spanish bilinguals than in their monolingual Spanish-learning peers (Lleó, 2003, 2006). This suggests that the high frequency of coda consonants in the ambient language may encourage early inclusion of coda consonants in children's early PWs.

Again, this development is interesting for several reasons. First, from a constraint-based perspective, it suggests that children are demoting the highly ranked constraint against coda consonants (NoCoda) earlier in languages where codas abound, allowing for coda consonants to appear earlier that in languages where coda consonants are less frequent. Although some children may initially epenthesize early CVC (e.g. big /ˈbɪg/ > [ˈbigə], *cup* /ˈkʌp / > [ˈkopi]), most appear to wait until they can produce coda consonants in a more adult-like way (cf. Demuth, Culbertson, & Alter, 2006).

Cross-linguistic evidence suggests that children's early use of coda consonants, and consonant clusters more generally, again reflect the tendencies of the ambient language. For example, Prieto (2006) shows that children learning Catalan, which has many final consonant clusters, begin to produce coda consonants earlier than in closely related Spanish. She suggests that the fact that these coda clusters occur in a stressed syllable further facilitates early production. Kirk and Demuth (2006) also show that simple codas are more likely to occur in stressed and/or final syllables in children's early speech, probably due to the increased duration of such syllables, providing sufficient time to produce all the coda consonants (cf. Hsieh, Leonard, & Swanson, 1999). This then also has an impact on when and how word-final inflectional morphemes are acquired (e.g., Song, Sundara, & Demuth, 2009). We revisit this issue below.

Levelt, Schiller, and Levelt (2000) show that Dutch-learning children begin to use consonant clusters either at the beginnings or ends of words, but not both at the same time. Levelt et al. (2000) show that the frequency of onset and coda clusters is the same, again suggesting that learners are acutely attuned to the frequency with which certain prosodic structures are used in the ambient language. Coordinating all such segments of a PW in a short amount of time (one syllable) can be an extremely complex undertaking, requiring very fine tongue movements – a challenge for early learners (Kent, 1976). Kirk and Demuth (2006) show that children are most likely to include production of segments within a PW if they occur at the same place of articulation, thus requiring fewer articulatory gestures.

But what happens before coda consonants are acquired? It was implied above that children then produce either reduplicated CVCV PWs, or merely CV. However, in many of the world's languages, CV PWs are not permitted. The following section discusses Minimal Word constraints, and how and when children come to realize that the target language has such constraints.

Minimal prosodic words and moras

Many languages in the world observe a Minimal Word constraint, where well-formed PWs must contain a minimal amount of phonological structure. Typically this is realized in terms of two moras of structure, or a Foot. In languages that have only monomoraic syllables (i.e. no long vowels/diphthongs and no coda consonants), Minimal Words must be composed of a disyllabic foot. Bantu languages such as Sesotho have such forms (Doke & Mofokeng, 1985). Thus, for monosyllabic verb stems produced in the imperative, where no other prosodic clitics are available, the verb is augmented with an epenthetic syllable to ensure PW well-formedness (e.g. *ja* 'eat' > *jaa*! ~ *eja*! 'eat!').

A language such as English requires that, in the case of a monosyllable, there be at least two moras of structure. Thus, word formation processes such as forming nicknames (e.g., *Lou* from *Louis*, *Joe* from *Joseph*, *Bec* from *Rebecca*) all take a tense/long vowel (CVV) or a coda consonant (CVC): forms that are only monomorphemic, such as *Be* [bɛ] instead of *Bec*, are ill-formed. Thus, although Spanish permits sub-minimal forms in expletives *sí!* 'yes!', it does not permit subminimal open class lexical items. Similarly, English function words can occur with only one mora (e.g., articles *the, a*, etc.), whereas open class lexical items cannot. There are a few languages, however, that permit sub-minimal PWs for open class lexical items such as nouns, verbs, adjectives, etc., and this includes languages like French (*lait* /lɛ/ 'milk', *eau* /o/ 'water') and Japanese (/me/ 'eye').

However, Ota (1999) suggests that words in Japanese child-directed speech are typically augmented to form disyllables with the addition of diminutive suffixes. Thus, Japanese leaners may rarely hear subminimal PWs during development. It also appears that Japanese learners are keenly attuned to the moraic structure of their language from an early age, employing compensatory lengthening as a means of preserving moraic structure when they cannot produce word internal coda consonants (Ota, 1999).

This raises the question of when and how learners in general become sensitive to the fact that PWs in the language they are learning must have a certain amount of prosodic structure. At a stage of development where they cannot yet produce coda consonants, is there any evidence that they produce Sesotho-like 'repairs', enhancing their coda-less productions with an epenthetic vowel, or lengthening the vowel to constitute two moras of structure?

Demuth and Fee (1995) suggested that English learners do show early compensatory strategies, lengthening CVC targets with a short/lax vowel to a long, bimoraic form (e.g, *cat* /kæt/ > [kæ:]). However, as the data were limited, it was difficult to verify these findings. This issue was therefore revisited once data from six English-speaking 1–3-year-olds and their mothers become available as part of the Providence Corpus (Demuth et al., 2006). Although there was no evidence of vowel lengthening from IPA transcriptions, the authors found that coda consonants were more accurately produced in monosyllabic target words with monomoriac vowels, suggesting earlier use of coda consonants in contexts where they could be prosodified as part of a bimoraic PW, or Foot. Song and Demuth (2008) then carried out an acoustic study of three children's spontaneous speech productions from the Providence Corpus (Demuth et al., 2006), and found that target CVC words produced without a coda consonant *did* show lengthening of the vowel. Interestingly, however, this lengthening took place on words with both long and short vowels, suggesting that this was evidence of compensatory lengthening (making up for the missing coda consonant), rather than lengthening the vowel to ensure word minimality. That is, young children were exhibiting vowel lengthening across the board, for both target short and long vowels, rather than selectively for only the short vowels. This suggests that this compensatory process was carried out for *segmental*, rather than prosodic reasons. It thus remained unclear if and when young children might learn that well-formed English PWs must also be minimal words, containing two moras of structure.

Previous study by Kehoe and Stoel-Gammon (1997, 2001) had also suggested that children tend to produce coda consonants earlier in contexts where these are prosodically licensed as part of a bimoraic foot. To explore this issue further Miles, Cox, Yuen, and Demuth (2016) carried out a carefully controlled elicited imitation experiment with Australian English-speaking children aged 2;3. The stimuli contained all monosyllabic words, half with a short vowel (CVC), and half with a long vowel (CVVC). The hypothesis was that the children would be more likely to produce codas with the short-vowel stimuli and allow coda-less forms for the stimuli with long vowels, and this was confirmed: Acoustic analysis showed that the coda consonant was preserved in 83% of the CVC targets, but in only 59% of the CVVC targets. Thus, the odds of a child exhibiting a coda consonant when preceded by a short vowel (CVC) was 1.9 times greater than when preceded by long vowel (CVVC). This suggests that, at least by the age of 2;3, children learning English have some sensitivity to the Minimal Word constraint. It would be very interesting to know when and how this awareness develops in other languages that exhibit these constraints.

The situation is somewhat different in French, where subminimal words constitute 20% of the open class lexical items found in CDS. In a study of a French-speaking child's early productions, Demuth and Johnson (2003) found that target CVC words initially surfaced as CV, and were then reduplicated as CVCV, similar to the early unmarked patterns widely found, as discussed above. However, the child then selectively truncated these (as well as some disyllabic target words) to CV, producing *subminimal* words, (*sel* [sɛl] > [tɛ] 'salt', *coquille* [kɔˈkij] > [tɔti] > [ti] 'shell'). This highly unusual developmental path is not predicted by the theory that proposes a steady increase in prosodic development, where children's PW structures are thought to increase in complexity over time (Demuth, 1996). However, Demuth, and Johnson (2003) showed that the French findings could be explained by appealing to interactions between segmental and prosodic constraints, where unmarked consonants (often /t/) are initially mapped in to output CV forms, but then deleted as the grammar becomes more 'faithful' to mapping target consonants into the output. However, Demuth and Johnson (2003) also argued that children's subminimal CV outputs, which make up a significant portion of the French lexicon, are also prosodically licensed by a mature grammar of French. Thus, truncation to subminimal words in a language like French does not violate the general constraints on prosodic word structure in this language. Interestingly, Prieto (2006) reports similar subminimal CV forms in early Catalan, where the final stressed syllable is attempted, resulting in only a CV PW (*madame* [maˈdam] > [ˈda] 'Mrs.' (Suzanne, 1;6), *pernil* [pərˈnil] [ˈni] 'prosciutto' (Lluís 1;4.10)). Similar findings are reported for European Portuguese (*papa* /ˈpapɐ/ > [ˈpa] 'food' (João 1;3.4) (Vigário, Freitas, & Frota, 2006). For children learning some languages, then, subminimal words may be produced at early stages of development, whether prosodically licensed by the target grammar or not. However, in terms of prosodic development, this may actually represent an advance, gradually becoming more faithful to the segmental content of target form, even at the cost of omitting the occasional (typically unstressed) syllable. This is explored further below.

Truncation of unfooted syllables

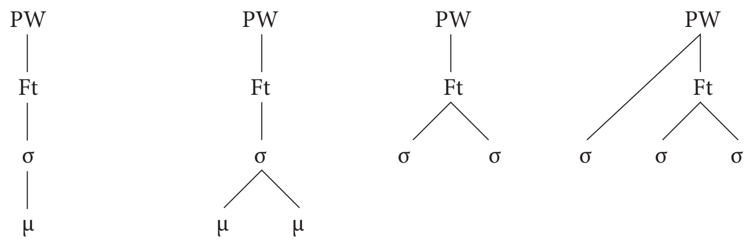
As mentioned above, typical (and/or permitted) forms of the ambient lexicon helps to explain early truncation patterns. This is true also for the truncation of unfooted syllables, which tends to occur in many languages until around 2 years and 6 months of age. Thus, given the typical shape of a PW in the target language, children tend to either omit pretonic unfooted syllables, or shift the stress to match the more typical lexical form. This is clearly seen in examples from English (e.g. *banana* > *nana*) and Dutch (*balon* /baˈlon/ > [ˈbo:mi] 'ballon'), respectively (cf. Demuth, 1996a, 1996b, 2001; Fikkert, 1994). Sometimes words like *banana* are truncated to the more unmarked (obstruent) onset *bana*, providing clear evidence that the unfooted syllable is perceived, just not produced. The same is found for Spanish, where children reduce a trisyllable to a disyllabic trochaic foot, omitting (part of) the syllable that contains segments they cannot produce (e.g., *muñeca* > *meka* 'doll') (Gennari & Demuth, 1997).

Further evidence that children 'know' that these syllables exist in the target form is that they leave acoustic traces for the syllables they omit. This was nicely illustrated in an elicited production experiment by Carter and Gerken (2004).

Thus, given a sentence such as *I like Lucinda*, children aged 2;3 left an acoustic interval/timing unit for the first syllable of the proper name (*I like _cinda).* This suggests that children 'know' that the missing syllable should be in this slot, but lack the speech planning resources to implement it. This raises many questions about the nature of children's early lexical and PW representations. If these are adult-like, then this raises further questions about how these are retrieved from the lexicon, and implemented during production. In the following section we explore this issue in greater detail, along with the early production of prosodic clitics. The development of early PW structures, at least for English (but probably many other languages as well), can therefore be understood in terms of the prosodic structures found in (3) (Demuth, 1996b).

(3) Prosodic Word Structures

a. Subminimal Word b. i. Bimoraic Foot ii. Disyllabic Foot c. Initial Unfooted Syllable



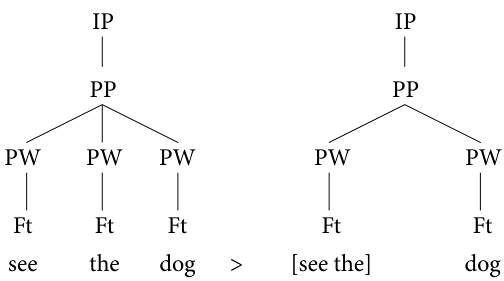
In the next section we see that this has further implications for the acquisition of function morphemes, which often take the form of prosodic clitics.

The development of prosodic clitics: Interactions with PW shape

Of course, children do eventually begin to use PWs that contain more than just a foot of prosodic structure, producing words such as *banana* in English, *balon* 'ballon' in Dutch, and *muñeca* 'doll' and *escalera* 'stairs' in Spanish. This tends to happen by 2;6 for many children learning English, but earlier for Spanish (cf. Lleó, 2001) – again influenced by the frequency with which these more complex prosodic structures occur in the ambient language. Along with the appearance of PWs that constitute more than a Foot of structure comes the ability to produce the beginnings of prosodic clitics such as determiners as well (e.g., *muñeca* > *meca* > *muñeca*, *la muñeca* > *ameca*) (e.g., Gennari & Demuth, 1997; Lleó & Demuth, 1999). This then raises the question of what level of prosodic structure these forms take.

Recall that Selkirk (1996) suggested that clitics are prosodified at different levels of structure depending on both the nature of the clitic and the language, as shown in (2) above. Thus, Gerken, and McIntosh (1993) and Gerken (1996) proposed that English-speaking children's early variable omission of articles is due to prosodic constraints: Those determiners following a monosyllabic word can be prosodified as part of a foot and PW, as in (2b) above, whereas those following a disyllabic word are left unfooted, to be prosodified at the higher level of the PP or IP, as in (2c) above. This suggests that, before 2;3 years, many English-speaking children might not yet represent higher-level IP structure, or at least do not/cannot prosodify function words at this level, and thus omit unfooted articles. (See Snow (1994), however, who provides evidence of phrase-final lengthening at 2 years, suggesting a *boundary* awareness of IP structure at this age). The results of Gerken's (1996) elicited imitation findings are confirmed in analysis of spontaneous speech (Demuth & McCullough, 2009). Further acoustic analysis of one of the children's productions in this latter study even showed development in this respect, first producing the article *the* as an independent (stressed) PW at 1;10, but then as part of a Foot with the previous monosyllabic word by the age of 2;0, as shown in (4).

(4) Prosodic Incorporation



Thus, although many researchers have proposed that the omission of grammatical morphemes in children's early productions is due to syntactic or semantic limitations, it appears that many early aspects of children's variable production of grammatical morphemes may be due to their limited abilities with prosodic representations. In particular, those grammatical morphemes that are 'prosodically licensed' as part of the child's prosodic grammar tend to be more reliably produced (cf. Demuth, 2014).

Crosslinguistic research has shown that children's early production of articles and determiners (English, French, Spanish, German, Italian), as well as Bantu noun class prefixes (Demuth & Ellis, 2009) is thus prosodically constrained, interacting with children's level of prosodic competence at the level of PP and IP structure. However, the driving force behind this expansion to higher levels of prosodic structure appears to be the lexicon itself. Thus, children learning languages such as Spanish and Italian, where most words are 2 or 3 syllables in length (e.g., often more than a Foot), appear to represent higher levels of PP and IP structure earlier than children acquiring primarily monosyllabic English (e.g., Demuth, Patrolia, Song, & Masapollo, 2012; Giusti & Gozzi, 2006). Children learning languages with a large number of disyllabic stems, such as French and Sesotho, fall somewhere in between. Evidence for this comes from examining when and how increasingly large prosodic structures, including words with a greater number of syllables, and the production of prosodic clitics (e.g. articles, determiners, noun class prefixes, etc.), begin to be produced (e.g. Spanish: *la eskalera* 'the stairs': *lera* > *kalera* > *akalera* > *la-eskalera*) (cf. Demuth et al., 2012). Thus, the nature of the ambient lexicon plays a major role in how and when grammatical morphemes such as determiners are acquired.

This is part of a much larger picture of how children's acquisition of grammar interfaces with their growing competence with prosodic structure more generally. Children's early use of phrase final lengthening (Snow, 1994) provides the context for the first use of inflectional morphology in a language like English. Thus, recent research has shown that children are much more likely to produce plural and tense/agreement inflections on nouns and verbs that occur phrase finally (*Now she <u>cries</u>, I like <u>dogs</u>*) compared to phrase medially (*She <u>cries</u> now, The <u>dogs</u> run*) (e.g., Song et al., 2009, 2012; see also Hsieh, Leonard, & Swanson, 1999, and Demuth, 2014, for review). We suspect that many other cases of function morpheme omission are also due to limitations in children's use of prosodic structure, but this has yet to be more fully investigated crosslinguistically. This is part of a larger developmental process of gradually becoming more adult-like in many different aspects of speech planning and production (cf. Yuen, Demuth, & Johnson, 2011), an issue that still requires much future research.

What might a developmental model of speech planning look like? Adult models like that of Levelt and colleagues (Levelt et al., 1999) assume an adult-like lexicon, adult-like representation of PWs, and access to a full prosodic hierarchy. It also assumes an adult-like working memory capacity, and the concomitant ability to construct a lemma, access words from the mental lexicon, formulate motor commands, and execute them using language-specific prosodic organization. If the first pieces of this process are not yet adult-like, the ultimate prosodic organization of an utterance, as well as the words within the utterance, may not be completely well-formed. Much research is still needed to better understand the various steps in this process, and crosslinguistic evidence will be critical, just as it has been for understanding the development of prosodic units such as the syllable and PWs. This will hopefully provide a better understanding of why children omit certain syllables and grammatical morphemes in early speech, why child speech often sounds like it is a collection of PWs combined into an IP, with somewhat flat rather than hierarchical prosodic structure, and how and when children become prosodically more adult-like.

Conclusion

Much is still to be learned about how children's prosodic structure develops, including at higher levels of prosody (see, for example, Chen & Fikkert, 2007; Frota et al., 2016; Prieto, Estrella, Thorson & Vanrell, 2012; and Yuen et al., 2016, amongst others). Critically important to this enterprise is an acoustic understanding of how different levels of the prosodic hierarchy are instantiated across languages, and how children begin to perceive and produce the acoustic cues to these different prosodic structures in their own speech productions as these become syntactically and prosodically more complex. It is hoped that this paper will provide the beginnings of a road map for further investigation in this direction. Ultimately, the findings will help inform a much-needed developmental model of speech planning and production, providing further insight into how and when child prosody begins to approximate more fluent speech.

Funding

This work was funded in part by the following grants: ARC FL130100014, ARC CE110001021.

Acknowledgements

I thank Ivan Yuen for assistance with the figures, and reviewers for their helpful comments.

References

Boersma, P., & Weenink, D. (2005). *Praat: Doing phonetics by computer* (Version 4.4.07). Retrieved from < http://www.fon.hum.uva.nl/praat/ >.

Carter, A., & Gerken L. A. (2004). Do children's omissions leave traces? *Journal of Child Language*, 31, 561–586. doi:10.1017/S030500090400621X

Chen, A., & Fikkert, P. (2007). Intonation of early two-word utterances in Dutch. *Proceedings of the 16th International Congress of Phonetic Sciences* – Saarbrücken, 315–320.

Demuth, K. (1995). Markedness and the development of prosodic structure. *Proceedings of the North East Linguistic Society* 25 – Amherst, 13–25.

Demuth, K. (1996a). The prosodic structure of early words. In J. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 171–184). Mahwah, NJ: Lawrence Erlbaum Associates.

Demuth, K. (1996b). The prosodic structure of early words. In J. Morgan, & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 171–184). Mahwah, NJ: Lawrence Erlbaum Associates.

Demuth, K. (2001). Prosodic constraints on morphological development. In J. Weissenborn, & B. Höhle (Eds.), *Approaches to bootstrapping: Phonological, syntactic and neurophysiological aspects of early language acquisition* (pp. 3–21). Amsterdam: John Benjamins. doi:10.1075/lald.24.02dem

Demuth, K. (2006). Crosslinguistic perspectives on the development of prosodic words. Guest Editor, Special Issue, *Language and Speech*, 49, 129–297. doi:10.1177/00238309060490020101

Demuth, K. (2014). Prosodic Licensing and the development of phonological and morphological representations. In A. Farris-Trimble, & J. Barlow (Eds.), *Perspectives on phonological theory and acquisition: Papers in honor of Daniel A. Dinnsen* (pp. 11–24). Amsterdam: John Benjamins. doi:10.1075/lald.56.04dem

Demuth, K., Culbertson, J., & Alter, J. (2006). Word-minimality, epenthesis, and coda licensing in the acquisition of English. *Language & Speech*, 49, 137–174. doi:10.1177/00238309060490020201

Demuth, K., & Ellis, D. (2009). Revisiting the acquisition of Sesotho noun class prefixes. In J. Guo, E. Lieven, N. Budwig, S. Ervin-Tripp, K. Nakamura, & S. Ozçalikan (Eds.). *Crosslinguistic approaches to the psychology of language: Festschrift for Dan Slobin* (pp. 93–104). New York, NY: Psychology Press.

Demuth, K., & Fee, J. (1995). *Minimal prosodic words in early phonological development*. Unpublished manuscript. Brown University and Dalhousie University.

Demuth, K., & Johnson, M. (2003). Truncation to subminimal words in early French. *Canadian Journal of Linguistics*, 48, 211–241. doi:10.1353/cjl.2004.0025

Demuth, K., & McCullough, E. (2009). The prosodic (re)organization of children's early English articles. *Journal of Child Language*, 36, 173–200. doi:10.1017/S0305000908008921

Demuth, K., Patrolia, M., Song, J. Y., & Masapollo, M. (2012). The development of articles in children's early Spanish: Prosodic interactions between lexical and grammatical form, *First Language*. *Special Issue on Linguistic Interfaces and Language Acquisition in Childhood*, 32, 17–37.

Demuth, K., & Tremblay, A. (2008). Prosodically-conditioned variability in children's production of French determiners. *Journal of Child Language*, 35, 99–127. doi:10.1017/S0305000907008276

Doke, C. M., & Mofokeng, S. M. (1985). *Textbook of Southern Sotho grammar*. Cape Town: Longman.

Fikkert, P. (1994). *On the acquisition of prosodic structure*. Unpublished PhD dissertation. University of Leiden.

Frota, S., M. Cruz, N. Matos, & M. Vigário. (2016). Early prosodic development: Emerging intonation and phrasing in European Portuguese. In M. Armstrong, N. C. Henriksen, & M. M. Vanrell (Eds.), *Intonational grammar in Ibero-Romance: Approaches across linguistic subfields* (pp. 295–324). Amsterdam: John Benjamins.

Gennari, S., & Demuth, K. (1997). Syllable omission in Spanish. In E. Hughes, M. Hughes, & A. Greenhill (Eds.), *Proceedings of the 21st Annual Boston University Conference on Language Development* (Vol. 1, pp. 182–193). Somerville, MA: Cascadilla Press.

Gerken, L. A. (1996). Prosodic structure in young children's language production. *Language*, 72, 683–712. doi:10.2307/416099

Gerken, L. A., & McIntosh, B. (1993). The interplay of function morphemes and prosody in early language. *Developmental Psychology*, 29, 448–457. doi:10.1037/0012-1649.29.3.448

Gnanadesikan, A. (2004). Markedness and faithfulness constraints in child phonology. In R. Kager, J. Pater, & W. Zonneveld (Eds.), *Constraints in phonological acquisition* (pp. 73–109). Cambridge: Cambridge University Press.

Giusti, G., & Gozzi, R. (2006). The acquisition of determiners: Evidence for the Full Competence Hypothesis. *Language acquisition and development. Proceedings of Generative Approaches to Language Acquisition 2005* – Siena, 232–238.

Hansson, G. (2001). *Theoretical and typological issues in consonant harmony*. Unpublished PhD dissertation. University of California at Berkeley.

Hsieh, L., Leonard, L. B., & Swanson, L. (1999). Some difference between English plural noun inflections and third singular verb inflections in the input: The contributions of frequency, sentence position, and duration, *Journal of Child Language*, 26, 531–543. doi:10.1017/S030500099900392X

Hyams, N. (2007). Aspectual effects on interpretation in early grammar. *Language Acquisition*, 14, 231–268. doi:10.1080/10489220701470992

Kehoe, M., & Stoel-Gammon, C. (1997). The acquisition of prosodic structure: An investigation of current accounts of children's prosodic development. *Language*, 73, 113–144. doi:10.2307/416597

Kehoe, M., & Stoel-Gammon, C. (2001). Development of syllable structure in English-speaking children with particular reference to rhymes. *Journal of Child Language*, 28, 393–432. doi:10.1017/S030500090100469X

Kent, R. D. (1976). Anatomical and neuromuscular maturation of the speech mechanism: Evidence from acoustic studies. *Journal of Speech Language and Hearing Research*, 19, 421–447. doi:10.1044/jshr.1903.421

Kern, S., & Davis, B. (2009). Emergent complexity in early vocal acquisition: Cross-linguistic comparisons of canonical babbling. In I. Chitoran, F. Pellegrino, & E. Marsico (Eds.), *Approaches to phonological complexity* (pp. 353–375). Berlin: Mouton de Gruyter. doi:10.1515/9783110223958.353

Kirk, C., & Demuth, K. (2006). Accounting for variability in 2-year-olds' production of coda consonants. *Language Learning and Development*, 2, 97–118. doi:10.1207/s15473341lld0202_2

Levelt, C. C. (1995). Segmental structure of early words: Articulatory frames or phonological constraints. *Proceedings of the Twenty-seventh Annual Child Language Research Forum* – Stanford, 19–27.

Levelt, C. C., Schiller, N. O., & Levelt, W. J. (2000). The acquisition of syllable types. *Language Acquisition*, 8, 237–264. doi:10.1207/S15327817LA0803_2

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1–75. doi:10.1017/S0140525X99001776

Lleó, C. (2001). The interface of phonology and morphology: the emergence of the article in the early acquisition of Spanish and German. In J. Weissenborn, & B. Höhle (Eds.), *Approaches to bootstrapping: Phonological, syntactic and neurophysiological aspects of early language acquisition* (pp. 23–44). Amsterdam: John Benjamins. doi:10.1075/lald.24.03lle

Lleó, C. (2003). Prosodic licensing of codas. *Probus*, 15, 257–281. doi:10.1515/prbs.2003.010

Lleó, C. (2006). The acquisition of prosodic word structures in Spanish by monolingual and Spanish-German bilingual children. *Language and Speech*, 49, 207–231. doi:10.1177/00238309060490020401

Lleó, C., & Demuth, K. (1999). Prosodic constraints on the emergence of grammatical morphemes: Crosslinguistic evidence from Germanic and Romance languages. In A. Greenhill, H. Littlefield, & Cheryl Tano (Eds.), *Proceedings of the 23rd Annual Boston University Conference on Language Development* (pp. 407–418). Somerville, MA: Cascadilla Press.

MacNeilage, P. F. (1980). The control of speech production. In G. Yeni-Komshian, J. Kaveanagh, & C. Ferguson (Eds.), *Child phonology: Production* (Vol. 1, pp. 9–21). New York, NY: Academic Press.

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk. Vol 2: The Database* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Miles, K., Cox, F., Yuen, I., & Demuth, K. (2016). The prosodic licensing of coda consonants in early speech: Interactions with vowel length. *Journal of Child Language*, 43, 265–283.

Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris.

Ota, M. (1999). *Phonological theory and the acquisition of prosodic structure: Evidence from child Japanese*. Unpublished PhD dissertation. University of Georgetown.

Prieto, P. (2006). The relevance of metrical information in early prosodic word acquisition: A comparison of Catalan and Spanish. *Language & Speech*, 40, 231–259. doi:10.1177/00238309060490020501

Prieto, P., Estrella, A., Thorson, J., & Vanrell, M. M. (2012). Is prosodic development correlated with grammatical development? Evidence from emerging intonation in Catalan and Spanish. *Journal of Child Language*, 39, 221–257. doi:10.1017/S030500091100002X

Prince, A., & Smolensky, P. (2004). *Optimality Theory: Constraint interaction in generative grammar*. Malden, MA: Blackwell. doi:10.1002/9780470759400

Roark, B., & Demuth, K. (2000). Prosodic constraints and the learner's environment: A corpus study. In S. C. Howell, S. A. Fish, & T. Keith-Lucas (Eds.), *Proceedings of the 24th Annual Boston University Conference on Language Development* (pp. 597–608). Somerville, MA: Cascadilla Press.

Rose, S., & Walker, R. (2004). A typology of consonant agreement as correspondence. *Language*, 80, 475–531. doi:10.1353/lan.2004.0144

Rose, Y., & MacWhinney, B. (2014). The PhonBank project: Data and software-assisted methods for the study of phonology and phonological development. In J. Durand, U. Gut, & G. Kristoffersen (Eds.), *The Oxford handbook of corpus phonology* (pp. 380–401). Oxford: Oxford University Press.

Selkirk, E. O. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: The MIT Press.

Selkirk, E. O. (1996). The prosodic structure of function words. In J. L. Morgan, & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 187–213). Mahwah, NJ: Lawrence Erlbaum Associates.

Snow, D. (1994). Phrase-final syllable lengthening and intonation in early child speech. *Journal of Speech, Language, and Hearing Research*, 37, 831–840. doi:10.1044/jshr.3704.831

Stevens, K. N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. *Journal of the Acoustical Society of America*, 111, 1872–1891. doi:10.1121/1.1458026

Song, J. Y., & Demuth, K. (2008). Compensatory vowel lengthening for omitted coda consonants: A phonetic investigation of children's early representations of prosodic words. *Language & Speech*, 51, 382–399.

Song, J. Y., Demuth, K., & Shattuck-Hufnagel, S. (2012). The development of acoustic cues to coda contrasts in young children learning American English. *Journal of the Acoustical Society of America*, 131, 3036–3050. doi:10.1121/1.3687467

Song, J. Y., Sundara, M., & Demuth, K. (2009). Phonological constraints on children's production of English third person singular *-s*. *Journal of Speech, Language, and Hearing Research*, 52, 623–642. doi:10.1044/1092-4388(2008/07-0258)

Spencer, A. (1986). Towards a theory of phonological development. *Lingua*, 68, 3–38. doi:10.1016/0024-3841(86)90021-5

Vigário, M., Freitas, M. J., & Frota, S. (2006). Grammar and frequency effects in the acquisition of prosodic words in European Portuguese. *Language and Speech*, 49, 175–203. doi:10.1177/00238309060490020301

Wexler, K. (1994). Optional infinitives, head movement, and the economy of derivation in child language. In D. Lightfoot, & N. Hornstein (Eds.), *Verb movement* (pp. 305–350). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511627705.016

Yuen, I., Demuth, K., Johnson, M. (2011). Prosodic structure in child speech planning and production, *Proceedings of ICPhS XVII* – Hong Kong, 2248–2251.

Yuen, I., Xu Rattanasone, N., Schmidt, E., Macdonald, G., Holt, R., & Demuth, K. (2016). Accentual lengthening in 5-year-old AusE-speaking children: Preliminary results. *Proceedings of the 16th Australasian International Conference on Speech Science and Technology* – Sydney, 21.

PART 3

Moving to meaning: Prosody and pragmatic development

CHAPTER 12

Early development of the prosody-meaning interface

Núria Esteve-Gibert[1,2] & Pilar Prieto[3,4]
[1]Aix Marseille Université, CNRS, LPL UMR 7309 / [2]Universitat de Barcelona / [3]Institució Catalana de Recerca i Estudis Avançats (ICREA) / [4]Universitat Pompeu Fabra, Department of Translation and Language Sciences

This chapter reviews evidence on how infants up to 18 months of age develop the ability to use prosody as a sign of the expression of pragmatic meanings, from both a comprehension and a production point of view. Developmental research reveals that pre-lexical infants use prosodic information not only to comprehend emotions in the speech of their communicative partners, the intentional value of the partners' speech, and their speech act motivation, but also to express these same pragmatic meanings when they communicate with others. In essence, before the emergence of lexical and grammatical skills, infants use prosody to communicate intentionally with the world around them.

Introduction

Human languages resort to a variety of linguistic strategies for the expression of meaning, including not only grammatical and lexical encoding but also prosody. Prosodic patterns in human speech convey a set of pragmatic meanings that a young child must learn to comprehend. Research on the prosody-meaning interface has traditionally distinguished between two types of information that can be expressed through prosodic patterns: on the one hand, attitudes and emotions (what has been traditionally called paralinguistic or affective meaning) and, on the other hand, speech act information, information structure, and speakers beliefs (what has been called linguistic meaning) (e.g., Gussenhoven, 1984; Ladd, 1996; see Prieto, 2015, for a review). Traditionally, prosody has been separated into two types, emotional/affective prosody and linguistic/pragmatic prosody. More recent accounts, however, have rejected this simple dichotomy on the grounds that prosody is used by speakers to convey a wider range of dimensions

DOI 10.1075/tilar.23.12est

of meanings than just these two (such as emotions and affect, politeness, speech acts, agreement or disagreement between interlocutors, epistemic positioning, or information structure), and, furthermore, these dimensions are not mutually exclusive to each other anyway (Portes, Beyssade, Michelas, Marandin, & Champagne-Lavau, 2014; Prieto, 2015; for a claim that emotions can be seen as pragmatic and epistemic actions, see Wilutzky, 2015). In the present chapter we will refer to all of such dimensions as meanings or pragmatic meanings indistinguishably, since the debate about whether they are all linguistic or not lies beyond the scope of this chapter.

The central question we will address here is when and how infants start to develop the ability to understand and express meanings through prosody. Much of the answer must lie in the human environment with which a newborn has first contact. Indeed, through the exaggerated prosodic properties of infant-directed speech, caregivers make the relationship between prosodic form and communicative function uniquely salient to children (see Saint-Georges et al., 2013, for a review), thus paving the way for infants to start learning the prosody-meaning correspondences of their mother tongue. Over the last few decades, developmental research has focused heavily on how infants use sensitivity to prosody as a tool to discriminate among the rhythmic patterns, pitch cues, and accentual patterns that are typical of the language that surrounds them (e.g., Bhatara, Boll-Avetisyan, Höhle, & Nazzi, this volume). This early sensitivity to prosody has been claimed to function as a bootstrapping mechanism for early speech segmentation as well as word learning (e.g., Bhatara et al., this volume; de Carvalho, Dautriche, Millotte, & Christophe, this volume; Thorson, this volume; Teixidó, François, Bosch, & Männel, this volume). However, work on children's use of this early sensitivity to prosody to access the pragmatic meaning dimension is still relatively sparse, and little is known about how children exploit the various meanings conveyed by prosody, especially in the early stages of language development.

Being able to map incoming speech to meaning is a fundamental skill in early language acquisition because it is crucial for successful social interactions. From the moment they are born, infants show social skills that act as important precursors and predictors of their later linguistic and cognitive abilities. Newborns prefer to look at human faces more than other types of stimuli (e.g., Johnson, Dziurawiec, Ellis, & Morton, 1991), and they prefer to listen to their mother's voice over the voices of other females or other people in general (DeCasper & Fifer, 1980). Thanks to these early social abilities, infants know when they are being addressed, have expectations about what will come next in social interactions, and begin to infer relevant and generalizable information about communicative acts (Csibra & Gergely, 2009).

Around the middle of the first year of life, infants develop communicative skills that are crucial for the emergence of intentional communication. They start to engage in joint attentional frames, following a communicative partner's gaze or gesture towards an event or an action or directing the other person's attention towards it by using gaze or gesture strategies themselves (e.g., Bruner, 1975). Infants as young as 5 months also notice that actions can have different goals (Woodward, 1998), and by the end of their first year of life they can attribute a goal to an action even if the action is not completed (e.g., Meltzoff, 1995). In order to process and express meanings in interaction, infants use contextual information such as the physical presence of the objects being referred to or the social actions that preceded the interaction (e.g., Clark, 2004, among many others).

This chapter will review evidence that reveals how infants (from 0–18 months) use prosody as a tool to comprehend and express pragmatic meanings in interactions. The evidence will be drawn from a set of studies that show infants' abilities to successfully map prosody to pragmatic meaning in three separate areas, namely (a) the comprehension and expression of emotion-related meanings, (b) the distinction between intentional and non-intentional speech, and (c) the expression and comprehension of speech act motives. The motivation for this three-way division comes from the fact that, although these areas interact with and feed off each other, previous research has traditionally studied them separately. On the one hand, some studies have investigated when and how infants learn that actions are goal-directed (i.e., intentional, as opposed to accidental), and on the other hand, other studies have examined when infants learn to produce and comprehend actions with specific social motivations (in speech, these would be regarded as speech acts). In essence, the existing evidence indicates that infants use prosody as a tool to infer and express pragmatic meaning well before they can use lexical and grammatical tools for these purposes.

The prosody-meaning interface in infancy: Comprehension

There is a vast literature on language development showing that babies are sensitive to the prosodic features of speech such as pitch and duration, even before birth (see Bhatara et al., this volume; de Carvalho et al., this volume). Four-month-old infants prefer to listen to infant-directed speech (i.e., speech with exaggerated prosodic features such as higher pitch range, higher mean pitch, and slower speech tempo) than to adult-directed speech (e.g., Fernald, 1985). Similarly, babies can discriminate between languages which display timing differences such as utterance-final lengthening or languages which belong to distinct rhythm classes (e.g., Nazzi, Bertoncini, & Mehler, 1998), and they can also distinguish between different stress patterns in the ambient language. Infants between 4 and 9 months of age prefer to listen to the rhythmic and prosodic patterns of their ambient language and can use this information to segment the speech stream into words (e.g., Höhle, Bijeljac-Babic, Herold, Weissenborn, & Nazzi, 2009).

Infants' sensitivity to pitch direction also emerges very early in life. Infants between 4 and 9 months of age discriminate between rising and falling intonation patterns in the case of intonational languages like Portuguese or English (Frota, Butler, & Vigário, 2014) and several different tonal shapes in the case of tonal languages like Mandarin or Cantonese (Yeung, Chen, & Werker, 2013). These studies suggest that language-specific perceptual sensitivity emerges earlier for prosodic patterns (e.g., pitch and stress) than for segmental patterns (e.g., vowels and consonants). However, less is known about how young infants use this early sensitivity to prosodic cues in social contexts to infer pragmatic meanings expressed by their interlocutors. The following sections discuss the current findings on how infants develop the ability to map prosodic cues to social meaning, focusing on how prosody is used by infants to understand a communicative partner's (a) emotional and affective state, (b) goal to act intentionally, and (c) motivation for a particular speech act.

Infants' early understanding of others' emotional states through prosody

The sensitivity to the emotional valences expressed by prosodic cues develops very early in infants. By 5 months of age, infants are found to react differently to different affective vocal expressions (for instance, smiling in response to approval vocalizations), can discriminate between sad or angry vocalizations if the vocalizations are accompanied by facial expressions, and can match positive and negative affective vocalizations with the appropriate facial expression (e.g., Fernald, 1993; Vaillant-Molina, Bahrick, & Flom, 2013). These behavioral findings have been confirmed by evidence coming from electrophysiological and brain imaging studies. Electrophysiologically measured responses have revealed that newborns are sensitive to threat-related emotions, with different activations of the fronto-central scalp region being found for angry and fearful voices respectively (Zhang et al., 2014). Similarly, evidence coming from near infrared spectroscopy (NIRS) (e.g., Grossmann, Oberecker, Koch, & Friederici, 2010) and from ERP responses (e.g., Blasi et al., 2011) has shown that 7-month-old infants react differently to semantically-neutral words depending on whether they are produced with a happy, angry, or neutral tone of voice (see Teixidó, François, Bosch, & Männel, this volume, for more information about infants' brain responses to prosody for speech segmentation and word learning).

Before their first birthday, infants develop a strong understanding of emotional states based on their analysis of prosodic cues. Mumme, Fernald, and Herrera (1996) asked mothers to produce vocal cues of positive, negative, or neutral affect directed towards a referent located in front of their 12-month-old infant. In the positive affect condition, the mothers were asked to say 'Oh, how delightful!', the *Oh* being high-pitched and smooth, and the *how delightful* being produced in a relaxed voice with a rise-fall pitch. In the negative affect condition, the mothers said 'Oh, how frightful!', the *Oh* here produced with a gasping inhalation, and the *how frightful* produced rapidly with a tense voice slightly high in pitch. In the neutral affect condition, the mothers said 'Oh, how insightful!', the entire utterance being produced in a monotone voice. The results revealed that the infants showed negative-affect behaviors towards the referent after mothers had expressed vocal cues of negative affect, but not when they had expressed vocal cues of positive or neutral affect. Interestingly, they also revealed that not all affects seem to have the same impact, since vocal cues of positive affect did not induce positive-affect behaviors in the infants.

Despite the fact that Mumme et al.'s (1996) results showed that infants could infer an adult's emotions by relying on vocal cues, the experimental design left it unclear as to whether it was the lexical or prosodic content of the vocalization (or both) that induced this effect. Although at such a young age the infants presumably did not yet understand words like *frightful* or *delightful,* the possibility that lexical knowledge might have influenced their responses cannot be completely ruled out. More recent studies have managed to isolate prosody as a variable, with results showing that 15-month-old infants are able to associate pitch values in a communicative partner's speech with the emotions (s)he is intending to convey. Hoicka and Wang (2011) presented 15-month-old infants with an experimenter producing either humorous or 'sweet' vocal cues, the only difference between the two being pitch mean values (the humorous cues having a higher pitch mean than the 'sweet' ones). After being exposed to the vocal cues, infants observed the experimenter producing either a matching behavior or the opposite one (i.e., either a humorous action or a sweet action). Importantly, the experimenter maintained a neutral facial expression in all conditions in order to prevent infants from using facial cues to drive their inferences. Infants looked longer at the behavior if it mismatched the vocal cues than if it matched them, revealing that they had developed expectations about the adult's behavior given a specific prosodic feature. These results show that already at 15 months of age infants can make prosody-meaning associations independently of situational context and lexical cues.

The studies reviewed so far show an early mapping of prosodic cues to emotional states. This early comprehension, however, continues to develop in later stages up to adolescence. Preschool children can match speech bearing different acoustic cues of emotion to the appropriate facial expressions (Berman, Chambers, & Graham, 2016), and by the time they enter school children will rely on prosodic cues for the comprehension of an interlocutor's emotion when these cues compete with lexical information or situational context, such as in ironic speech (see Armstrong & Hübscher, this volume, for a review of the research on the expression of affect through prosody at these later stages).

Early understanding of prosody as a marker of intentional communication

The set of cognitive abilities that enable infants to comprehend a communicative partner's intentions is called Theory of Mind. Theory of Mind, also known as 'mind-reading' or perspective-taking, is the capacity to understand that other people have beliefs, intentions, and desires, and that these may differ from one's own (Premack & Woodruff, 1978, and many others thereafter). Research has shown that Theory of Mind emerges only gradually in development, the simplest forms being observed in early infancy and the most complex ones during the school years (see Liszkowski, 2013, for a review). At the outset, infants have flexible expectations about other people's actions depending on the preceding shared actions or the situational context in which these actions occur. Seven-month-old infants, for instance, are surprised when an event does not fulfill their own and someone else's expectations (Kovács, Téglás, & Endress, 2010), and toddlers make predictions about another person's beliefs (Onishi & Baillargeon, 2005). Only much later do other more complex forms of Theory of Mind emerge. During the preschool years, for example, children develop the ability to hold first-order representations, that is, an awareness that somebody else can hold a belief that may differ from their own (Baron-Cohen, Leslie, & Frith, 1985). By about 5 to 7 years of age, or even later, children are capable of holding second-order false belief representations, the realization that one person may hold a belief that contrasts with another person's belief about something (e.g., Miller, 2009; Sullivan, Zaitchik, & Tager-Flusberg, 1994).

Intentional communication typically emerges around 9–12 months of age, and studies have shown that by around their first birthday infants know that speech is used to refer to abstract intentions (Martin, Onishi, & Vouloumanos, 2012; Vouloumanos, Onishi, & Pogue, 2012). In these studies infants were exposed to act-preceding contextual information shared between the interlocutors, and this helped them to interpret the speech as meaningful. Although most of the existing research on the mechanisms by which young infants infer an interlocutor's intentions highlight the role played by information that the infants extract from the context prior to the shared action (e.g., Tomasello, Carpenter, & Liszkowski, 2007), there is also recent evidence that infants rely heavily on act-accompanying signals like prosody and gesture to understand that the other person's speech is intentional. In order to test whether infants detect intentionality from speech cues alone, Carpenter, Akhtar, and Tomasello (1998) designed a study in which intentional and non-intentional actions were preceded by the same shared action context and only speech features differed between the two conditions. The authors tested 14- to 18-month-old infants using an action imitation paradigm. They found that infants imitated an action more if it was followed by a verbal sign of intentionality (the expression 'There!' with appropriate prosody) than if followed by a verbal sign indicating that the action was accidental (the expression 'Whoops!' with appropriate prosody). Since the two actions were performed by the experimenter in identical fashion with the only difference between them being the verbal message (even the facial expressions were neutralized as much as possible), the authors concluded that at the beginning of their second year of life infants can use prosodic properties in speech to infer that an action is intentional.

In order to tap into the specific role of prosody in infant's detection of intentional and non-intentional actions, Sakkalou and Gattis (2012) tested whether 14- and 18-month-old infants could detect the intentional nature of an action by relying on the prosodic cues of the word accompanying the action. Using Carpenter et al.'s (1998) methodology, infants were first exposed to intentional actions that were accompanied by the word 'There!', produced with high amplitude and long duration, and accidental actions that were accompanied by the word 'Whoops!', produced with low amplitude and short duration; in a second experiment, the authors removed the lexical information by testing L1 English-learning infants with the Greek equivalents for 'There!' and 'Whoops!'. In both experiments the authors found that infants imitated the intentional actions more often than the accidental ones, and they observed an age-related effect when lexical cues were removed (i.e., older infants performed better than younger infants). Since infants could only have based their imitative behavior on the prosody accompanying the action, it was concluded that infants comprehended the pragmatic value of prosodic cues and were able to relate them to speech intentionality at 14 months of age. These findings were confirmed and further explored in a later study examining individual differences (Sakkalou, Ellis-Davies, Fowler, Hilbrink, & Gattis, 2013). Interestingly, the authors found that the infants who at 13 months imitated intentional actions most frequently where the ones who were better at detecting intentional actions on the basis of prosody at 14 months of age.

Early understanding of prosody as a marker of speech act information

In communicative interactions, interlocutors detect not only that the speaker is communicating intentionally but also the specific intention behind the communicative act (i.e., they can differentiate between informative/assertive, requestive/directive, and expressive speech acts). Caregivers use a different set of linguistic patterns, including distinct melodies, depending on whether they are uttering a statement, greeting someone, asking a question, making a request, or refusing something (e.g., Armstrong, 2012; Esteve-Gibert, Liszkowski, & Prieto, 2016). Esteve-Gibert et al. (2016) analyzed the speech of caregivers as they played with infants. The experiment was designed to elicit pointing actions on the part of the caregiver either to ask the infant to give them an object (imperative condition), share interest in the object with the infant (expressive condition), or inform the infant about a specific feature of that object (informative condition). Their results showed that in each of the three conditions caregivers accompanied their lexical message with a specific combination of prosodic patterns (intonation contour, pitch range, and mean syllable duration) and gesture features (hand shape, gesture duration) in order to distinguish the intention they wished to transmit.

Again, most of the previous research on infants' early comprehension of speech acts has focused on how infants use the contextual cues in which these acts are embedded to comprehend other people's intentions. Although 12-month-old infants do use social contextual cues to understand whether an adult has produced an attention-directing action (such as pointing) in order to request an object, inform about its presence, or express interest in it (see Liszkowski, 2014, for a review), often times social contextual cues are ambiguous or underspecified. Since prosody and gesture constitute another marker of adults' pragmatic intentions, it is possible that young infants can resort to them to infer an interlocutor's intended meaning. This hypothesis was tested in a subsequent study in which prosodic features (alongside pointing gesture cues) were the only cues that infants could use to infer the motivation behind a communicative partner's speech act. Esteve-Gibert, Prieto, and Liszkowski (2017) carried out two experiments to determine whether 12-month-old infants were capable of successfully associating specific prosodic patterns (and hand gestures) with the speaker's underlying speech act intention. The first experiment had the goal of testing the role of prosody in conjunction with lexical and gestural cues, and the second one was intended to test exclusively the role of prosody and gestures, without lexical input. Infants saw an adult who produced gestures to direct the infant's attention towards an object with either an imperative, expressive, or informative intent. The measure of interest was whether, for each condition, the infants reacted in a way that was consistent with the adult's intention by either giving the object to the adult in the imperative condition, sharing interest in the object with the adult in the expressive condition, or discovering the feature of the object that they had been informed about in the information condition. Importantly, the situational context was the same across the three possible conditions to rule out the possibility that infants would use contextual information to infer the motive behind the adult's speech act. The results of the first experiment showed that infants could use lexical information, prosody, and gesture as a whole to understand the speaker's intention behind the attention-directing action. The results of the second experiment showed that when the lexical information was neutralized across intentions (by having the adult say 'Hey, this, this!' in all three conditions), infants could still understand the adult's intention and react accordingly.

In sum, evidence to date shows that a few months after being born and within their first year of life infants develop a fine perceptual sensitivity to the differences in prosodic cues in speech that can be used to understand a communicative partner's affective internal state and the intentional value of their speech. Later on, around 12–15 months of age, they seem to be able to associate specific variations in the prosodic cues (pitch, duration, and intonation) of their caregiver's speech with specific speech act meanings.

The prosody-meaning interface in infancy: Production

In this section, we review the literature on the pathways followed by infants in their early expression of pragmatic meaning through prosodic cues. As mentioned above, babies are very closely attuned to the prosodic patterns of the ambient language(s) during their first year of life. There is evidence that newborn's early production of prosodic cues like intonation is not random but instead progressively linked to the patterns they are exposed to in their target language (e.g., Mampe, Friederici, Christophe, & Wermke, 2009; Wermke et al., 2016). Mampe et al. (2009), for instance, found that the intonation pattern of newborns' cries is tuned to the most common pattern of the language to which they have been exposed while still inside their mother's womb. The authors tested French and German newborns only a few hours after being born, and found that French newborns produced rising contours when crying, while German newborns' cries had a falling contour. This difference reflects the most frequent tonal pattern of each language, since French is mostly a rising-intonation language and German a falling-intonation one. Some months later, infants' babbling has also been found to reflect the prosodic correlates of stress – namely higher F0, intensity, and duration values – of their ambient language (Davis, MacNeilage, Matyear, & Powell, 2000), as well as the pitch range values of the ambient language (Snow, 2006). Similarly, the form of infants' first words is found to be constrained by the structure of prosodic words and phrases in the language they hear around them (Kehoe, this volume). Again, this raises questions about when and how infants start to produce the target prosodic features of the language to express their emotional and intentional states. The following sections explore this issue.

Infants' early use of prosody to mark their emotional status in speech

In general, studies of infants' emerging control of prosody for the expression of affective and emotional meaning show that from 2 months of age onwards, infants start to differentiate positive and negative emotions in their vocalizations on the basis of acoustic parameters (e.g., Lindová, Špinka, & Nováková, 2015; Oller, Buder, Ramsdell, Warlaumont, & Chorna, 2013; Papousek, 1992; Scheiner, Hammerschmidt, Jürgens, & Zwirner, 2002). Oller et al. (2013), for instance, investigated the relation between the speech sounds produced by 3- to 4-month-old infants and the emotional meanings they seemed to express with them. The authors found that infants used cries to express mostly a negative emotion and laughs to express a positive emotion, and, interestingly, that other types of speech sounds (squeals, vowel-like sounds, and growls) were used indistinguishably to express positive, negative, and neutral emotions. This functional flexibility had also been observed in previous studies focusing on the acoustical properties of infants' emotional vocalizations. One study reported, for example, that 2-month-old infants use variations in pitch height, pitch range, and duration in a flexible way to indicate positive or negative emotions (Scheiner et al., 2002), while an earlier one noted that these acoustic properties increase with increasing negative arousal (Papousek, 1992).

The flexible use of early vocal sounds is claimed to signal an infant's first step towards an adult linguistic system. Crucially, from a very early stage, infants carry out this de-functionalization by controlling the prosodic properties of speech. This early control of prosodic parameters is the precursor of later more sophisticated prosodic patterns in infants' vocalizations which allow them to convey intentionality and make basic speech act distinctions. In the next section we review the literature that addresses these issues.

Early production of prosody as a marker of intentional communication

Goldstein, Schwade, and Bornstein (2009) tested 5-month-old infants in a task where an experimenter interacted normally with infants and then suddenly stopped responding to the infants' vocalizations. When the experimenter interrupted the interaction in this fashion, the infants increased their vocalization rate. This showed that as early as 5 months infants use speech as a social tool to try to modify their environment. As we will see in the following paragraphs, the prosody of such early vocalizations reflects whether or not an infant's speech is intentional and goal-directed.

Beaumont and Bloom (1993) asked adults to judge 3-month-old infants' vocalizations as being an attempt to interact socially or not. The authors selected two types of vocalizations for the judgment task, consonant-vowel syllabic sounds with pitch variation, and vocalic sounds with uniform pitch. Results showed that raters judged syllabic vocalizations as being more intentional and socially favorable, and thus demonstrated that adults assign intentionality to infants' vocalizations on the basis of how infants modulate prosodic features (and specifically, pitch variation). However, the fact that at 3 months of age infants have not yet developed the ability to engage in joint attentional frames and produce goal-directed actions (e.g., Woodward, 1998, 1999) suggests that the intentionality perceived in infant vocalizations may be simply an adult mental construction reflecting adults' propensity to automatically project intention on infant behavior rather than true evidence that infants can modify prosodic features to achieve a specific goal.

Some months later, between 7–12 months of age, infants start to engage in patterns of joint attention and are able to produce goal-directed actions (as we saw in previous sections of this chapter), a sign that they can communicate intentionally with others. In this time period, infants have been shown to control prosodic features like pitch range and duration in their own productions to differentiate among emotionally expressive vocalizations, goal-directed vocalizations, and vocalizations which have no specific goal. Papaeliou, Minadakis, and Cavouras (2002) examined whether 7- to 11-month-old infants used different duration, fundamental frequency (F0), and intensity values during vocalizations to express intentionality. Infants at all ages were found to control these prosodic features, with emotional vocalizations being longer and showing a higher F0 and lower intensity than vocalizations with pragmatic intent. The results of an interesting follow-up study by Papaeliou and Trevarthen (2006) revealed that intentional vocalizations (defined as those produced in a playing-together situation) were shorter and had a higher pitch range than non-intentional vocalizations (defined as those produced in a playing-alone situation). The expression of emotional and intentional speech (as well as the ability to establish and maintain a state of joint attention) constitutes the pragmatic antecedent of speech act distinctions, which will be discussed in the next subsection.

Early production of prosody as a marker of speech act information

There is evidence that young infants are able to specify the intention of their communicative acts to their interlocutors. Most of the results come from studies examining non-speech communicative acts by young infants. Bates, Camaioni, and Volterra (1975) were among the first to propose that before the emergence of their first spoken words infants express proto-versions of speech acts by using deictic gestures. Since then abundant evidence has been gathered showing that 12-month-old infants point with a specific pragmatic intention in mind, be it to ask for information (interrogative pointing), request an object or action (requestive pointing), share their interest in something (expressive pointing), or inform another person about something that might be relevant (informative pointing) (e.g., Tomasello et al., 2007).

In this context, various recent studies have shown that, like deictic gestures, prosodic features such as pitch and duration are used by young infants to signal speech act distinctions (see Cameron-Faulkner, 2014, for a review of research on the development of speech acts in infancy). For example, Esteve-Gibert and Prieto (2013) examined the prosodic markers of speech act motivation in video recordings of more than 2,000 vocalizations by pre-lexical infants at 7, 9, and 11 months of age. To do so, the pitch range and duration values of the vocalizations were analyzed. After classifying vocalizations as being either intentional or non-intentional, the authors then further classified intentional vocalizations as either expressing emotions like discontent or satisfaction, or signaling a speech act like response, statement, or request. The study yielded two main results. First, the types of pragmatic meaning changed with age: while at younger ages there were a high number of non-intentional productions, intentional vocalizations became more frequent as the infants grew older. Notably, it was found that the proportion of non-intentional vocalizations decreased significantly between 9 and 11 months of age. Furthermore, at younger ages the expressions of discontent and satisfaction were the most frequent pragmatic functions, but by the time the infants were 11 months of age the frequency of statements, requests, and responses had substantially increased. Second, the results showed that non-intentional vocalizations were much longer and more monotonic (with a narrower pitch range) than intentional vocalizations, and that expressions of discontent and requests showed a significantly wider pitch range and longer duration than responses and statements.

The direction of the intonation contour is also used by pre-lexical infants to signal speech act information. Here the evidence comes from studies analyzing the intonation contours that accompany pointing gestures. For example, Grünloh and Liszkowski (2015) tested 14-month-old Dutch-learning infants by eliciting imperative (requestive) and declarative (expressive and informative) pointing acts in laboratory-controlled situations and then analyzed the prosodic and gestural characteristics of the infants' behavior. They found that infants produced flat intonation (defined as having less than 2 semitones of difference between a high and a low F0 target) with requestive pointing gestures, and rising contours with declarative (i.e., expressive and informative) pointing gestures. Murillo and Capilla (2016) explored a longitudinal corpus in which 9- to 15-month-old L1 Spanish-learning infants were recorded during semi-structured play. They found that, at the age of 15 months, infants produced flat intonations for declarative functions and rising contours for imperative functions. Although these two studies did not detect the same patterns, this might be explained by age differences in the respective subject populations, language-specific effects (given that Spanish and Dutch have a different distribution of intonation contours per pragmatic function), or even methodological issues (since it is more difficult to identify the possible social intention of pointing gestures when they are produced spontaneously compared with when they occur in a controlled shared-action context). In any case, what the two studies reveal is that, beyond the social cues, the prosodic features of vocalizations (together with the pointing gestures used) give clues to pre-lexical infants' underlying intentions.

Infants continue to gain competence in the use of prosodic features (and specifically, intonation contours) to signal speech act type and pragmatic meaning as their lexical and socio-cognitive skills develop further. Prieto, Estrella, Thorson, and Vanrell (2012) analyzed the intonation patterns produced by four Catalan-speaking children and two Spanish-speaking children between 11 and 28 months. Their results revealed that the children's use of intonation was largely independent of grammatical development, and it emerged well before the appearance of two-word combinations. Both Catalan- and Spanish-speaking children produced the basic pragmatically appropriate and phonologically differentiated intonation contours (i.e., the contours used to signal vocatives, statements, requests, and expressive intentions) of their respective ambient languages between 12 and 15 months of age, that is, during the period when some infants go from the onset of word production to having a small lexicon of 25 words (see Armstrong & Hübscher, this volume; Chen, this volume; Frota & Butler, this volume).

All in all, the results reported in this section reveal that infants start to command the prosodic properties of speech very early on, well before they produce their first words. Evidence suggests that pre-lexical infants use prosody to express their emotional states, and some months later they can use it to communicate intentionally with the world around them and even signal the specific intention they want to transmit.

Conclusions and future directions

The mapping between prosody and meaning constitutes a central ability in early language acquisition, an ability which is important for successful social interactions. The overview presented in this chapter has revealed that infants' sensitivity to prosodic features emerges very early in their lives and that these early perception skills are employed to begin mapping prosodic patterns onto the pragmatic meanings intended by others.

An important conclusion from the studies reviewed in this chapter is that very young infants use prosodic patterns to express and comprehend meaning well before they use lexical and grammatical marking. We do not yet know, though, whether higher pragmatic prosodic skills in infancy will be predictive of lexical and grammatical skills later in development, in other words, whether pragmatic prosody is paving the way for lexical and grammatical acquisition in later stages or not. If this were the case, it would suggest that prosody may have a bootstrapping effect on the development of pragmatics (see Hübscher, Esteve-Gibert, Igualada, & Prieto, 2017, for a proposal in this regard). Indeed, while the role of prosody as a facilitator of syntactic bootstrapping has been highlighted in language acquisition research (see de Carvalho et al., this volume), the role of prosody as a pragmatic decoder has been neglected. On a related note, researchers interested in gestural development have already found evidence that children's use of gestures seems to favor their acquisition of linguistic skills (e.g., Iverson & Goldin-Meadow, 2005; see Goldin-Meadow & Alibali, 2013, for a review).

An important research topic which merits further study is the relationship between the emergence of prosodic abilities and infants' later development of linguistic and grammatical skills, and, more specifically, whether the acquisition of prosodic meaning can serve as a predictor of later language abilities. Despite the evidence of considerable variability at the individual level, there have been several attempts to relate the emergence of phonological and communicative strategies to children's subsequent linguistic abilities. Igualada, Bosch, and Prieto (2015), for instance, found evidence that the multimodal ability of 12-month-old children to appropriately combine pointing gestures with speech to express an assertive speech act is predictive of later lexical and syntactic abilities at 18 months. For their part, McGillion et al. (2017) found that babble onset predicted the emergence of first words and infants' expressive vocabulary at 18 months of age, and that the onset of pointing predicted infants' receptive vocabulary at 18 months of age. Future studies could try to correlate the emergence of infants' ability to use and understand prosodic cues with their later communicative skills, with the goal of assessing the possible role of prosody in predicting later linguistic and cognitive impairments. In this regard, one line of research that could be strengthened is the study of atypical populations, in which specific cognitive or linguistic abilities are impaired (see Peppé, this volume, for a review of the research on prosodic development in such populations).

From a developmental perspective, the evidence points to a tight relation between the pragmatic meanings that infants can access through prosody at a certain developmental stage and the cognitive skills that emerge at that point in time (see Matthews, 2014, for a review). Thus, at a very early stage infants are able to interpret prosody in adult speech only to distinguish between basic emotions such as fear, anger, or happiness. Some months later, once they have learned that actions can be goal-directed and that people act intentionally towards objects and events, infants can comprehend the intentional value of somebody else's action (that is, they can tell whether it is intentional or accidental) by processing the prosodic cues of that person's speech. And as soon as infants can comprehend that actions can have specific and varying goals (e.g., pointing at an object can have different purposes), they show evidence that they associate specific prosodic patterns with certain simple speech acts like making a request, offering information, or expressive a feeling. However, 12-month-olds are still not able to understand and produce speech acts that involve complex mental skills. As we have noted, first-order false-belief representations (the realization that somebody can hold a belief that may differ from one's own) emerge during the preschool years (Baron-Cohen et al., 1985), and second-order false belief representations (the realization that somebody can hold a false belief of someone else's false belief about something) do not appear until age 5 or later (Miller, 2009). It stands to reason that only when children have developed these more complex mental skills can they make use of prosody to understand and signal belief states (see Armstrong & Hübscher, this volume).

Furthermore, the directionality of the relation between cognitive and linguistic development is still not fully understood. Some results certainly suggest that cognitive skills are a precursor of and/or a prerequisite for linguistic skills. As we have noted, for example, Sakkalou et al. (2013) found that infants with higher cognitive skills at 13 months of age are the earliest to use prosody to map intentions with actions at 14 months of age. However, other accounts propose an inverse causal relationship, suggesting it is linguistic skills that impact and determine infants' development of cognitive abilities, especially at later stages of language development (see De Villiers, 2007, for a discussion of this debate). The studies reported so far indicate that at least at the pre-lexical stage the emergence of cognitive skills coincides with or slightly precedes the emergence of pragmatic prosody. Further studies could assess the potential correlations between individual differences in cognitive skills (and also Theory of Mind abilities related to an understanding of mental states) and the development of pragmatic prosody.

The evidence outlined in this chapter has also suggested a parallel temporal development between infants' early comprehension and production of pragmatic prosody. First, around 5 months of age infants react by smiling to prosody conveying positive affect, such as when a mother expresses approval (Fernald, 1993). At around the same age, infants begin to produce vocal cues to express emotions in a flexible way, so that the same vocal cues can be used to express different emotions (Oller et al., 2013). Later on, around their first birthday, infants understand that prosody (and speech in general) is used by their interlocutors as a sign of intentional communication (Carpenter et al., 1998; Sakkalou & Gattis, 2012; Vouloumanos et al., 2012). Simultaneously, as they gain control over prosodic cues like duration and pitch range, they begin to employ it for this purpose themselves (Esteve-Gibert & Prieto, 2013; Papaeliou et al., 2002; Papaeliou & Trevarthen, 2006). At about the same time (12 months), infants learn to decode prosodic signals to distinguish between informative, requestive, and expressive attention-directing acts (Esteve-Gibert et al., 2017), and they begin to signal such simple forms of speech acts in their own prosodic output (Esteve-Gibert & Prieto, 2013; Grünloh & Liszkowski, 2015; Murillo & Capilla, 2016).

Broadening the experimental procedures and methodologies employed in the acquisition of the prosody-meaning interface is essential to make progress in the field. On the one hand, more production studies should be conducted in laboratory-controlled settings, where shared-action contexts and act-preceding interactions can be precisely controlled and infant subjects' social intent thus more clearly determined. Larger sample sizes with more infants should be tested, thereby identifying potential individual effects, and bi- and multilingual populations could be further explored. On the other hand, online processing techniques like eye-tracking or neuroimaging can tap into infants' implicit language and cognitive skills before they are able to demonstrate them explicitly (see Teixidó et al., this volume, for a review). Using such techniques will be crucial to investigate the point at which infants start to make the connection between the suprasegmental cues they perceive and their intended meanings. All in all, a variety of empirical techniques can be deployed in a complementary fashion to assess infants' understanding of prosody as their linguistic and cognitive capacities unfold.

References

Armstrong, M. E. (2012). *The development of yes-no question intonation in Puerto Rican Spanish*. Unpublished PhD dissertation. Ohio State University.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "Theory of Mind"? *Cognition*, 21, 37–46. doi: 10.1016/0010-0277(85)90022-8

Bates, E., Camaioni, L., & Volterra, V. (1975). The acquisition of performatives prior to speech. *Merrill-Palmer Quarterly of Behavior and Development*, 21(3), 205–226.

Beaumont, S. L., & Bloom, K. (1993). Adults' attributions of intentionality to vocalizing infants. *First Language*, 13, 235–247. doi: 10.1177/014272379301303805

Berman, J. M. J., Chambers, C. G., & Graham, S. A. (2016). Preschoolers' real-time coordination of vocal and facial emotional information. *Journal of Experimental Child Psychology*, 142, 391–399. doi: 10.1016/j.jecp.2015.09.014

Blasi, A., Mercure, E., Lloyd-Fox, S., Thomson, A., Brammer, M., Sauter, D., Deeley, Q., Barker, G. J., Renvall, V., Deoni, S., Gasston, D., Williams, S. C. R., Johnson, M. H., Simmons, A., & Murphy, D. G. M. (2011). Early specialization for voice and emotion processing in the infant brain. *Current Biology*, 21(14), 1220–1224. doi: 10.1016/j.cub.2011.06.009

Bruner, J. S. (1975). The ontogeny of speech acts. *Journal of Child Language*, 2(1), 1–19. doi: 10.1017/S0305000900000866

Cameron-Faulkner, T. (2014). The development of speech acts. In D. Matthews (Ed.), *Pragmatic development in first language acquisition* (pp. 37–52). Amsterdam: John Benjamins.
Carpenter, M., Akhtar, N., & Tomasello, M. (1998). Fourteen- through 18-month-old infants differentially imitate intentional and accidental actions. *Infant Behavior & Development*, 21(2), 315–330. doi: 10.1016/S0163-6383(98)90009-1
Clark, E. V. (2004). Pragmatics and language acquisition. In L. Horn, & G. Ward (Eds.), *The handbook of pragmatics* (pp. 562–577). Malden, MA: Blackwell.
Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, 13(4), 148–153. doi: 10.1016/j.tics.2009.01.005
Davis, B. L., MacNeilage, P. F., Matyear, C. L., & Powell, J. K. (2000). Prosodic correlates of stress in babbling: An acoustical study. *Child Development*, 71(5), 1258–1270. doi: 10.1111/1467-8624.00227
DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers voices. *Science*, 208,1174–1176. doi: 10.1126/science.7375928
De Villiers, J. (2007). The interface of language and Theory of Mind. *Lingua. International Review of General Linguistics. Revue Internationale de Linguistique Générale*, 117(11), 1858–1878.
Esteve-Gibert, N., Liszkowski, U., & Prieto, P. (2016). Prosodic and gestural features distinguish the intention of pointing gestures in child-directed communication. In M. E. Armstrong, N. Henriksen, & M. M. Vanrell (Eds.), *Intonational grammar in Ibero-Romance. Approaches across linguistic subfields* (pp. 249–276). Amsterdam: John Benjamins.
Esteve-Gibert, N., & Prieto P. (2013). Prosody signals the emergence of intentional communication in the first year of life: Evidence from Catalan-babbling infants. *Journal of Child Language*, 40(5), 919–944. doi: 10.1017/S0305000912000359
Esteve-Gibert, N., Prieto, P., & Liszkowski, U. (2017). Twelve-month-olds understand social intentions based on prosody and gesture shape. *Infancy*, 22(1), 108–129. doi: 10.1111/infa.12146
Fernald, A. (1993). Approval and disapproval: Infant responsiveness to vocal affect in familiar and unfamiliar languages. *Child Development*, 64, 657–674. doi: 10.2307/1131209
Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development*, 8(2), 181–195. doi: 10.1016/S0163-6383(85)80005-9
Frota, S., Butler, J., & Vigário, M. (2014). Infants perception of intonation: Is it a statement or a question? *Infancy*, 19(2), 194–213. doi: 10.1111/infa.12037
Goldin-Meadow, S., & Alibali, M. W. (2013). Gestures role in speaking, learning, and creating language. *Annual Review of Psychology*, 64, 257–283. doi: 10.1146/annurev-psych-113011-143802
Goldstein, M. H., Schwade, J. A., & Bornstein, M. H. (2009). The value of vocalizing: Five-month-old infants associate their own noncry vocalizations with responses from caregivers. *Child Development*, 80(3), 636–644. doi: 10.1111/j.1467-8624.2009.01287.x
Grossmann, T., Oberecker, R., Koch, S. P., & Friederici, A. D. (2010). The developmental origins of voice processing in the human brain. *Neuron*, 65(6), 852–858. doi: 10.1016/j.neuron.2010.03.001
Grünloh, T., & Liszkowski, U. (2015). Prelinguistic vocalizations distinguish pointing acts. *Journal of Child Language*, 42(6), 1312–1336. doi: 10.1017/S0305000914000816
Gussenhoven, C. (1984). *On the grammar and semantics of sentence accents*. Dordrecht: Foris.
Höhle, B., Bijeljac-Babic, R., Herold, B., Weissenborn, J., & Nazzi, T. (2009). Language specific prosodic preferences during the first half year of life: Evidence from German and French infants. *Infant Behavior & Development*, 32(3), 262–274. doi: 10.1016/j.infbeh.2009.03.004

Hoicka, E., & Wang, S. H. (2011). Fifteen-month-old infants match vocal cues to intentional actions. *Journal of Cognition and Development*, 12(3), 299–314. doi: 10.1080/15248372.2010.542215

Hübscher, I., Esteve-Gibert, N., Igualada, A., & Prieto, P. (2017). Intonation and gesture as bootstrapping devices in speaker uncertainty. *First Language*, 37(1), 24–41.

Igualada, A., Bosch, L., & Prieto, P. (2015). Language development at 18 months is related to communicative strategies at 12 months. *Infant Behavior and Development*, 39, 42–52. doi: 10.1016/j.infbeh.2015.02.004

Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science*, 16(5), 367–371. doi: 10.1111/j.0956-7976.2005.01542.x

Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, 40(1–2), 1–19. doi: 10.1016/0010-0277(91)90045-6

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others beliefs in human infants and adults. *Science*, 330, 1830–1834. doi: 10.1126/science.1190792

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

Lindová, J., Špinka, M., & Nováková, L. (2015). Decoding of baby calls: Can adult humans identify the eliciting situation from emotional vocalizations of preverbal infants? *PLoS ONE*, 10(4), 1–16. doi: 10.1371/journal.pone.0124317

Liszkowski, U. (2013). Using Theory of Mind. *Child Development Perspectives*, 7(2), 104–109. doi: 10.1111/cdep.12025

Liszkowski, U. (2014). Two sources of meaning in infant communication: Preceding action contexts and act-accompanying characteristics. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 369, 20130294. doi: 10.1098/rstb.2013.0294

Mampe, B., Friederici, A. D., Christophe, A., & Wermke, K. (2009). Newborns cry melody is shaped by their native language. *Current Biology*, 19(23), 1994–1997. doi: 10.1016/j.cub.2009.09.064

Martin, A., Onishi, K. H., & Vouloumanos, A. (2012). Understanding the abstract role of speech in communication at 12 months. *Cognition*, 123(1), 50–60. doi: 10.1016/j.cognition.2011.12.003

Matthews, D. (Ed.). (2014). *Pragmatic development in first language acquisition*. Amsterdam: John Benjamins.

McGillion, M., Herbert, J. S., Pine, J., Vihman, M. M., Depaolis, R., Keren-Portnoy, T., & Matthews, D. (2017). What paves the way to conventional language? The predictive value of babble, pointing, and socioeconomic status. *Child Development*, 88(1), 156–166.

Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology*, 31(5), 838–850. doi: 10.1037/0012-1649.31.5.838

Miller, S. A. (2009). Childrens' understanding of second-order mental states. *Psychological Bulletin*, 135(5), 749–773. doi: 10.1037/a0016854

Mumme, D. L., Fernald, A., & Herrera, C. (1996). Infants' responses to facial and vocal emotional signals in a social referencing paradigm. *Child Development*, 67, 3219–3237. doi: 10.2307/1131775

Murillo, E., & Capilla, A. (2016). Properties of vocalization- and gesture-combinations in the transition to first words. *Journal of Child Language*, 43(4), 890–913.

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. doi: 10.1037/0096-1523.24.3.756

Oller, D. K., Buder, E. H., Ramsdell, H. L., Warlaumont, A. S., & Chorna, L. (2013). Functional flexibility of infant vocalization and the emergence of language. *PNAS*, 110(16), 6318–6323. doi: 10.1073/pnas.1300337110

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false belief? *Science*, 308, 255–258. doi: 10.1126/science.1107621

Papaeliou, C. F., & Trevarthen, C. (2006). Prelinguistic pitch patterns expressing "communication" and "apprehension". *Journal of Child Language*, 33(1), 163–178. doi: 10.1017/S0305000905007300

Papaeliou, C., Minadakis, G., & Cavouras, D. (2002). Acoustic patterns of infant vocalizations expressing emotions and communicative functions. *Journal of Speech, Language, and Hearing Research*, 45(2), 311–317. doi: 10.1044/1092-4388(2002/024)

Papousek, M. (1992). Early ontogeny of vocal communication in parent-infant interactions. In H. Papousek, U. Jurgens, & M. Papousek (Eds.), *Nonverbal vocal communication: Comparative and developmental approaches* (pp. 230–261). Cambridge: Cambridge University Press.

Portes, C., Beyssade, C., Michelas, A., Marandin, J. M., & Champagne-Lavau, M. (2014). The dialogical dimension of intonational meaning: Evidence from French. *Journal of Pragmatics*, 74, 15–29. doi: 10.1016/j.pragma.2014.08.013

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 1, 515–526. doi: 10.1017/S0140525X00076512

Prieto, P. (2015). Intonational meaning. *Wiley Interdisciplinary Reviews: Cognitive Science*, 6, 371–381.

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. doi: 10.1017/S030500091100002X

Saint-Georges, C., Chetouani, M., Cassel, R., Apicella, F., Mahdhaoui, A., Muratori, F., Laznik, M. C., & Cohen, D. (2013). Motherese in interaction: At the cross-road of emotion and cognition? (A systematic review). *PloS One*, 8(10), e78103. doi: 10.1371/journal.pone.0078103

Sakkalou, E., Ellis-Davies, K., Fowler, N. C., Hilbrink, E. E., & Gattis, M. (2013). Infants show stability of goal-directed imitation. *Journal of Experimental Child Psychology*, 114(1), 1–9. doi: 10.1016/j.jecp.2012.09.005

Sakkalou, E., & Gattis, M. (2012). Infants infer intentions from prosody. *Cognitive Development*, 27(1), 1–16. doi: 10.1016/j.cogdev.2011.08.003

Scheiner, E., Hammerschmidt, K., Jürgens, U., & Zwirner, P. (2002). Acoustic analyses of developmental changes and emotional expression in the preverbal vocalizations of infants. *Journal of Voice*, 16(4), 509–529. doi: 10.1016/S0892-1997(02)00127-3

Snow, D. (2006). Regression and reorganization of intonation between 6 and 23 months. *Child Development*, 77(2), 281–296. doi: 10.1111/j.1467-8624.2006.00870.x

Sullivan, K., Zaitchik, D., Tager-Flusberg, H. (1994). Preschoolers can attribute second-order beliefs. *Developmental Psychology*, 30, 395–402. doi: 10.1037/0012-1649.30.3.395

Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A new look at infant pointing. *Child Development*, 78(3), 705–722. doi: 10.1111/j.1467-8624.2007.01025.x

Vaillant-Molina, M., Bahrick, L. E., & Flom, R. (2013). Young infants match facial and vocal emotional expressions of other infants. *Infancy*, 18(1), 1–15.

Vouloumanos, A., Onishi, K. H., & Pogue, A. (2012). Twelve-month-old infants recognize that speech can communicate unobservable intentions. *PNAS*, 109(32), 12933–12937. doi: 10.1073/pnas.1121057109

Wermke, K., Teiser, J., Yovsi, D., Kohlenberg, P. J., Wermke, P., Robb, M., Keller, H., & Lamm, B. (2016). Fundamental frequency variation within neonatal crying: Does ambient language matter? *Speech*, *Language and Hearing*, 19(4), 211–217.

Wilutzky, W. (2015). Emotions as pragmatic and epistemic actions. *Frontiers in Psychology*, 6, 1593. doi: 10.3389/fpsyg.2015.01593

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, 69(1), 1–34. doi: 10.1016/S0010-0277(98)00058-4

Woodward, A. L. (1999). Infants ability to distinguish between purposeful and non-purposeful behaviors. *Infant Behavior & Development*, 22(2), 145–160. doi: 10.1016/S0163-6383(99)00007-7

Yeung, H. H., Chen, K. H., & Werker, J. F. (2013). When does native language input affect phonetic perception? The precocious case of lexical tone. *Journal of Memory and Language*, 68(2), 123–139. doi: 10.1016/j.jml.2012.09.004

Zhang, D., Liu, Y., Hou, X., Sun, G., Cheng, Y., & Luo, Y. (2014). Discrimination of fearful and angry emotional voices in sleeping human neonates: A study of the mismatch brain responses. *Frontiers in Behavioral Neuroscience*, 8, 422. doi: 10.3389/fnbeh.2014.00422

CHAPTER 13

Gradual development of focus prosody *and* affect prosody comprehension

A proposal for a holistic approach

Kiwako Ito
Ohio State University

Excellence in communication skills requires an ability to appropriately represent the discourse structure including focus, as well as good comprehension of speaker affect. Both focus and affect are communicated in large part through prosody, so comprehension and production of the accompanying prosody is essential. However, past studies on focus prosody have been both theoretically and methodologically separated from the research on affect prosody. (In this chapter, I use the term 'focus prosody' to refer to prosodic phenomena that are either produced or perceived as the cue to a specific part of speech that conveys the focal content of a message. This includes 'narrow focus', which is defined in terms of the informational scope (e.g., answers to Wh-questions), and 'contrastive focus', which is a subtype of narrow focus that evokes interpretational alternatives.) This chapter argues that the suggested difference in the developmental trajectory (i.e., focus prosody develops slower as compared to affect prosody) may be an artifact of the perspective divergence, and points out that the mastery of prosodic skills in both these domains must be necessarily *gradual* – though they may not develop hand-in-hand. A holistic approach that considers the interaction between affect prosody and focus prosody is proposed as a future direction of the research on prosodic development within and across individuals.

Introduction

Prosodic prominence provides dynamicity in speech acts at multiple levels of communication: it may accompany facial expressions and gestures for displaying affect during face-to-face interaction, convey the speaker's mood, level of excitement or formality in the presence or absence of visual cues, reflect the speaker's attitude toward the topic of a conversation or the listener, and signal the

DOI 10.1075/tilar.23.13ito

relative importance of particular parts of speech, indicating to the listener what is worth paying attention to and what is a simple reminder or a confirmation. Not surprisingly, understanding prosodic development in children has been a challenge for linguists, psychologists, speech therapists, neuroscientists and psychiatrists due to its intertwined relationship to individuals' social cognition and pragmatic knowledge. Experts in various scientific disciplines have studied prosody to describe its linguistic structure (Beckman, 1996; Beckman & Pierrehumbert, 1986; Jun, 2005; Ladd, 2008), to identify neuro-anatomical specificity of brain function (Belyk & Brown, 2013; Grossmann, Striano, & Friederici, 2005; Ross, 1981; Wildgruber, Ackermann, Kreifelts, & Ethofer, 2006), or to achieve better machine recognition of speaker's emotions (Fernandez & Picard, 2011; Oudeyer, 2003; Schuller et al., 2010; Schuller, Steidl, & Batliner, 2009). Researchers have also tried to assess the prosodic skills in individuals with communication disorders and identify the similarities and differences in their processing mechanisms as compared to those in their typically-developing peers (Edwards, Pattison, Jackson, & Wales, 2001; Hargrove, 2013; Ito & Martens, 2017; Peppé, 2009, this volume; Peppé & McCann, 2003; Wells, Peppé, & Goulandris, 2004). While efforts in each of these fields have certainly advanced our understanding of prosody in the last few decades, the extreme theoretical and methodological diversity in their approaches may have prevented the construction of a model of prosodic development that incorporates empirical evidence gathered across this wide range of research fields.

In the past, I have argued that the acquisition of contrast-marking prosody takes time, and have proposed some accounts as to why (Ito, 2014; Ito, Bibyk, Wagner, & Speer, 2014a; Ito, Jincho, Minai, Yamane, & Mazuka, 2012). This chapter reviews a wider range of related research to advance this view and adds a claim that while comprehension of focus prosody develops gradually through childhood (and possibly through adulthood), so do other abilities that involve perception and processing of prosodic prominence – such as affect recognition. I argue that accurate mapping of prosodic prominence onto message structure takes many years to develop, whether it is for computing the discourse structure (e.g., focus) or for identifying the intention or emotion of the speaker (e.g., affect). This is because human oral communication gains complexity and sophistication through social experience, yet its elements are susceptible to cognitive constraints and informational screening. The intentions and nuances that can be expressed via prosody may receive more or less weight according to the reliability of the cues produced by the speaker (Kurumada, 2013) and the listeners' cognitive capacities such as attention or memory (Fraundorf, Watson, & Benjamin, 2010, 2012). In addition, listeners may tune differently to particular aspects of prosodic cues according to their percept of the discourse context and communicative demands (Kurumada, Brown, Bibyk, Pontillo, & Tanenhaus, 2014; for speaker's prosodic adjustment according to informational predictability, see Turnbull, 2016). While processing of prosodic prominence in any discourse context involves both the computation of informational weight and the computation of speaker's emotional status, there has been a clear division between these subfields of prosody research: when listeners' comprehension of focus prosody is studied, their understanding of speaker affect is rarely discussed, whereas when listeners' emotion recognition is investigated, how they represent the informational structure tends to be neglected. To date, research on the interaction of so-called 'linguistic' prosody and affect prosody is extremely sparse (Pihan, Tabert, Assuras, & Borod, 2008). However, simultaneous consideration of those two aspects is critical for modeling prosodic development, as correct identification of speaker's affect may facilitate the comprehension of informational foci and vice versa.

While the ability to perceive affect in speech seems to start developing from a preverbal stage (Flom & Bahrick, 2007; Grossmann et al., 2005; Walker-Andrews & Grolnick, 1983; Walker-Andrews & Lennon, 1991; see Esteve-Gibert & Prieto, this volume, for an updated summary), the ability to use prosodic cues to comprehend the informational structure seems to take a much slower, gradual developmental trajectory (Cruttenden, 1985; Cutler & Swinney, 1987; Grassmann & Tomasello, 2007, 2010; Ito, 2014; Ito et al., 2012, 2014; Sekerina & Trueswell, 2012; Wells et al., 2004). This apparent contrast in the developmental time course, however, may have resulted from differences in the theoretical assumptions and research directions across these 'sub-fields'. The first section of the present chapter reviews how scholars have framed their research questions and described their study implications to elicit two distinct traditions: while scientists who study affect prosody focus on *how early children start exhibiting their sensitivity to vocal cues to speaker emotion*, those who study focus prosody attempt to demonstrate *when children start representing the referential statuses of linguistic entities (e.g., morphemes, words and phrases) in an adult-like manner*. There, I argue that a particular assumption about cognitive development (i.e., that understanding affect prosody must precede comprehending prosodic cues to informational structure) may have driven research on affect prosody to focus on infancy and studies on focus prosody to explore later developmental stages.

Following a summary of the theoretical trends, the second section of the chapter discusses the methodological constraints that may have led to the contrast in the research outcome – early affect prosody and late focus prosody. Here, I emphasize that we must carefully evaluate the age-appropriateness of tasks and the meta-linguistic skills required by the tasks before making conclusions about the onset and completion of prosodic development. The third section of the chapter will discuss the cognitive factors that underlie the development of prosodic semantics and processing efficiency. It will also explore the danger in assuming prosodic categories that are composed of fixed sets of acoustic cues that directly map onto semantic categories (whether affect or informational status). In the final section, I will propose a holistic approach that investigates how children's perception of speaker affect influences their discourse structure representation. In this approach, individual differences in the executive function as well as social cognition must be considered together. The chapter will conclude with a set of questions and research topics for future studies.

Theoretical division: Fundamental affect function vs. complex information-structuring function of prosody?

Studies on affect prosody development often emphasize how essential the processing of affect is for the general development of social cognition and communication skills. An example comes from the first line of the article by Vaillant-Molina, Bahrick, and Flom (2013): "Perception of emotional expressions is fundamental to social development." The view that the sensitivity to emotional expressions and its early comprehension are universal seems to have collected support from cross-linguistic studies on infant-directed speech (Fernald, 1985, 1993, 2004; Fernald et al., 1989; Fernald & Mazzie, 1991; Katz, Cohn, & Moore, 1996), as well as from modern brain science which provides evidence for spontaneous neuronal responses to affective prosody in both adults and children (Ethofer et al., 2012, 2009; Grandjean et al., 2005; Grossmann et al., 2005). Research on primates' distinctive responses to affective vocalization (e.g., Owren & Rendall, 1997; Scherer & Kappas, 1988) has also supported the view that the ability to learn to recognize emotion in voice may have an evolutionarily traceable genetic basis. In a review of studies on affect prosody, Grossmann et al. (2005: 1828) state, "it has been proposed that enhanced sensory responses to emotional facial and vocal stimuli might be a fundamental neural mechanism". This echoes the earlier claim by Walker-Andrews (1997: 437): "although processing only rudimentary capacities to detect, discriminate, and recognize others' emotional expressions, the human infant is born well prepared to rapidly develop these competencies during the first year." Developmental language researchers are generally driven by the question of 'how early children start responding to linguistic stimuli' and are expected to advance methodologies for detecting the behavioral traits earlier than previously reported. It is therefore not surprising that many studies of affective vocal perception test infants younger than 5 months to examine whether they can respond to changes in affect prosody (Flom & Bahrick, 2007; Vaillant-Molina, Bahrick, & Flom, 2013; Walker, 1982; Walker-Andrews, 1997).

In comparison, the field of research on focus prosody seems to be less bound to this 'how early' question. This may be due to the notion that the understanding of prosody that signals the information structure of words (such as novelty or givenness, narrow focus and contrast) must await the development of abilities to segment words, allocate attention to relevant referents, and represent referential status of linguistic entities. In other words, we tend to think that understanding which words bear relative importance in a given discourse demands a wider range of cognitive prerequisites than understanding affect. Numerous studies have shown that the use of language-specific rhythmic structure for word segmentation develops in the latter half of the first year (Bull, Eilers, & Oller, 1984; Eilers, Bull, Oller, & Lewis, 1984; Höhle, Bijeljac-Babic, Herold, Weissenborn, & Nazzi, 2009; Jusczyk, 1997; Jusczyk, Cutler, & Redanz, 1993; Nazzi, Jusczyk, & Johnson, 2000; see also Frota & Butler, this volume, for a more detailed review). This coincides with the phase where infants exhibit the ability to extract transitional probabilities (Marcus, Vijayan, Bandi Rao, & Vishton, 1999; Saffran, Aslin, & Newport, 1996) and gain further sensitivity to their native phonemic categories while losing sensitivity to foreign categories (Kuhl et al., 2006; Werker & Tees, 2002). Interestingly, a shift in the gaze more to the mouth than to the eyes takes place in this narrow time window as well (Tomalski, 2015). Thus, infants who will start producing one-word utterances in a few months are busy forming their native phonemic categories, learning the language-specific rhythmic patterns and phonotactics, and mapping the sound strings to visually identifiable oral gestures. While this seems a lot to deal with for their very young brains, they may be able to handle more. In theory, infants who are learning to detect word boundaries in running speech may be simultaneously developing the ability to compute gross informational weight for parts of speech that bear prosodic prominence. While young infants become capable of tuning to durational, pitch, and spectral cues for segmenting words, they may be able to benefit from exaggerated, emphatic cues to more clearly represent the referential relationships.

Fernald and Mazzie (1991) once hypothesized that focus prosody, which tends to mark word newness, must bear a critical role in lexical learning in infants who cannot rely on a rich semantic network or contextual knowledge. Fernald's research on infant-directed speech (IDS) emphasizes the effect of IDS on word learning via infants' preferences for motherese (Cooper & Aslin, 1990; Fernald & Kuhl, 1987). Later studies provided direct and indirect evidence for this view. For example, Thiessen, Hill, and Saffran (2005) confirmed that statistical learning of word-forming syllables is facilitated by IDS. Ma, Golinkoff, Houston, and Hirsh-Pasek (2011) showed that 21-month-old children learned novel words better with IDS than with adult-directed speech (ADS), a finding which was replicated in 17-month-olds by Graf Estes and Hurley (2013). Importantly, Graf Estes and

Hurley (2013) demonstrated that IDS was effective only when target words exhibited prosodic variability across tokens: mere repetitions of identical IDS tokens did not lead to word learning, and performance was no better than for children tested with ADS. Shukla, White, and Aslin (2011) showed that 6-month-old infants gazed longer at an object that was labeled with a disyllabic string (e.g., mu:ra:) when the string was grouped into the same intonational phrase (IP) than when it straddled an IP boundary. In addition, Sakkalou and Gattis (2012) revealed that 18-month-old English-learning toddlers better followed the actions labeled by foreign words (Greek "*Ochi*" and "*Nato*") when they were presented with prosody that expressed a speaker's intention to guide them (i.e., a tune that is similar to what is produced with the English directive utterance "There.") than when they were accompanied by prosody that expressed an accidental event (the tune that accompanies "Whoops"). Thus, studies suggest that children younger than two years are not only sensitive to various prosodic cues (on this topic see Esteve-Gibert & Prieto, this volume; Frota & Butler, this volume), but they are also able to *use* prosody to allocate their attention to novel events and learn object names. What remains unclear is whether they can use prosodic prominence to *represent discourse structure* even at a very rudimentary level (e.g., whether some entities or events are more worth paying attention to than others for the current communicative purpose).

As for the development of the ability to map prosodic prominence to discourse structure, studies show mixed results regarding when it emerges and how it is mastered. The age of participants also has varied widely, and the conclusion as to whether children can or cannot comprehend focus prosody seems largely dependent on the experimental tasks and measures. In one of the early production studies, Hornby and Hass (1970) tested 4-year-old children, who had not been explicitly taught the use of contrastive prosody in school. Because their participants used contrastive stress more often for the contrastive element of scenes they were asked to describe than for other non-contrastive elements, Hornby and Hass concluded that production of contrastive stress was mastered by 4 years of age. Hornby (1971) put 1st, 3rd , and 5h graders (approximately 6, 8, and 10 years old) through a task where participants must provide an explicit correction of the experimenter's scene narration, and found that the use of contrastive stress reduced as children gained a wider range of syntactic means to express focus (such as passive and cleft structures). Later, a study by MacWhinney and Bates (1978) reported that discourse newness, which was elicited by a switched element across otherwise identical scenes (e.g., the recipient in a scene where a person gives something to another person), increased the use of sentential stress in children between ages 3 and 5 years. Their results also showed that English-speaking children used stress a lot more often than Italian and Hungarian-speaking children, who preferred to use other means such as word order to express newness. The steady increase of stress with age was found only in English-speaking children, indicating that the frequency of focus prosody in child language may depend on the language-specific repertoire of focus expressions children develop with age. Recent production studies summarized by Frota and Butler (this volume) and by Chen (this volume) suggest that the developmental trajectory of prosodic production is even more complex than what these early studies depicted decades ago. These studies show that both language-specific constraints and individual variability impact the use of particular phonetic cues, while leaving little room to dispute the early onset of functional prosodic production in child speech. The detection of early onset, however, should not be interpreted as the evidence of immediate mastery.

As for the comprehension of focus prosody, the findings from Solan (1980), Cutler and Swinney (1987), Cruttenden (1974, 1985), and Wells et al. (2004) all point to difficulties for preschoolers and school-age children in understanding the meaning of sentential stress. These findings therefore provide counter-evidence to the general observation of language development that comprehension precedes production. In fact, this paradox of focus prosody acquisition was explicitly discussed by Cruttenden (1985) and Cutler and Swinney (1987). Cruttenden (1985) acknowledges the studies by Hornby and Hass (1970) and Crystal (1979) that argue for an early mastery of production of focus prosody, yet claims that his experimental results, in which 10-year-old children underperformed adults in comprehension of prosody, "dispel the myth that children master the adult intonation system very early in their linguistic life [which Cruttenden had already claimed in 1974] p. 657)". Cruttenden defies Bolinger's (1978) view of 'intonation as an innate gesture-like reflex' as rather too strong and argues that certain intonation meanings that require mature grammatical and contextual knowledge (e.g., 'indignation' meaning for a fall-rise with a word 'might') should appear later in production. Cruttenden also points out that children's underperformance in some comprehension tasks (e.g., distinction between *hotdog* and *hot dog*) does not necessarily reflect their lack of knowledge (of stress assignment), but may instead show their uncertainty in how to weigh various cues available at the moment. Cutler and Swinney (1987) also argue that children younger than age 5 to 6 have yet to develop the skill to map discourse structure onto linguistic structure, and that truly intentional use of focus prosody must await the development of semantic and pragmatic knowledge: "Only once this [referring to the development of discourse representation] has occurred can the prosodic production system approximate the adult system, in which the underlying physiological basis has become 'socialized' (p. 163)." Considering Bolinger's view mentioned above, Cutler and Swinney suggest that the early production of sentential stress or focus prosody is 'qualitatively different' from later productions that are based on the computation of discourse-level factors.

In sum, earlier studies that discussed the discrepancy between children's production and comprehension of focus prosody emphasized that the integration of semantic and pragmatic knowledge with the prosodic structure may not be acquired early, because sufficient development of those components is necessary before they can be linked effectively. Since we cannot address when children acquire adult-like skills to express and interpret prosodic meaning without a fair grasp of adults' prosody and its function, researchers' attention may have been drawn more toward a finer-grained description of adult prosody and simultaneously drifted away from the development of focus prosody during early childhood. In contrast, researchers of affect prosody tend to assume that perception of prosodic cues to emotion must develop (and be mastered) early, and thus they have not explored how children gain prosodic means to express and understand emotion in later stages of development. However, both tracing the symptomatic behavior of information structuring in early developmental stages and investigating the processing of affect in later developmental stages are essential for achieving a more accurate description of prosodic development within and across individuals. The methodological challenges are, of course, non-trivial.

Methodological division: Passive tasks to test affect detection vs. interactive tasks to test focus comprehension

Due to the focus on early infancy, the majority of traditional investigations of affect prosody in infancy has used preferential looking paradigms (Flom & Bahrick, 2007; Vaillant-Molina et al., 2013; Walker-Andrews, 1998; Walker-Andrews & Grolnick, 1983; Walker-Andrews & Lennon, 1991). There, the experimental effects are inferred from the gaze duration and its proportion, which is assumed to reflect the degree of infants' interest in the stimuli. Since the values of these dependent measures themselves do not reveal the processing mechanisms that lead to the differences across conditions, statistical inference must be necessarily indirect: for example, if tested infants spent 54% of stimuli presentation time attending to the target in condition A and 46% of time in condition B and their difference is statistically significant, researchers may conclude, despite the overall quasi-chance-level performances, that infants preferred condition A to condition B or that they at least discriminated the tested affect categories. Designs of infant studies are typically constrained by the small number of trials, and the interpretation of small effect sizes requires caution due to close-to-chance-level baselines. Research with preschoolers that investigates the mapping between prosodic tunes and affect types also tends to be constrained by a small number of stimuli (see Armstrong & Hübscher, this volume, for a review of studies testing children aged

3 years and older). In many studies, the experimental task often forces one-to-one mapping between prosodic tunes and labels or facial expressions (e.g., Berman, Chambers, & Graham, 2010, 2016). While this seems a feasible strategy, forced-choice tasks entail a problem of methodological adequacy because the prosody-to-meaning mapping is 'context-dependent and defeasable' in nature (Hirschberg, 2002), and the performance may rely on the participants' understanding of the experimenter's intention (see below for a discussion on a similar problem with focus prosody research). If the task measures the skill for building mutually exclusive links between stimulus sets and response options within an experiment, the findings may or may not directly reflect children's spontaneous interpretations of the prosodic cues in question.

The indirectness of inference also applies to a neurolinguistic approach. For example, a study with the event-related brain potential (ERP) technique by Grossmann et al. (2005) reports larger positive wave shifts for the temporal electrodes for happy and angry prosody as compared to neutral prosody, and negative wave shifts for angry, as compared to happy and neutral prosody for the frontal-central sites in 7-month-old infants. These differences in the size of brainwave components show that the infants responded to the prosodic changes and indicate that they may attend to a specific category of valence differently. However, the differences in the size or direction of brainwaves do not show whether the infants recognized the emotion, i.e., whether the happy and angry prosody were *interpreted* by the infants as the expression of happiness and anger. Thus, implications of the studies with preverbal infants are often limited to the indication of the sensitivity to prosodic manipulation.

In contrast, methods for testing focus prosody research in older children may be easier to evaluate against study hypotheses, because unlike deducing affect recognition from indirect physiological measures such as gaze duration and ERPs, the semantics of prosody can be more directly observed with interactive tasks. Solan (1980), Cruttenden (1985), Cutler and Swinney (1987), and Wells et al. (2004) all employed tasks that required behavioral responses from children that can be coded as either accurate or inaccurate. Except for Cutler and Swinney who used a word-monitoring task, these studies used visual stimuli or real world objects that had to be selected or evaluated according to speech input. Solan (1980) used toy animals in a game where the experimenter and a child participant took turns in acting out the pre-recorded narratives (e.g., *The camel hit the lion, and then HE hit the elephant*). While 5-, 6-, and 7-year-old children all performed correctly when they heard the stress on the pronoun (e.g., by grabbing the lion), 5-year-olds performed incorrectly when the unstressed pronoun was expected to lead to a parallel interpretation (while 6- and 7-year-olds grabbed the previous agent camel upon hearing *he*, 5-year-olds grabbed the lion). Since the contrastive stress did not change the behavior in 5-year-olds, Solan (1980 : 696) speculates that they are in a phase in which they "assume that reciprocity governs events in the world". While this proposal remains to be empirically attested, I suspect that this outcome may also have to do with a difference in task comprehension between the youngest and older children. It is possible that the youngest children simply thought that the animals were supposed to take turns in this particular game, and thus focused on the action to be repeated and paid little attention to prosody. Once they believed that they had learned the rule of the game, the perseverance tendency in younger children (Trueswell, Sekerina, Hill, & Logrip, 1999) may have made them respond in a consistent manner for the remaining trials. The older children, in contrast, may have better guessed what the experimenter expected them to do in the task: differentiate actions according to different speech input.

While measures of interactive tasks may more directly speak to the research questions than indirect physiological measures, earlier studies on focus prosody (e.g., Cruttenden, 1985; Wells et al., 2004) were not free of methodological problems. Since I have discussed them elsewhere (Ito, 2014; Ito et al., 2012, 2014), I refrain from repeating the details here. In a nutshell, children have shown difficulty selecting the picture that matches the intended interpretation of the spoken stimuli, yet their poor performance may have resulted from (1) a failure to comprehend out-of-the-blue prosodic prominence that was presented in isolated sentences, and (2) a problem linking the speaker's intention to one of the multiple contrastive relations among the visual prompts, which were susceptible to ranking according to visual salience (for example, a boy holding four oranges makes a more salient contrast with a girl holding four bananas than with a girl holding two oranges, thus the former picture set is more likely to lure children's attention when they hear *John's got FOUR oranges*). Importantly, although successful performances in these interactive tasks may suggest children's correct contrastive interpretation of the prominence, their incorrect responses do not necessarily indicate a lack of contrastive interpretation. It is possible that children noticed the emphasis in the narrative, but decided to weigh visual salience more. Ironically, these interactive tasks for gauging the semantics of prosody in older children are not adequate for detecting their sensitivity to prosody, which may be identified quickly by passive preferential looking paradigms.

While performance in offline interactive tasks often depends on child participants' interpretation of the task, using eye-tracking techniques such as the visual-world paradigm can overcome the problem of individual differences in meta-linguistic skills and improve methodological adequacy for investigating children's responses to the presence or absence of prosodic prominence (Arnold, 2008; Ito et al., 2012, 2014; Sekerina & Trueswell, 2012). The primary advantage of the visual-world paradigm is its capacity to trace participants' spontaneous reactions to speech input before they follow the commands or make selections for a visual search (Trueswell & Tanenhaus, 2005, for a summary). Visual-world studies commonly report that child participants (aged between 4 and 11 years) respond to prosodic manipulations within a short period of time after the critical speech input. For example, Arnold (2008) detected fixations to the anaphoric (or already mentioned) visual target within 1000 ms from the onset of the unaccented target word in 4- and 5-year-old children. Sekerina and Trueswell (2012) found facilitation of target detection for nouns that immediately followed accented color adjectives in 6-year-old Russian-speaking children. Ito and colleagues found both facilitative and misleading effects of prominence on pre-nominal color adjectives within 400–600 ms after the offset of the adjectives, in Japanese-speaking (Ito et al., 2012) as well as English-speaking (Ito et al., 2014) 6-year-old children: a prominent adjective in a sequence that repeated the noun (pink cat → GREEN cat) led to faster detection of the target animal set (cat), whereas the prominence in the sequence that switched both the adjective and the noun (purple lion → ORANGE monkey) led to initial fixations on the previously mentioned animal set (lion), which resulted in a slower detection of the target set (monkey). In eye-tracking data, the timing of fixations can reveal participants' sensitivity to prosodic (as opposed to segmental) cues, while the direction of fixations can determine whether prosodic prominence is appropriately interpreted as the cue to contrast. In Ito et al. (2014), the robust immediate effect of prominent accent was confirmed in all age groups (6- and 7-year-olds, 8- and 9-year-olds, 10- and 11-year olds, and adults), but importantly, the fixation timings of child participants approached those of adults *gradually* with age, and even the oldest child group (10- and 11-year-olds) were not as swift as adults. In addition, the oldest group's recoveries from the misguided fixations were clearly delayed as compared to those of adults. For example, while adults' initial incorrect fixations to the lion cell in (purple lion → ORANGE monkey) peaked at the midpoint of the noun 'monkey' and decreased from there on, the incorrect fixations kept increasing throughout the noun in 10- and 11-year-olds.

In sum, eye-tracking techniques with better temporal resolution have advanced our understanding of prosodic processing in children: the data demonstrate immediate responses to prosodic cues, spontaneous detection of particular visual targets that reflects the interpretation of the cues, and a clear effect of age on processing efficacy. The developmental trajectory that these findings depict, however, is not very different from the claims of three decades ago: comprehension of focus prosody may emerge early, but takes time to develop (Cruttenden, 1985; Cutler & Swinney, 1987). To date, use of interactive eye-tracking paradigms such as Arnold (2008), Sekerina and Trueswell (2012), and Ito et al. (2012, 2014) to investigate younger children (e.g., 2- and 3- year olds) is rather sparse, because experimental materials are limited by such factors as the child participants' vocabulary size, ability to comprehend instructions, and attention span. Future studies must explore whether using a smaller number of within-subject conditions with between-subject designs can overcome the problems of data loss and low statistical power that typically challenge the studies with toddlers.

Even though the use of online methods has enabled fine-grained examination of spontaneous responses to prosodic cues in children, such studies have yet to overcome a methodological pitfall that is common across experimental paradigms: the use of acted or read speech. In most of the aforementioned studies on affect prosody and focus prosody, participants were presented with a set of carefully handpicked speech stimuli produced by an actor or trained phonetician. This has been the methodological norm, in order to overcome the problem of using live face-to-face interactions that do not control for prosodic consistency within and across experimental sessions (e.g., Cruttenden, 1985). Using pre-recorded speech stimuli facilitates objective assessment of prosodic skills and reduces inter-experimenter and environmental variability (e.g., computerized version of Profiling Elements of Prosodic Systems-Child (PEPS-C), originally developed by Peppé and McCann (2003) and advanced by Peppé and colleagues, see http://www.peps-c.com/). However, as long as we use non-spontaneous stimuli, we cannot escape from the question of whether the experimental findings are generalizable to the daily oral communication that children experience. One strategy we may adopt is the use of measurable materials from spontaneous speech corpora. Some studies have used speech from pre-existing corpora and tested the prominence ratings of naïve adult listeners (Cole, Mo, & Hasegawa-Johnson, 2010; Turnbull, Royer, Ito, & Speer, 2017), as well as their eye-movement responses to tree-decoration instructions (Ito, Turnbull, & Speer, 2017), in order to confirm listeners' sensitivities to natural speech. With the growing number of open resources such as CHILDES <http://childes.psy.cmu.edu/>, and advanced recording devices such as the LENA system <https://www.lena.org/>, experimentation with more naturalistic (yet controlled) stimuli is certainly becoming feasible.

In sum, affect prosody and focus prosody have been studied with different experimental paradigms, largely due to differences in the age of targeted populations. Measures of responses to affect prosody in pre-verbal children are necessarily indirect (e.g., gazes and brainwaves), whereas responses to focus prosody can be more directly observed with interactive tasks with older children. While applications of a visual-world paradigm have demonstrated children's immediate interpretation of contrastive focus prosody and its gradual development, experimental paradigms can be further advanced with more naturalistic discourse speech materials.

Slow development of prosodic skills and slow development of developmental theory: Why does it take so long?

To explain the factors underlying the gradual developmental trajectory of contrastive prosody comprehension, and Ito et al. (2012, 2014) have proposed data-driven accounts that extend the views of Cruttenden (1985) and of Cutler and Swinney (1987). First, these studies found that 6-year-olds, both Japanese- and English-speaking, have a perseveration tendency that makes them take time to shift attention from the previous referential set to the new referential set. Thus, the efficacy of prosodic processing seems tightly related to the development of attention allocation that controls the speed of discourse representation (e.g., new/less accessible vs. old/accessible), which in turn affects the effectiveness of prosody-to-discourse mapping. Second, the fact that children take longer to recover from prosodic garden-pathing indicates that they are not as efficient as adults in detecting conflicts in the signal, revising prosody-based expectations, and letting segmental information guide the referential resolution. This requires inhibition of salient percepts of contrastive prosody and a switch of attention to segmental cues, which demands general cognitive flexibility. Executive function that includes attention allocation, inhibition, and cognitive flexibility is known to develop slowly throughout childhood and adolescence (Prencipe et al., 2011; Zelazo & Müller, 2002). Thus, it is reasonable to assume that it takes years to achieve the level of ability that mature processors have to integrate prosodic cues with referential information and to revise analyses when necessary.

Assuming that executive function has an intertwined relationship with various aspects of oral communication skills, it does not make sense to believe that the ability to process affect prosody is mastered in early childhood. As summarized in Armstrong and Hübscher (this volume), studies show that young children rely more on salient lexical information, facial expressions and gestural cues than on prosodic cues for interpreting affect, and even older children often exhibit difficulty understanding a speaker's emotion expressed by prosody. These findings may reflect an overall shortage of cognitive resources (such as memory and attention span) for multi-dimensional information processing in children, and their tendency to ignore less direct and more variable cues. A cognitive capacity constraint on audio-visual processing is also suggested by unique eye-movement patterns in infants with Autism Spectrum Disorder, who look less at the face when speech signals accompany the visual stimuli than when they are presented silently (Shic, Macari, & Chawarska, 2014). As for the link between memory function and general language skills, a study by Vulchanova, Foyn, Nilsen, and Sigmundsson (2014) reports strong correlations between verbal working memory capacity (measured with a forward digit recall task) and grammar, vocabulary, and L2 spoken sentence comprehension measures in 10-year-old Norwegian children. Individual differences in verbal working memory are also known to affect adult sentence processing of complex syntactic structures such as object-extracted relative clauses (Fedorenko, Gibson, & Rohde, 2006; King & Just, 1991). While prosody conveys information about lexical identity, phrase structure, discourse context and social dynamics, it is likely that processing of signals related to information weight or not-so-evident expressions of affect is compromised in processors with limited capacity. Children must learn to attend to referential contexts and social factors (such as who is talking to whom) while holding on to the linguistic contents of utterances during conversation. The ability to quickly map prosodic cues to context-dependent informational structures and complex affect statuses must develop, to some degree, with the growth of cognitive resources.

Another fundamental factor that tends to be oversimplified in research on affect prosody is the granularity of affect categories. While many studies of infant affect detection test the distinction between canonical categories such as happy/positive vs. sad/negative (Grossmann et al., 2005; Vaillant-Molina, Bahrick, & Flom, 2013; Walker-Andrews & Grolnick, 1983), real-world communication often requires much finer distinctions of emotions along the valence and arousal dimensions (e.g., tired vs. bored, miserable vs. depressed, delighted vs. pleased, content vs. calm, etc.: Russell, 1980). Research has shown that 5-year-olds can reliably recognize happy and sad faces yet have difficulty identifying expressions for fear, disgust and anger (Durand, Gallay, Seigneuric, Robichon, & Baudouin, 2007). Complex affect categories beyond these five basic emotions are therefore predicted to develop even later.

As for the auditory processing of affect, Flom and Bahrick (2007) have shown that 5- and 7-month-old infants discriminate happy, sad, and angry prosodies, and Vaillant-Molina, Bahrick, and Flom (2013) found that 5-month-olds can correctly map prosody to positive and negative facial expressions. It is yet to be discovered, however, when children acquire other affect categories such as fear and disgust and recognize them with particular prosodic cues. A study by Demenescu, Kato, and Mathiak (2015) showed that adults recognize happiness and anger better than sadness, disgust and fear in vocal expressions, and negative emotional recognition generally deteriorates with aging. Their fMRI data suggest that the reduction of sensory function (in the superior temporal gyrus) may underlie this decline. While the development of sensory function is considered as the primary neurological component of automatic affect detection, another recent study by Voyer, Thibodeau, and Delong (2014) showed that the interpretation of sarcasm in adults heavily relies on context as well as prosody. Thus, processing of affect prosody requires both quasi-automatic perception of acoustic cues (such as energy, voice quality and tune) and a fair grasp of discourse context, which may also rely on executive function abilities such as working memory and attention.

Finally, for the construction of a more empirically adequate theory of prosodic development, it is important to remember that different affect categories are expressed by different acoustic cues, which are not uniformly salient across different types and levels of emotions (Banse & Scherer, 1996; Juslin & Laukka, 2001; Laukka, Juslin, & Brestin, 2005; Sauter, Eisner, Calder, & Scott, 2010). For example, while Banse and Schere (1996) show that panic, hot anger, happiness, sadness are respectively expressed with increased mean F0, mean energy, low frequency energy, and duration, Juslin and Laukka (2001) reveal that emotional intensity affects the use of the same acoustic cues differently for different emotion types: strong emotional intensity boosts the F0 floor for both fear and happiness, but it boosts the F0 expansion and F0 maxima only for fear but not for happiness, which already reaches the speaker's ceiling F0 level with weak emotional intensity. Sauter et al. (2010) provide a diagram of which acoustic cues are mainly used for which types of emotion (e.g., pitch for surprise, anger, achievement, and relief but not for amusement, sadness, contentment, pleasure and disgust, which are expressed by spectrum and amplitude instead. Fear is expressed solely by spectrum according to this diagram). However, these cues do not always discern between different affect types: Sauter et al.'s (2010) experimental data demonstrate that both human listeners and discriminant analysis of acoustic measures tend to confuse anger with disgust, and contentment with pleasure or relief. Furthermore, while humans may misidentify fear as amusement (13%), discriminant analysis of acoustics mistakes fear for achievement (31%) and anger (25%).

These studies point to the fact that prosodic characterization of affect types is difficult, and acoustic cues can be unreliable and misleading at times. Based on the currently available empirical data, we may not want to assume any fixed bundle of acoustic features for any given affect category, as an expression of a particular emotional status can be achieved via various combinations of multi-modal cues. We may in fact achieve a better scientific description of human cognition by identifying the conditions under which particular prosodic features are processed as dominant cues to the observed emotional status rather than by seeking evidence for prosodic categories that directly map onto affect categories. Affect is a primary source for changes in speech rate, overall intensity, voice quality, and other articulartory gestures as well as body language and facial expressions of the speaker: these behavioral signals in turn lead to the percept of a specific emotional status in the listener, who may not necessarily share the ranges of valence and arousal with the speaker. Thus, affect, either as the speaker's emotional status or the listener's percept, can never be defined by a finite set of absolute prosodic cues. This basic observation is often overlooked in the studies of affect prosody acquisition. To advance the theory of prosodic development, we must walk away from assumptions about direct mappings between affect categories and prosodic categories (for the acoustic difference between lab speech and spontaneous speech that lead to similar contrastive interpretation, see Ito et al., 2017).

In this section, I reviewed studies that suggest affect recognition is not easy once it requires more than a simple discrimination between happiness and sadness. I argue that maturity of executive functions underlies the slow development of both focus prosody and affect prosody, which are orchestrated flexibly with cues expressed through other modalities according to each communicative purpose and context. Importantly, we should bear in mind that seeking for a set of acoustic cues that invariably labels a type of speech act – whether it is for marking information structure or recognizing affect – is not a very rewarding approach, as compared to efforts to identify the mechanism of cue tuning or balancing.

Interaction and integration of affect prosody and focus prosody

At the beginning of this chapter, I suggested a direction of future studies that considers developmental trajectories of affect prosody and focus prosody simultaneously. This was in fact inspired by a comment from an individual with Williams syndrome (WS), who participated in our eye-tracking study (Ito, Martens, & McKenna, 2014b, March): "This person sounds very loud. I felt she was scolding me". The stimuli were identical to those of Ito et al. (2014), for which a young female phonetician produced questions such as "*Where is the pink monkey? Now, where is the GREEN monkey?*" imagining talking to a young child. To everybody up to that point, she sounded like a happy preschool teacher rather than a grumpy lady, and none of the typically developing child participants had expressed discomfort with her voice. While the above comment from an adult participant with WS was thus unexpected, it certainly provided food for thought. Individuals with WS are known to be hypersensitive to social cues (Dykens, 2003), and thus this participant may have paid particular attention to cues to speaker affect. However, this may also happen in everyday conversations among typically developing individuals. If rudimentary sensitivity to affect prosody develops during infancy, young children may well be capable of automatically detecting basic affect status (e.g., positive vs. negative) while attending to prosodic signals related to discourse structure.

To date, studies that simultaneously examine more than one function of prosody have been sparse. Pihan et al. (2008) asked participants to listen to a pair of sentences and indicate which one sounds more like a question. Three statements were produced by two speakers (one male, one female) in happy, neutral and fearful prosody, and their pitch contours were artificially modified such that each sentence ends with a rising, falling or level tone. Pihan et al. (2008) found that the rising tones generally increased the 'question' responses. However, the happy prosody that was characterized by the largest F0 changes throughout the utterances interfered with the speaker intention judgments, resulting in the lowest 'question' responses within the rising tone set. (The predicted right-lateralization for happy and fear prosody was not detected in the EEG signals.) The findings of Pihan et al. (2008) suggest that unconsciously detected speaker emotion can affect the perception of illocutionary force. Thus, while laboratory experimenters typically make participants attend to particular dimensions of speech signals, listeners may assess the emotional state of a speaker automatically and make responses accordingly. This implication needs to be considered carefully for the study of focus prosody in toddlers and infants. On one hand, the function of accentuation in highlighting specific discourse entities may be blurred by happy or overly excited prosody with exaggerated pitch excursions. On the other hand, accents embedded in a cold angry tone with a compressed pitch range may be more efficiently processed if they stand out acoustically, although some listeners may interpret the emphasis as part of the expression of anger.

The interaction between affect and focus prosody may also modulate the effect of memory encoding in children. Fraundorf et al. (2010, 2012) have shown that contrastive accent leads to better recall of narratives in both younger and aging adults. Lee and Snedeker (2016) have partially replicated this effect of contrastive accent in 5-year-old children. Another recent study by Lee and Fraundorf (2016) confirmed some sensitivity to contrastive accent only in high-proficiency, but not in low-proficiency, L1-Korean learners of English. This finding adds to a general observation of difficulty in using prosodic cues for interpreting the speaker's intention in language learners (see Armstrong & Hübscher, this volume, for a summary of research on children's comprehension of prosody for belief state). A recent study by Igualada, Esteve-Gibert, and Prieto (2017) reports the effect of 'beat gestures' (gestures that highlight a part of speech) on the recall of verbal information in 3- to 5-year-old children, suggesting that informational retrieval can be facilitated by associated visual cues. Since this finding demonstrates preschoolers' ability to bind cues for retrieving spoken information, we may hypothesize that prosodic prominence can also function as a retrieval cue in young children. However, the critical question is whether prosodic cues alone can flag the discourse status of a referential expression and facilitate the retrieval of it. Based on the results of Ito et al. (2014), I doubt that prosodic prominence is as effective as visually-associated cues for memory retrieval in young children, because segmental and prosodic cues must compete for the limited cognitive resources available for auditory processing. In addition, if the listener happens to attend to the emotional state of the speaker while processing the segmental information for syntactic and lexical semantic structures of the spoken message, the resources for auditory input (or the 'phonological loop' component in Baddeley and Hitch's (1974) model of working memory) may be overloaded with an underdeveloped processor. This 'limited resource hypothesis' needs to be tested with careful experimental designs, because particular emotional percepts may encourage or discourage memory encoding as well as informational representation in listeners. Caution is required when preparing experimental stimuli, because accents may not stand out in an overly expanded pitch range that may express a high level of arousal (as in IDS), while a specific combination of acoustic cues for expressing a particular degree of valence may sharpen or dilute the effect of prosodic emphasis.

While these hypotheses about the interaction of affect prosody and focus prosody with limited cognitive resources remain to be explored, the research outcome for such questions would be very beneficial not only to the field of developmental intonational phonology, but also to the research fields of educational psychology, language pedagogy and communication disorders. One common goal across these applied research fields is to find a way to improve individuals' communication skills. There is little room to debate whether excellence in communication skills comprises of good understanding of speaker emotion and a fair comprehension of discourse context and structure. Since prosody provides robust cues to both these components, developmental research in this area should take a more holistic approach in which the bi-directional interaction between affect prosody and focus prosody is examined across multiple developmental stages. This approach should investigate whether and how children's percept of speaker affect impacts their understanding of message structure, and how this interaction changes as they acquire fine-grained affect categories and discourse representations. It is also important to examine how efficacy of sensory processing, sensitivity to social cues, cognitive flexibility, and memory impact the way emotion recognition influences message comprehension. Although it is impossible to assess all these inter-related abilities of an individual in one study, strategic experimental designs and inter-disciplinary collaborations that draw on multiple expertises have strong potential to overcome methodological challenges and advance our understanding of communication development.

Acknowledgements

I thank the two anonymous reviewers who gave me very encouraging and constructive comments.

References

Arnold, J. E. (2008). THE BACON not the bacon: How children and adults understand accented and unaccented noun phrases. *Cognition*, 108, 69–99. doi:10.1016/j.cognition.2008.01.001

Baddeley, A. D., & Hitch, G. (1974). Working memory. *The Psychology of Learning and Motivation: Advances in Research and Theory*, 8, 47–89. doi:10.1016/S0079-7421(08)60452-1

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, 70, 614–636. doi:10.1037/0022-3514.70.3.614
Beckman, M. E. (1996). The parsing of prosody. *Language and Cognitive Processes*, 11, 17–67. doi:10.1080/016909696387213
Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English, *Phonology Yearbook*, 3, 255–309. doi:10.1017/S095267570000066X
Belyk, M., & Brown, S. (2013). Perception of affective and linguistic prosody: An ALE meta-analysis of neuroimaging studies. *Social Cognitive and Affective Neuroscience*, 9, 1395–1403. doi:10.1093/scan/nst124
Berman, J. M., Chambers, C. G., & Graham, S. A. (2010). Preschoolers' appreciation of speaker vocal affect as a cue to referential intent. *Journal of Experimental Child Psychology*, 107(2), 87–99. doi:10.1016/j.jecp.2010.04.012
Berman, J. M., Chambers, C. G., & Graham, S. A. (2016). Preschoolers' real-time coordination of vocal and facial emotional information. *Journal of Experimental Child Psychology*, 142, 391–399. doi:10.1016/j.jecp.2015.09.014
Bolinger, D. (1978). Intonation across languages. In J. H. Greenberg, C. A. Ferguson & E. Moravcsik (Eds.), *Universals of human language: Phonology* (Vol. 2, pp. 471–524). Stanford, CA: Stanford University Press.
Bull, D. H., Eilers, R. E., & Oller, D. K. (1984). Infants' discrimination of intensity variation in multisyllabic stimuli. *Journal of the Acoustical Society of America*, 76, 13–17. doi:10.1121/1.391110
Chen, A. (2014). Production-comprehension Asymmetry: Individual differences in the acquisition of prosody focus-marking. *Proceedings of the 7th International Conference on Speech Prosody* – Dublin, 423–427.
Cole, J., Mo, Y., & Hasegawa-Johnson, M. (2010). Signal-based and expectation-based factors in the perception of prosodic prominence. *Laboratory Phonology*, 1(2), 425–452. doi:10.1515/labphon.2010.022
Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development*, 61, 1584–1595. doi:10.2307/1130766
Cutler, A., & Swinney, D. A. (1987). Prosody and the development of comprehension. *Journal of Child Language*, 14, 145–167. doi:10.1017/S0305000900012782
Cruttenden, A. (1974). An experiment involving comprehension of intonation in children from 7 to 10. *Journal of Child Language*, 1, 221–232. doi:10.1017/S0305000900000660
Cruttenden, A. (1985). Intonation comprehension in ten-yea-olds. *Journal of Child Language* 12, 643–661. doi:10.1017/S030500090000670X
Crystal, D. (1979). Prosodic development. In P. Fletcher, & M. Garman (Eds.), *Language acquisition: Studies in first language development* (pp. 33–48). Cambridge: Cambridge University Press.
Edwards, J., Pattison, P. E., Jackson, H. J., & Wales, R. J. (2001). Facial affect and affective prosody recognition in first-episode schizophrenia. *Schizophrenia Research*, 48, 235–253. doi:10.1016/S0920-9964(00)00099-2
Demenescu, L. R., Kato, Y., & Mathiak, K. (2015). Neural processing of emotional prosody across the adult lifespan. *BioMed Research International*. Article ID 590216.
Durand, K., Gallay, M., Seigneuric, A., Robichon, F., & Baudouin, J. Y. (2007). The development of facial emotion recognition: The role of configural information. *Journal of Experimental Child Psychology*, 97, 14–27. doi:10.1016/j.jecp.2006.12.001
Dykens, E. M. (2003). Anxiety, fears, and phobias in persons with Williams syndrome. *Developmental Neuropsychology*, 23, 291–316. doi:10.1080/87565641.2003.9651896
Eilers, R. E., Bull, D. H., Oller, D. K., & Lewis, D. C. (1984). The discrimination of vowel duration by infants. *Journal of the Acoustical Society of America*, 75, 1213–1218. doi:10.1121/1.390773

Ethofer, T., Bretscher, J., Gschwind, M., Kreifelts, B., Wildgruber, D., & Vuilleumier, P. (2012). Emotional voice areas: anatomic location, functional properties, and structural connections revealed by combined fMRI/DTI. *Cerebral Cortex*, 221, 191–200. doi:10.1093/cercor/bhr113

Ethofer, T., Kreifelts, B., Wiethoff, S., Wolf, J., Grodd, W., Vuilleumier, P., & Wildgruber, D. (2009). Differential influences of emotion, task, and novelty on brain regions underlying the processing of speech melody. *Journal of Cognitive Neuroscience*, 217, 1255–1268. doi:10.1162/jocn.2009.21099

Fedorenko, E., Gibson, E., & Rhode, D. (2006). The nature of working memory in linguistic, arithmetic, and spatial integration processes. *Journal of Memory and Language*, 56, 246–269. doi:10.1016/j.jml.2006.06.007

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development*, 8, 181–195. doi:10.1016/S0163-6383(85)80005-9

Fernald, A. (1993). Approval and disapproval: Infant responsiveness to vocal affect in familiar and unfamiliar languages. *Child Development*, 64, 657–674. doi:10.2307/1131209

Fernald, A. (2004). Hearing, listening, and understanding: Auditory development in infancy. In G. Bremner, & A. Fogel (Eds.), *Blackwell handbook of infant development* (pp. 35–70). Oxford: Blackwell. doi:10.1002/9780470996348.ch2

Fernald, A., & Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development*, 10, 279–293. doi:10.1016/0163-6383(87)90017-8

Fernald, A., & Mazzie, C. (1991. Prosody and focus in speech to infants and adults. *Developmental Psychology*, 27(2), 209–221. doi:10.1037/0012-1649.27.2.209

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*, 16(3), 477–501. doi:10.1017/S0305000900010679

Fernandez, R., & Picard, R. (2011). Recognizing affect from speech prosody using hierarchical graphical models. *Speech Communication*, 53, 1088–1103. doi:10.1016/j.specom.2011.05.003

Flom, R., & Bahrick, L. E. (2007). The development of infant discrimination of affect in multimodal and unimodal stimulation: The role of intersensory redundancy. *Developmental Psychology*, 431, 238–252. doi:10.1037/0012-1649.43.1.238

Fraundorf, S. H., Watson, D. G., & Benjamin, A. S. (2010). Recognition memory reveals just how CONTRASTIVE contrastive accenting really is. *Journal of Memory and Language*, 63, 367–386. doi:10.1016/j.jml.2010.06.004

Fraundorf, S. H., Watson, D. G., & Benjamin, A. S. (2012). The effects of age on the strategic use of pitch accents in memory for discourse: A processing-resource account. *Psychology and Aging*, 27(1), 88–98. doi:10.1037/a0024138

Graf Estes, K., & Hurley, K. (2013). Infant-directed prosody helps infants map sounds to meanings. *Infancy*, 18(5), 797–824. doi:10.1111/infa.12006

Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M., Scherer, K. R., & Vuilleumier, P. (2005). The voices of wrath: Brain responses to angry prosody in meaningless speech. *Nature Neuroscience*, 82, 145–146. doi:10.1038/nn1392

Grassmann, S., & Tomasello, M. (2007). Two-year-olds use primary sentence accent to learn new words. *Journal of Child Language*, 34, 677–687. doi:10.1017/S0305000907008021

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

Grossmann, T., Striano, T., & Friederici, A. D. (2005). Infants' electric brain responses to emotional prosody. *Neuroreport*, 16, 1825–1828. doi:10.1097/01.wnr.0000185964.34336.b1
Hargrove, P. M. (2013). Pursuing prosody interventions. *Clinical Linguistics & Phonetics*, 27, 647–660. doi:10.3109/02699206.2013.777121
Hirschberg, J. (2002). The pragmatics of intonational meaning. *Proceedings of Speech Prosody 2002* – Aix-en-Provence, 65–68.
Höhle, B., Bijeljac-Babic, R., Herold, B., Weissenborn, J., & Nazzi, T. (2009). Language specific prosodic preferences during the first half year of life: Evidence from German and French infants. *Infant Behavior and Development*, 32, 262–274. doi:10.1016/j.infbeh.2009.03.004
Hornby, P. (1971). Surface structure and the topic-comment distinction: A developmental study. *Child Development*, 42, 1975–1988. doi:10.2307/1127600
Hornby, P., & Hass, W. (1970). Use of contrastive stress by preschool children. *Journal of Speech and Hearing Research*, 13, 395–399. doi:10.1044/jshr.1302.395
Igualada, A., Esteve-Gibert, N., & Prieto, P. (2017). Beat gestures improve word recall in 3- to 5- year- old children. *Journal of Experimental Child Psychology,* 156, 99–112.
Ito, K., & Martens, M. (2017). Contrast-marking prosodic emphasis in Williams syndrome: results of detailed phonetic analysis. *International Journal of Language and Communication Disorders*, 52, 46–58.
Ito, K. (2014). Children's pragmatic use of prosodic prominence. In D. Matthews (Ed.), *Pragmatic development in first language acquisition* (pp. 199–218). Amsterdam: John Benjamins.
Ito, K., Bibyk, S., Wagner, L., & Speer, S. R. (2014a). Interpretation of contrastive pitch accent in 6- to 11-year-old English speaking children and adults. *Journal of Child Language*, 41(1), 84–110. doi:10.1017/S0305000912000554
Ito, K., Jincho, N., Minai, U., Yamane, N., & Mazuka, R. (2012). Intonation facilitates contrast resolution: Evidence from Japanese adults & 6-year olds. *Journal of Memory and Language*, 66(1), 265–284. doi:10.1016/j.jml.2011.09.002
Ito, K., Martens, M., & McKenna, E. (2014b, March). *Processing of pitch prominence in Williams syndrome*. Talk presented at 27th Annual CUNY Conference on Human Sentence Processing, Columbus, USA.
Ito, K., Turnbull, R., & Speer, S. R. (2017). Allophonic tunes of contrast: Lab and spontaneous speech lead to equivalent fixation responses in museum visitors. *Laboratory Phonology*, 8(1): 6, 1–29.
Jun, S-A. (Ed.) (2005). *Prosodic typology: The phonology of intonation and phrasing*. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199249633.001.0001
Jusczyk, P. W. (1997). *The discovery of spoken language*. Cambridge, MA: The MIT Press.
Jusczyk, P. W., Cutler, A., & Redanz, N. (1993). Preference for the predominant stress patterns of English words. *Child Development*, 64, 675–687. doi:10.2307/1131210
Juslin, P. N., & Laukka, P. (2001). Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion. *Emotion*, 1, 381–412. doi:10.1037/1528-3542.1.4.381
Katz, G., Cohn, J., & Moore, C. (1996). A combination of vocal F0 dynamic and summary features discriminates between three pragmatic categories of infant-directed speech. *Child Development*, 67, 205–217.
King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, 30, 580–602. doi:10.1016/0749-596X(91)90027-H

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, 9, F13–F21. doi:10.1111/j.1467-7687.2006.00468.x

Kurumada, C. (2013). *Navigating variability in the linguistic signal: Learning to interpret contrastive prosody*. Unpublished PhD dissertation. Stanford University.

Kurumada, C., Brown, M., Bibyk, S., Pontillo, D., & Tanenhaus, M. K. (2014). Is it or isn't it: Listeners make rapid use of prosody to infer speaker meanings. *Cognition*, 133, 335–342. doi:10.1016/j.cognition.2014.05.017

Laukka, P., Juslin, P., & Bresin, R. (2005). A dimensional approach to vocal expression of emotion. *Cognition & Emotion*, 19, 633–653. doi:10.1080/02699930441000445

Lee, E-K., & Fraundorf, S. (2016). *Effects of contrastive accents in memory for L2 discourse*. Bilingualism: Language and Cognition, 20(5), 1063–1079. doi: 10.1017/S1366728916000638

Lee, E-K., & Snedeker, J. (2016). Effects of contrastive accents on children's discourse comprehension. *Psychonomic Bulletin & Review*, 23, 1589–1595.

Ladd, R. D. (2008). *Intonational phonology* (2nd ed.). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511808814

Ma, W., Golinkoff, R. M., Houston, D., & Hirsh-Pasek, K. (2011). Word learning in infant- and adult-directed speech. *Language Learning and Development*, 7, 209–225. doi:10.1080/15475441.2011.579839

MacWhinney, B., & Bates, E. (1978). Sentential devices for conveying givenness and newness: A cross-cultural development study. *Journal of Verbal Learning Verbal*, 17, 539–558. doi:10.1016/S0022-5371(78)90326-2

Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton P. M. (1999). Rule learning by 7-month-old infants. *Science*, 283, 77–80. doi:10.1126/science.283.5398.77

Nazzi, T., Jusczyk, P. W., & Johnson, E. K. (2000). Language discrimination by English-learning 5-month-olds: Effects of rhythm and familiarity. *Journal of Memory and Language*, 43, 1–19. doi:10.1006/jmla.2000.2698

Oudeyer, P-Y. (2003). The production and recognition of emotion in speech: Features and algorithms. *International Journal of Human-Computer Studies*, 59, 157–183. doi:10.1016/S1071-5819(02)00141-6

Owren, M. J., & Rendall, D. (1997). An affect-conditioning model of nonhuman primate vocalizations. In D. W. Owings, M. D. Beecher, & N. S. Thompson (Eds.), *Perspectives in ethology* (Vol. 12, pp. 299–346). New York, NY: Plenum Press.

Peppé, S. J. E. (2009). Aspects of identifying prosodic impairment. *International Journal of Speech-Language Pathology*, 11(4), 332–338. doi:10.1080/17549500903003102

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. doi:10.1080/0269920031000079994

Pihan, H., Tabert, M., Assuras, S., & Borod, J. (2008). Unattended emotional intonations modulate linguistic prosody processing. *Brain and Language*, 105, 141–147. doi:10.1016/j.bandl.2007.08.001

Prencipe, A., Kesek, A., Cohen, J., Lamm, C., Lewis, M. D., & Zelazo, P. D. (2011). Development of hot and cool executive function during the transition to adolescence. *Journal of Experimental Child Psychology*, 108, 621–637. doi:10.1016/j.jecp.2010.09.008

Ross, E. D. (1981). The aprosodias: functional-anatomic organization of the affective components of language in the right hemisphere. *Archives of Neurology*, 38, 561–569. doi:10.1001/archneur.1981.00510090055006

Russell J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39, 1161–1178. doi:10.1037/h0077714
Sakkalou, E., & Gattis, M. (2012). Infants infer intentions from prosody. *Cognitive Development*, 27, 1–16. doi:10.1016/j.cogdev.2011.08.003
Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926–1928. doi:10.1126/science.274.5294.1926
Sauter, D. A., Eisner, F., Calder, A. J., & Scott, S. K. (2010). Perceptual cues in nonverbal vocal expressions of emotion. *The Quarterly Journal of Experimental Psychology*, 63(11), 2251–2272. doi:10.1080/17470211003721642
Scherer, K. R., & Kappas, A. (1988). Primate vocal expression of affective state. In D. Todt, P. Goedeking, & D. Symmes (Eds.), *Primate vocal communication* (pp. 171–194). Berlin: Springer. doi:10.1007/978-3-642-73769-5_13
Schuller, B., Steidl, S., & Batliner, A. (2009). The INTERSPEECH 2009 emotion challenge. *Proceedings of the 10th International Speech Communication Association Annual Conference* – Brighton, 312–315.
Schuller, B., Steidl, S., Batliner, A., Burkhardt, F., Devillers, L., Müller, C., & Narayanan, S., (2010). THE INTERSPEECH 2010 paralinguistic challenge. *Proceedings of the 11th International Speech Communication Association Annual Conference* – Chiba, 2794–2797.
Sekerina, I. E., & Trueswell, J. C. (2012). Interactive processing of contrastive expressions by Russian children. *First Language*, 32(1–2), 63–87. doi:10.1177/0142723711403981
Shic, F., Macari, S., & Chawarska, K. (2014). Speech disturbs face scanning in 6-month-old infants who develop Autism Spectrum Disorder. *Biological Psychiatry*, 75, 231–237. doi:10.1016/j.biopsych.2013.07.009
Shukla, M., White, K. S., & Aslin, R. N. (2011). Prosody guids the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 6038–6043.
Solan, L. (1980). Contrastive stress and children's interpretation of pronouns. *Journal of Speech and Hearing Research*, 23, 688–698. doi:10.1044/jshr.2303.688
Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, 7, 53–71. doi:10.1207/s15327078in0701_5
Tomalski, P. (2015). Developmental trajectory of audiovisual speech integration in early infancy. A review of studies using the McGurk paradigm. *Psychology of Language and Communication*, 19, 77–100. doi:10.1515/plc-2015-0006
Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, 73, 89–134. doi:10.1016/S0010-0277(99)00032-3
Trueswell, J. C., & Tanenhaus, M. K. (2005). *Approaches to studying world-situated language use*. Cambridge, MA: The MIT Press.
Turnbull, R. (2016). The role of predictability in intonational variability. *Language and Speech*, 60, 123–153.
Turnbull, R., Royer, A., Ito, K., & Speer, S. R. (2017). Prominence perception is dependent on phonology, semantics, and awareness of discourse. *Language, Cognition and Neuroscience*. doi:10.1080/23273798.2017.1279341
Vaillant-Molina, M., Bahrick, L. E., & Flom, R. (2013). Young infants match facial and vocal emotional expressions of other infants. *Infancy*, 18(suppl. 1), E97–E111. doi:10.1111/infa.12017

Voyer, D., Thibodeau, S-H., & Delong, B. J. (2014). Context, contrast, and tone of voice in auditory sarcasm perception. *Journal of Psycholinguistic Research*, 45, 29–53.

Vulchanova, M., Foyn, C. H., Nilsen, R. A., & Sigmundsson, H. (2014). Links between phonological memory, first language competence and second language competence in 10-year-old children. *Learning and Individual Differences*, 35, 87–95. doi:10.1016/j.lindif.2014.07.016

Walker A. S. (1982). Intermodal perception of expressive behaviors by human infants. *Journal of Experimental Child Psychology*, 33, 514–535. doi:10.1016/0022-0965(82)90063-7

Walker-Andrews, A. S. (1997). Infants' perception of expressive behaviors: Differentiation of multimodal information. *Psychological Bulletin*, 121, 437–456. doi:10.1037/0033-2909.121.3.437

Walker-Andrews, A. S. (1998). Emotions and social development: infants' recognition of emotions in others. *Pediatrics*, 102(5 Suppl E), 1268–1271.

Walker-Andrews, A. S., & Grolnick W. (1983). Discrimination of vocal expression by young infants. *Infant Behavior & Development*, 6, 491–498. doi:10.1016/S0163-6383(83)90331-4

Walker-Andrews A. S., & Lennon, E. (1991). Infants' discrimination of vocal expressions: Contributions of auditory and visual information. *Infant Behavior & Development*, 14, 131–142. doi:10.1016/0163-6383(91)90001-9

Wells, B., Peppé, S., & Goulandris, N. (2004). Intonation development from five to thirteen. *Journal of Child Language*, 31, 749–778. doi:10.1017/S030500090400652X

Werker, J. F., & Tees, R. (2002). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 25, 121–133. doi:10.1016/S0163-6383(02)00093-0

Wildgruber, D., Ackermann, H., Kreifelts, B., & Ethofer, T. (2006). Cerebral processing of linguistic and emotional prosody: fMRI studies. *Progress in Brain Research*, 156, 249–268. doi:10.1016/S0079-6123(06)56013-3

Zelazo, P. D., & Müller, U. (2002). Executive function in typical and atypical development. In U. Goswami (Ed.), *Handbook of childhood cognitive development* (pp. 445–469). Oxford: Blackwell. doi:10.1002/9780470996652.ch20

CHAPTER 14

Children's development of internal state prosody

Meghan E. Armstrong & Iris Hübscher
University of Massachusetts Amherst / Universitat Pompeu Fabra

Infants have access to the prosodic aspects of their ambient language even prior to birth, but many aspects of prosody are produced and comprehended well after infancy. One of these aspects includes prosody related to *internal states* such as beliefs, desires, feelings and emotions. In this chapter, we review the literature on prosody related to internal states, paying special attention to prosodic meanings associated with *emotions* and *belief states* and drawing from production and comprehension studies of preschool- and school-aged children. We show that there are many parallels in the development of these two aspects of prosody, suggesting the usefulness of studying them in tandem. Implications for these findings are discussed as well as fruitful directions for future work.

Introduction

Speakers express their beliefs and desires using a variety of linguistic resources, and prosody is no exception. Pierrehumbert and Hirschberg's (1990) seminal work on intonational meaning brought to the forefront the idea that intonation can convey relationships between the propositional content of utterances and the mutual belief space between speaker and hearer. For instance, the L*+HL H% (rise-fall-rise) contour in American English has been shown to convey both uncertainty and incredulity, meanings that are disambiguated through pitch range (Ward & Hirschberg, 1988). Recently there has been a resurgence in the literature on the relationship between speaker belief states and prosody (Armstrong, 2015, for questions, inter alia; Armstrong & Prieto, 2015; Gravano, Benus, Hirschberg, German, & Ward, 2008, for declaratives; Gunlogson, 2003; Lai, 2010, for cue words; Vanrell, Mascaró, Torres-Tamarit, & Prieto, 2013). As we deepen our understanding of how adult speakers negotiate the mutual belief space through prosody, we are beginning to understand how children learn to do

DOI 10.1075/tilar.23.14arm

so as well. The literature on children's development of belief state or mental state language is enormously skewed such that the bulk of it focuses on lexical items that convey this type of information, rather than prosody (Barak, Fazly, & Stevenson, 2014; Booth, Hall, Robison, & Kim, 1997; de Villiers, 2005; Matsui, 2014; Shatz, Wellman, & Silber, 1983, among many others). Recent work has begun to pay more attention to the role prosody plays in children's development of belief state language.

It is clear, however, that prosody plays an important role in a child's development, even prior to birth. It has been claimed that newborns' cries reflect the prosodic characteristics of their mother's language (Mampe, Friederici, Christophe, & Wermke, 2009), and newborns have been shown to be able to discriminate language classes based on prosodic cues (Nazzi, Bertoncini, & Mehler, 1998). Infants make use of prosodic cues to intentionality; by 5 months of age infants are sensitive to their mothers' intentions as conveyed by prosody, responding differently to approving versus prohibiting maternal utterances (Fernald, 1993). Pitch discrimination abilities have also been shown for 5-month-olds – Frota, Butler, and Vigário (2014) showed that at this age there is evidence for the discrimination of pitch category use for questions versus statements in European Portuguese (see Frota and Butler, this volume, for a discussion of language-specific effects that emerge very early on in prosodic development). By the end of their first year of life, young children demonstrate an ability to accurately infer the intentions of their caregivers by making use of information conveyed through prosody and hand gesture in novel situations (Esteve-Gibert, Prieto, & Liszkowski, 2017). In production, infants between the ages of 7 and 11 months make prosodic distinctions in their communicative versus investigative vocalizations (Papaeliou & Trevarthen, 2006), and also produce prosodic distinctions based on intentionality – Esteve-Gibert and Prieto (2013) found that for vocalizations that required something of their caretaker (requests and expressions of discontent), infants produced vocalizations with wider pitch range and longer duration.

Sakkalou and Gattis (2012) have also shown evidence that infants can infer intentions through prosodic cues (see Esteve-Gibert & Prieto, this volume, for an overview of further studies). But some work suggests that things start to change during the second year. Friend (2001) showed evidence for a transition from affective to linguistic meaning for prosody; while preverbal infants showed behavior that was guided by prosody, their behavior became better regulated by lexical content as their receptive vocabulary grew, resulting in a lexical bias around the age of 15 months. Friend suggests that this begins a new stage for the "fine-tuning" of communication between the caretaker and the infant, in which a child must learn to pay attention to both *what* is said and *how* it is said, thus marking a change from a point where 'the melody is the message' (Fernald, 1989), when children are very young, to a state where prosody and lexical items work together, i.e., the message goes beyond the melody.

In this chapter, we offer a summary of existing research on the development of prosody that is related to *internal state* such as beliefs, feelings and emotions, focusing specifically on preschool- and school-aged children. We assume that these are largely related both to an individual's Theory of Mind – the ability to understand mental states and capacities to see people as psychological beings with different thoughts, desires and beliefs – as well as their emotion understanding – understanding of emotional expressions and internal feelings in self and others (LaBounty, Wellman, Olson, Lagattuta, & Liu, 2008). Importantly, both emotional prosody and belief state prosody reference the internal states of the self and others, as well as perspective-taking. For this reason, we use the term *internal state prosody* (LaBounty et al., 2008; Lemche, Kreppner, Joraschky, & Klann-Delius, 2007) in this chapter. In the sections below, we first provide a brief discussion of children's development of emotional prosody, and next focus on the work that has been done on belief state prosody to date. Finally, we highlight further avenues of investigation in this area.

Children's use of prosody for conveying and perceiving emotions

Sauter, Panattoni, and Happé (2013) assessed British English-speaking children between the ages of 5 and 10 in terms of their ability to match picture stimuli with auditorily communicated emotions. While many 5-year-olds succeeded at the task, they found that children continued to improve with age. In Friend (2000), 4-year-olds were able to identify happy versus angry reiterant speech (e.g. mamama), but when speech was low pass filtered, they were not successful. Berman, Chambers, and Graham (2010) sought to discover whether preschoolers would show greater sensitivity to emotional prosody in less cognitively-demanding tasks, compared to conflict tasks that pit prosody and situational context against each other. Three- and 4-year-olds were presented with referential descriptions while they saw arrays of photographs of three objects: two objects of the same category that differed in physical state (e.g. a deflated ball and an intact ball) as well as an object that was unrelated (e.g. a star). Children heard a phrase such as *Look at the ball* spoken with happy-, sad- or neutral-sounding prosody. While 4-year-olds' pointing responses (explicit measure) did not show sensitivity to emotional prosody, their gaze patterns did. For example, the deflated ball was fixated more often in the presence of sad emotional prosody, and this occurred as the ambiguous noun unfolded. On the other hand, no such sensitivity was found for the eye gaze patterns of 3-year-olds. In a follow-up study using the same paradigm (Berman, Graham and

Chambers 2013), 5-year-olds showed sensitivity through both pointing and eye gaze, i.e. both explicit and implicit measures. The authors concluded that 4-year-olds are at a transitional point in their ability to integrate emotional prosody with other linguistic information (i.e. lexical information). Berman, Graham, Callaway, and Chambers (2013) found more evidence for this transitional period in a set of experiments which assessed preschoolers' ability to use emotional prosody to learn new words. Four- and 5-year-olds were presented with novel objects, first in their original state and later in a changed state (broken or enhanced). Instructions to find the referent of a novel word were heard with either sad-, neutral- or happy-sounding prosody. Similar to what was found in the studies mentioned above, while 5-year-olds were shown to link the novel referent with the expected corresponding object (e.g. broken object when sad-sounding prosody was heard and the enhanced object when happy-sounding prosody was heard) in both their eye gaze patterns and their explicit referential decisions, 4-year-olds only showed this behavior in their eye gaze patterns. Later, Berman, Chambers, and Graham (2016) had 3- and 5-year-olds match emotional affect communicated auditorily with facial expressions, using both implicit (eyetracking) and explicit (pointing) measures. While sad-sounding speech triggered looks to a sad face from the earliest moments of the utterance for the children, they took longer to look at happy faces based on auditory stimuli (about 800ms into the utterance). Three-year-olds showed only a latent ability to map vocal affect to a corresponding face in the eyetracking data, but could not match voices to pictures in the pointing task, while 5-year-olds were able to do so. The authors took 3-year-olds' success on the implicit task but not the explicit task as evidence that their perception of emotional states through vocal cues were at an emergent state. This parallels what has been found for Theory of Mind reasoning tasks as well (Wimmer & Perner, 1983). Valence also matters –when preschoolers had only audio information, they were better at choosing the target label for sound clips that were sad or angry when compared to happy or fearful (Nelson & Russell, 2011), and tended to label a happy voice with a negative valence. They also note that while preschoolers are rather good at reading body language through the face or the body, the ability to read emotion through auditory cues alone lags behind. A facilitating effect for sad-sounding prosody over happy-sounding prosody was also found for 3- and 5-year-olds, with sad-sounding prosody facilitating looks to sad facial expressions more quickly than happy-sounding prosody facilitating looks to happy facial expressions (Berman et al., 2016). Quam and Swingley (2012) found that 4- and 5-year-olds were consistent in using happy or sad prosody to decide whether a puppet had succeeded or failed at a task, while 2- and 3-year-olds depended more on facial and body language cues. In their production, children in this study were able to produce the prosodic contrasts in question, and unlike in comprehension there were no age differences, echoing the 'production/perception paradox' (Cutler & Swinney, 1987). This paradox makes reference to various researchers' observations that children are quite adept at producing prosodic meaning but have more difficulties comprehending it, especially at the sentence level. Quam and Swingley (2012) call attention to the perhaps puzzling fact that infants respond to emotional cues from maternal speech, but preschoolers have trouble detecting emotion through prosody. They suggest that while happy and sad contours may be accessible to babies in infancy, they might lose their iconicity through reinterpretation during the language acquisition process, and that the late learning of connections between pitch and emotion could be due to the 'complexity of pitch-contour patterning in the language as a whole'. This supports Fernald's (1992) idea that children gain access to the different functions of pitch in language at different points depending on factors such as developmental relevance and cue validity.

Researchers have also been interested in understanding children's behavior when presented with multiple cues to emotion. When lexical cues and prosodic cues compete, younger children tend to give more weight to the lexical cue (Friend, 2000; Friend, 2003; Friend & Bryant, 2000; Morton & Trehub, 2001), and also to situational context (Aguert, Laval, Le Bigot, & Bernicot, 2010; Gil, Aguert, Bigot, Lacroix, & Laval, 2014). Going back to the idea proposed by Fernald (1989), it seems that children tend to focus on *what* is said rather than *how* it is said. This is likely to be related to children's pragmatic development, since intonation is known to generate implicature (Armstrong & Prieto, 2015), but early on children may show difficulty in calculating implicatures, and give great weight to the semantic interpretations of utterances (Huang & Snedeker, 2009; Papafragou & Musolino, 2003). Morton and Trehub (2001) found that while adults rely exclusively on prosody when prosodic and lexical cues are conflicting, 4-year-olds rely on lexical information only. In this study, there was a gradual increase in children's reliance on prosody. Even so, only half of the oldest children in the study (age 10) attended to prosody. However, Morton, Trehub, and Zelazo (2003) found that when 6-year-olds were primed to attend to prosody more than lexical content, they were able to do so. In this experiment, 6-year-olds had to decide whether a friend of the experimenter sounded happy or sad. Children heard utterances with neutral content accompanied by happy or sad prosody (called "paralanguage" in the study), as well as sentences with happy (e.g. *My mommy gave me a treat*) or sad (e.g. *My dog ran away from home*) emotional content, as well as happy or sad prosody. The emotional content utterance included both prosody that matched the content of the utterance (happy content with happy prosody; sad content with sad prosody) as well as prosody that was a mismatch given the content of the utterance (happy content with sad prosody; sad content with happy prosody). The two different content types were presented in blocks, with half the 6-year-olds receiving the neutral content condition first, and the other half receiving the emotional content first. The authors found an effect of condition such that there was a lower proportion of responses based on prosody for the emotional content condition and a higher proportion for the neutral content condition. The authors also showed that children were primed by having to rely on prosody – when the neutral content condition was given first, responses to prosody for the emotional content condition increased. When the emotional content condition was presented first, responses to prosody for the neutral content condition dropped.

Aguert, Laval, Lacroix, Gil, and Bigot (2013) suggest that prosody plays a 'subordinate' role when it is in competition with situational context, a phenomenon that persists even into children's early teenage years. Aguert et al. (2010) describe this slow process as a 'developmental shift in the processing of expressive speech acts' that starts off driven by situational context (at ages 5 and 7). According to these authors, expressive interpretation starts to evolve at age 9, when children are able to give prosody as much weight as situational context. Thus, it seems that the ability to consistently relate emotional prosody with emotional categories is beginning to emerge around age 3 (implicitly) and more consistently around age 4 or 5. However, the ability to use prosodic cues when in competition with other information does not seem to occur until very late.

Khu, Chambers, & Graham (2017) pointed out a gap in the literature on children's acquisition of emotional prosody, since these studies do not make clear that children are using emotional prosody to reason about the emotional perspective of another speaker. In order to test this, they carried out a task in which 4-year-olds played a competitive game with another individual. If the child won the game, the individual would lose, and vice versa. The children had to use the other player's emotional reactions to infer the outcome of the game – thus if the other player sounded sad, this meant a win for the child, if they sounded happy it meant a loss for the child. Children heard ambiguous statements with happy- or sad-sounding emotional prosody. Their responses were recorded and their eye gazes were tracked. While the effect of emotional prosody was quite weak for children's explicit responses, eye gaze patterns showed that the children were able to anticipate that they would lose when the speaker sounded happy, and that they would win when the speaker sounded sad. However, this effect was not found until after the utterance had ended, showing that children needed some time to process emotional prosody. In any case, 4-year-olds were shown to implicitly use emotional prosody to make inferences about the perspective of another individual.

Further evidence of children's ability to use prosody and also facial cues to access the affective perspective of another person has been demonstrated by Hübscher, Wagner, and Prieto (2016). They tested 3-year-old children's sensitivity to a speaker's polite affective stance encoded through audiovisual prosody in requests. The materials were presented in a Powerpoint presentation in three different conditions: (a) with prosody and facial cues available (audio-visual); (b) with just facial cues available (visual-only); (c) with only prosody available (audio-only). The children always saw/heard twins requesting an object. While the lexicon was kept constant in both polite and non-polite requests, pitch contours and facial cues were varied, ranging from falling L* L% (non-polite) to rising L+H* H% (polite) and frown (non-polite) to smile (polite) respectively. After they heard both requests, children were asked to place the requested object in the basket in front of the twin that asked more nicely. Results from 216 responses showed that 3-year-olds performed significantly above chance overall, and demonstrated clear sensitivity to a speaker's polite affective stance encoded through audio-visual prosody. Additionally, there was no significant difference between the performance for the individual conditions, showing that both prosody and facial cues are used as powerful indicators of a speaker's affective stance and are able to override the lexical meaning of the word *please*, which was present and thus controlled for in both polite and non-polite requests.

Taking these finding for preschool-aged children together, it is perhaps very surprising that 5-month-olds were able to match other infants' expressions of happiness/joy and frustration/anger with corresponding video recordings (Vaillant-Molina, Bahrick, & Flom, 2013). That is, when presented with side-by-side video recordings of a joyful and an angry baby, infants looked longer to the baby whose expressions better matched the audio stimulus (see further discussion in Esteve-Gibert & Prieto, this volume). Reconciling this finding with the rest of the literature, however, is perhaps not so difficult if we consider Friend's (2001) proposal that there is a transition from affective to linguistic prosody, and a new stage of "fine-tuning" of communication between caretaker and infant. Thus we could hypothesize that even though infants may be sensitive to the relationship between emotional speech and corresponding gestures at 5 months, there would be a stage where this ability is lost or attenuated. The age ranges discussed above, therefore, could be the points where these abilities are recuperated. On the other hand, we must consider that the lack of description of prosodic cues for many of these studies make it difficult to make direct comparisons. For instance, if infants heard happy cooing noises produced by another infant, this is quite different from hearing an adult produce an utterance with lexical items, higher overall pitch range and expanded excursions. We now turn to another aspect of reasoning about internal states through prosody, belief states.

Children's acquisition of belief state prosody

Production

Krahmer and Swerts (2005) investigated how Dutch-speaking children (ages 7–8) and adults produce and perceive audiovisual cues to uncertainty. They used the

Feeling of Knowing (FOK) paradigm (Hart, 1965) which is a method for accessing participants' metacognitive evaluation of to what degree they know the answer to factual questions when answering them. In other words, speakers are able to convey information about their varying degrees of certainty through the production of their answers. Adults were found to use mostly fillers, delays and high intonation to convey uncertainty, while children relied more on delay and high intonation. Thus children relied more on prosodic strategies than lexical fillers. Later in a second task, adults and children watched the previously recorded responses and were asked to judge the speaker's level of uncertainty. The results showed that adults' judgments about the recordings were more reliable than children's. Furthermore, adults also judged the older children's level of certainty better compared to the younger children's. Interestingly, children were better at inferring adults' degree of certainty when compared to children's, suggesting that children's cues to uncertainty are less reliable. Visser, Krahmer, and Swerts (2014) later assess the production of children (8–11) in their visual and auditory expressions of uncertainty using the FOK paradigm in the form of a quiz game in either a collaborative setting (two children answering questions as a team) or a competitive setting (two children playing against each other). Again children produced longer delay and higher intonation for lower FOK (in addition to eyebrow movements). However, the study showed an effect of age and context such that older children and children in the competitive setting were more expressive in their signalling of uncertainty when compared to younger children or collaborating children. These studies show that children aged 7–11 make use of pitch and speech rate when conveying degree of certainty in answers to questions, though both age and context are surely at play.

More recently, Armstrong (2016) was the first longitudinal study to investigate the production of mental state intonation. In this corpus study, she analyzed the speech of two toddlers (1;7–3;6) and their caretakers in Puerto Rican Spanish. Puerto Rican Spanish polar questions can be produced with a general belief-marking contour, a contour that marks that the speaker has a belief about propositional content (unspecified for the direction of the belief) or a contour that marks a speaker's disbelief. Caretakers used the general question-marking contour with the greatest frequency (93% of all polar questions), followed by the disbelief contour (5% of all polar questions) and the belief presence contour (2%). The two children only produced two instances each of the belief-presence contours, and never produced the disbelief contour felicitously. Infelicitous uses of disbelief intonation were identified in the production of one child, leading to a conversational impasse with the child's mother – she had a difficult time comprehending what the child meant in this situation. Such situations likely provide the child with feedback about the pragmatic restrictions on intonation contours. Additionally, felicitous productions of the belief presence contour appeared later in the corpus;

one child produced her first belief contour at 2;8, and the other at 3;0. These ages are in line with the ages at which lexical mental state language has been shown to emerge (Shatz et al., 1983). Armstrong attributes the absence of felicitous productions of the disbelief contour to the mental reasoning processes that are necessary for using the belief presence contour vs. the disbelief contour, which may be more complex for the case of disbelief. For the belief presence contour, the child must only be aware that she has a belief about a proposition, while for the disbelief contour she must identify a contrast between a belief she has held and information that has just become available in the discourse.

In discourse narratives, children show that they are sensitive to concepts such as finality and non-finality through their use of intonation, but parental input may also be a predictor of the types of contours children choose. American English displays a declarative rise in narratives that allows the speaker to make frequent checks about whether the hearer is following what s/he is saying (Warren, 2016). In this way, the speaker monitors the information that is part of the mutual belief space shared with the hearer. Armstrong, Piccinini, and Ritchart (2016) examined this use in the speech of American English-speaking mothers and their daughters (ages 6–7). Overall, daughters used more non-question rises in narratives than mothers did. When comparing production of rising vs. level tunes, however, mothers that produced more rises had daughters that produced more rises. Similarly, mothers who produced more level contours had daughters that produced more level contours. This tendency was not found when rises were compared to falls. Daughters always produced more rises than falls. Mothers used more mid-narrative falls than daughters, while daughters seemed to reserve falls for very final intonational phrases, such as the very last one in a narrative. It is unclear whether the children in this study were aware of the 'checking-in' function of rises, or whether they use them more generally to convey non-finality, since rises in American English can carry both meanings. Perception work will be key in better understanding how children understand rises in narratives to be related to the belief states of others.

Comprehension

In some of the earliest work on children's intonational comprehension, Cruttenden (1985) tested 10-year-olds' ability to assess speaker certainty based on falling versus rising intonation. He employed a picture-matching task where children heard a sentence, for instance *It's a very nice garden* produced with what would correspond to a high rise or a rise-fall. Participants could associate the test items with three possible pictures: (1) a nice garden, but the house is falling down (2) both the house and the garden are very nice and (3) an overgrown garden with a house that is not very nice. Participants had to match the statement with one of the pictures. While both the adults and children were able to match falling intonation with scenario (1), children had a harder time than adults associating sentences produced with rising intonation with scenario (2), where we would assume that the rising intonation indicated reservation or lack of commitment, suggesting that older children have difficulties using prosody as a cue to a speaker's degree of commitment. However, not all adults responded the same way, and the author admits possible issues with the task. Cruttenden also assessed speaker neutrality versus surprise about propositional content, hypothesizing that participants should relate the rise to surprise and the fall to neutrality with phrases like *She's gone away.* Participants were expected to relate the rising tune with a surprised face, and the falling tune with a neutral face. Ten-year-olds performed significantly worse than adults, though many adults associated both tunes with the surprised face, again indicating some task issues. Wells, Peppé, and Goulandris (2004) carried out a battery of perception and production tasks (the PEPS-C, (Peppé & McCann, 2003)) with children aged 5–13. The authors investigated various meanings using a perception task that tapped into children's comprehension of uncertainty. Children participated in a picture-naming task, where the child would state what they saw in each picture, and the experimenter would repeat that word with either rising or falling intonation. The authors hypothesized that falling intonation would indicate that the child was correct, and as such the child would continue on to the next picture. It was assumed that when the experimenter produced a rise it would indicate uncertainty about what the child said, and therefore the child would repeat the word. Few errors were identified for the affirming responses, but the authors reported 41.5% errors for 5-year-olds for the rises, where children interpreted questioning as sounding affirming, showing more difficulties when compared to 8-, 10-, or 13-year-olds. However, we point out that these meanings were quite specific to the particular context of the task, meaning that children would have had to calculate pragmatic meaning in the specific discourse context of the task, thus making the task more difficult for younger children. In this way, if a child produced the name *house* and the experimenter produced this word with rising intonation, the implicature might be +> *Did you say house?* When the experimenter repeated the word with falling intonation the implicature might be something like +> *I get it, house.* Even if these uses of prosody are highly conventionalized, we still take them to be quite context-specific in this case.

Moore, Harris, and Patriquin (1993) was the first and only study to compare children's comprehension of mental state lexicon to their comprehension of belief state prosody. They point out a problem with prior studies that focus on children's tendency to give more weight to lexical information versus prosodic information, arguing that 'The point of linguistic processing, however, especially for young children, is not to identify particular words, or even to comprehend sentence meaning, but to extract and respond to the speaker's intended meaning, and prosody plays an important part in this pragmatic function' (p. 55). They carried out a 'hiding game' where a candy was hidden in a red or a blue box over a series of trials. Children were told that two puppets wanted to tell them where the candy was. Children had to listen to the puppets to figure out where the candy was. Using a between subjects design, the authors tested children's (aged 3–6) ability to understand speaker certainty based on belief state verbs or prosody. Prosodically, they tested the use of falling (certainty) versus rising (uncertainty) intonation. For the lexical condition, the verbs *know* vs. *think* were pitted against each other, as well as *think* vs. *guess*. Three-year-olds could not use either cue to decide where the object was. Four-year-olds were significantly above chance in using prosody to decide where an object was, while older children responded better based on lexically-encoded linguistic items when compared to prosody. The *think* vs. *guess* condition was much more difficult for older children than the *know* vs. *think* condition. A follow-up experiment with the same age groups included 'matched' and 'mismatched' conditions. In the matched condition lexical items expressing more certainty were matched with falling intonation, and less certain lexical items were matched with rising intonation. The opposite was done for the mismatched condition. Five-year-olds performed much better this time for the *think* vs. *guess* distinction, presumably because the prosodic information reinforced the strength difference between these midscalar terms. Five-year-olds also scored significantly worse on the mismatched condition, showing that inappropriate prosody caused them to perform more poorly. This points to some sort of awareness about the acceptability of prosody related to speaker certainty with lexical items at the age of five. The authors propose that in general, 5-year-olds search for explicit lexical information about speaker's mental states, and consider intonation to be secondary. However, they are affected when the secondary information is infelicitous, given lexical information. Yet the study also shows that children are sensitive to lexically-encoded mental state information around the same age that they are aware of prosodically-encoded mental state intonation, contra prior accounts. This finding, as suggested by the authors, points to a cognitive account for children's comprehension of speaker certainty; their representational Theory of Mind must be developed to a certain degree (Moore, Pure, & Furrow, 1990) in order to comprehend mental state language, whether it be lexical or prosodic. Nonetheless, lexical information has a dominant role around 4 or 5, when prosody has a secondary role. Moore et al.'s (1990) study provided all the impetus necessary to thrust the study of children's prosodic comprehension forward, but strangely few studies advanced this research agenda in the 1990s or early 2000s. Studies assessing children's comprehension of belief states conveyed prosodically have gained more traction in recent years, and these studies are described below.

Armstrong (2014) investigated children's comprehension of intonationally-encoded disbelief in polar questions in Puerto Rican Spanish. Children between the ages of 4 and 6 participated in a comprehension task and were presented with a pair of twins and their friend. They were told that the twins' friend was telling them about the animals she saw on vacation, and that there was always one twin that did not believe what the friend said. For each trial, the friend told the twins which animal she saw, and each twin subsequently reacted with a pre-recorded utterance. Children heard the response of each twin in the form of an echo question. One twin produced the echo question *¿un mono?* 'a monkey?' with neutral question intonation, and the other with disbelief question intonation, *¡¿un mono!?* 'a monkey!?'. The child then had to point to the twin that did not believe the friend. All age groups performed significantly above chance, and no significant differences were found for the performance of 4- vs. 5-year-olds, who provided correct answers 85% of the time. 6-year-olds significantly outperformed the younger children, however, providing the desired response 92% of the time. Unexpectedly, some six-year-olds produced facial gestures that are known to be associated with both questioning (brow-raising) and incredulity (movement of head backwards, brow furrowing) when they heard the experimental stimuli (N.B., they were not producing the stimuli). This indicates that by age six, children have formed strong associations not only between intonational forms and their respective meanings, but also the facial gestures that often accompany these forms. Even though the 6-year-olds in this study outperformed the 4- and 5-year-olds, all groups show the ability to perceive intonationally-encoded disbelief, suggesting that the window between ages 4 and 6 (and probably earlier if 4-year-olds are performing at above-chance levels) is a very important one for intonational development. It also shows, contra Moore et al. (1993), that 4- and 5-year-olds do not necessarily need lexical information present in order to perceive a belief state distinction through intonation – that is to say, school-aged children are indeed capable of using intonation as a primary cue, if it is the only cue. However, it could be that the tune used in Puerto Rican Spanish, falling-rising-falling contour, is a more predictable cue to doubt than the rise is for uncertainty in American English. As has been shown for emotional prosody, younger children often make use of the visual modality better than they do for the audio modality.

Armstrong, Esteve-Gibert, and Prieto (2014) carried out a similar study to Armstrong's (2014) study on Puerto Rican Spanish-acquiring children, this time assessing 3- to 5-year-old children's ability to comprehend disbelief in Central Catalan. Children were tested with three types of modalities: audio (intonation-only), visual (facial gesture-only) and audiovisual (combination of the two). Facial gestures included cues known to convey question-marking or disbelief cross-linguistically (Crespo Sendra, Kaland, Swerts, & Prieto, 2013): brow-raising for neutral questions and brow-furrowing and a backwards movement of the head for disbelief. Overall, children's success was predicted by age. Results showed that the youngest children performed very poorly when no visual information was present: 3-year-olds performed at near-chance levels for the audio-only task. Four-year-olds showed a great deal of variability for the audio-only condition, while this variability was not found for 5-year-olds, who performed close to ceiling, and with little variability. On the other hand, 4-year-olds were quite successful when they had access to visual information. Thus while children may depend on cues from facial gesture in order to perceive disbelief early on, by around 5 years of age they no longer need to rely on visual cues.

Armstrong et al. (2014) also assessed Theory of Mind reasoning, by adding a modified version of Wimmer and Perner's (1983) Sally Anne task, where participants saw a video with puppets. First, a princess puppet put her ball in one of two containers, covered it and left for school. While the princess was gone, a lion came and moved the ball from one container to the other. He covered it as well, and left. When the princess came back from school, the child was asked *On buscarà la pilota, la nena?* 'Where will the girl look for the ball?'. A child passed the task if they answered that she will look for the ball where she left it, and failed if they answered that she will look in the container where the ball had been moved to by the lion. Across age groups, the authors found that the children that passed the false belief task were the ones that were the most successful at the comprehension task, regardless of the condition. This suggests, perhaps unsurprisingly, that the children with more sophisticated belief reasoning were the ones that tended to be most successful on the comprehension task, which also involves some degree of ToM reasoning since a child must use either visual or audio cues (or the combination of the two) to reason about the mental states of others.

Most recently, Hübscher, Esteve-Gibert, Igualada, & Prieto (2017) tested 3- to 5-year-olds Central Catalan acquiring children's ability to detect speaker uncertainty through various cues (intonational, gestural and lexical) when a speaker answered a question (divided into two groups: younger children and older children). Using a Powerpoint presentation, participants were introduced to different sets of twins that were playing a guessing game with their friend Bàrbara. Bàrbara would ask the twins a question, for example *Quin és el meu vegetal preferit?* 'What's my favorite vegetable?' with the answer revealed in a thought bubble. For example, if Bàrbara's favorite vegetable was a tomato, the child would see a thought bubble near Bàrbara's head with a tomato in it. The child was told that for each question that Bàrbara asked, there would be one twin that was sure of the answer, and another that was not. The child had to point to the uncertain twin. In the lexical condition, some fragments were produced with *potser* 'maybe' (e.g. *Potser el tomàquet* 'Maybe the tomato') versus *segur que* 'certainly' (e.g. *Segur que el tomàquet* 'Surely it's the tomato). Crucially, both of these were produced with the same falling intonation contour, L* L% in the Cat_ToBI system. The stimuli for the specific epistemic condition (certain vs. uncertain) were produced with corresponding gestures (i.e. a head-nod suggesting certainty vs. squinted eyes/raised eyebrows/head tilt for uncertainty). Meanwhile, in the intonation condition, no epistemic information was given in the sentence fragment. Rather, epistemic information was conveyed through intonation. Participants heard a sentence fragment such as *el tomàquet* 'the tomato' as an answer to a question with either falling (L* L%) or rising (L* H%) intonation, so as to convey certainty or uncertainty. The intonational stimuli had the same gestural cues as the lexical stimuli. Children were given either the lexical or the intonation condition, with trials that were audio-only, video-only or audiovisual. Both younger and older children performed better when some sort of visual information was present (i.e. the visual only and the audiovisual conditions). The authors point out that visual cues may be used for bootstrapping meaning, as has been claimed for other aspects of language development (Butcher & Goldin-Meadow, 2000; Kelly, 2001; McNeill, Cassell, & McCullough, 1994). Different from other studies, however, the authors showed that the younger children were more sensitive to intonational cues to uncertainty than they were to lexical cues. The 3-year-olds that received the intonation condition performed significantly better than the ones that received the lexical condition. The authors see the results as an example of prosodic and gestural bootstrapping, such that both prosodic and gestural cues guide children's acquisition of pragmatic meanings, such as a speaker's degree of certainty about a proposition.

Few studies have used implicit measures to assess children's comprehension of belief state intonation. Armstrong, Andreu, Esteve-Gibert, and Prieto (2016) used implicit and explicit measures to evaluate Central Catalan-speaking children's ability to override contextual information through morphosyntactic versus prosodic cues. Children aged 4–6 watched a series of videos which began with two actors, each playing with a stuffed rabbit on a chair. They then stopped playing with their rabbits, each leaving their respective rabbit on their respective chair. The actors left the room, shutting the door behind them. At this point, a puppet arrives. The puppet switches only *one* of the stuffed rabbits for a different animal, and leaves the scene. The actors then return and approach the chairs, at which point the screen freezes. Participants then heard a speech stimulus with one of two linguistic conditions. For the morphosyntactic condition, children heard either *Aquest **no** és el meu peluix* 'This is not my doll' or ***Aquest sí que** és el meu peluix* 'This sure is my doll'. For the prosodic condition they heard *Aquest és el meu peluix* 'This is my doll' produced with one of two intonation contours – a falling declarative (H+L* L% in Cat_TOBI) or a rising contour conveying disbelief (L* LH%). For each trial, the children heard one of the two epistemic possibilities (confirming or disbelieving) depending on the linguistic condition they were given in the between-subjects design. After hearing the spoken stimulus, children were asked 'Who said that? Point to them for me'. For a control group of adults, the presence of sentential negation and disbelief intonation guided them to choose the actor whose animal had been switched, while *sí que* and matter-of-fact intonation guided them to choose the actor whose animal was not switched. Children tended to rely on situational context and were biased towards choosing the actor whose animal was switched, since this was perhaps the most likely person to speak. Children often failed to override their context-based assumption, but got better with age. Eyetracking showed that 4-year-olds barely used prosodic information – looks to target actor were largely unaffected by prosodic cues though they did seem to make use of sentential negation. Five-year-olds were able to use both morphosyntax and intonation for confirming a hypothesis, but used only morphosyntax for overriding a hypothesis. There was evidence that the ability to use prosody for overriding a hypothesis was just emerging for 6-year-olds. The study demonstrated that, as has been claimed in the emotional prosody literature, children depend a great deal on situational content in interpreting meaning. Whether children's responses would change if they were trained to attend to prosodic cues, as was done in Morton et al. (2003), remains to be seen. One limitation mentioned in the study is that disbelief intonation is perhaps more grammaticalized in Central Catalan compared to matter-of-fact intonation, and therefore may have further complicated the task. The reliability of form-meaning mappings is an area that could shed light on the comprehension of belief state prosody.

Discussion and conclusions

This review reveals interesting parallels between children's development of two types of internal state prosody: emotional prosody and belief state prosody. The ability to explicitly match emotional prosody with a visually-manifested emotion is present at around 4 or 5 years, and beginning to emerge at 3 years. On the belief state side, the ability to detect meanings such as disbelief and uncertainty are also emerging at this time. The idea that 4- and 5-year-olds are more consistent at inferring internal states through prosody compared to 3-year-olds is perhaps not surprising, since this is also the age when children are known to predictably pass false belief tasks, with 3-year-olds only showing an implicit ability on such tasks (see Rubio-Fernández & Geurts, 2013, for a review; Wimmer & Perner, 1983). The research also shows that younger children (e.g. 3-year-olds) are more adept at using gestures to infer information about both emotions and beliefs, rather than lexical information. Valence effects have also been shown, with children identifying emotions with negative valence more readily than positive valence. The existing comprehension studies look at *dis*belief and *un*certainty, rather than the expression of positive belief. However, it is interesting to note that in production, preschool-aged children have been found to produce prosodic cues related to happiness (Quam & Swingley, 2012) and also positive belief (Armstrong, 2016).

The ability to weight a prosodic cue over lexical or situational information seems to take longer to develop. However, we should take into account that children likely do not encounter many situations where they are faced with discrepant cues. Rather, they are likely to encounter situations in which prosody reinforces lexical or situational information. To this end, it is perhaps unsurprising that younger children show difficulties when prosody is 'competing' with some other source of information. Their lexical biases (emerging once children learn that word meaning is important) lead children to focus more on *what* is said rather than *how* it is said. However, this is tightly linked to a child's pragmatic development, which is important to consider since children's reasoning about pragmatic implicature is also known to emerge later (Matthews, 2014). The idea that children rely on one cue (lexical) more than the other (prosodic) might also be explained by Bates and MacWhinney's (Bates & MacWhinney, 1982; MacWhinney, 1987) Competition Model. According to this model, when two reliable cues conflict, the one that wins out is higher in *conflict reliability*. For instance, MacWhinney (2013) points out that in L1 Dutch it is not until after age 8 that pronoun case takes precedence over the more frequent but slightly less reliable cue of word order. In addition to the fact that children's pragmatic competence is slow to emerge, we propose that lexical cues could have higher conflict reliability for children when compared to prosody. The emotional prosody research has shown that children as old as 10 may still struggle to assign appropriate weight to prosody, but we also find evidence that children's ability to use prosody to override other information (situational) is starting to emerge by age six. Hübscher et al.'s (2017) work showing 3-year-old's very early sensitivity to prosodic cues, even before epistemic adverbs, however, illustrates the importance of taking into consideration specific prosodic meanings and cue reliability in a given language. To this end, it could also be interesting to consider how language-specific versus cross-linguistically accessible certain cues are. For instance, rising intonation is extremely common crosslinguistically to express interrogativity or uncertainty (Gussenhoven, 2004), and so perhaps such meanings are available to children earlier on in the acquisition process, and by nature of this could be more predictable than some lexical items.

One issue in making sense of past studies is the ecological validity of the methods used. For instance, we know that prosodic meaning is often quite sensitive to context. Thus providing children with rich and realistic contexts is important for testing their knowledge of prosody meaning, which has been pointed out by Ito, Jincho, Minai, Yamane, and Mazuka (2012), who argue that early studies of prosodic comprehension have presented stimuli to children in ways that might seem too 'out-of-the-blue', and lacked adequate situational context. As we know, in real life, language is always very contextualized, and thus utterances lacking realistic contextualization that is accessible for a child (i.e. situations in which a child might find herself/himself) could prove confusing. In addition, as we have mentioned above, pitting prosody against other cues such as situational context may also not be the best measure of children's developmental patterns, since it is probably not the most frequent scenario a child will encounter. The low-pass filtering of speech is also helpful to isolate prosodic cues, but again, children likely do not encounter scenarios in which they have to interpret prosodic meaning in the absence of lexical cues, and thus we must question to what extent their success on such experiments are good measures of how children might produce and comprehend prosody in their daily lives. As we have mentioned briefly above, another issue with a good deal of the studies on emotional prosody is that the type of prosody is ill-defined. In fact, the prosodic characteristics being tested are often referred to as 'paralinguistic cues' or similar. This makes it difficult to understand which acoustic cues might be important for children to attend to. Thus studies with more accurate descriptions of prosodic patterns are needed.

The developmental literature on belief state prosody has benefited from the use of the Autosegmental Model (Jun, 2010; Pierrehumbert, 1980), in which phonological labels within different Tones and Breaks Indices systems can be used to refer to specific contours. This affords researchers the ability to have a clearer idea of which form-meaning functions are at play. However, it is likely the case that emotional meanings such as 'happy' or 'sad' are not typically grammaticalized intonationally in languages, while belief state meanings like 'disbelief' or 'uncertainty' are indeed grammaticalized in many languages. Further, belief state meanings may also be affected by discourse context (Armstrong & Prieto, 2015), and therefore it is important to control for the types of contexts that are used when testing children on contour meaning. For example, the meaning of falling-rising intonation in American English can vary widely based on context: does this present a problem for children?

While the bulk of the studies in our overview include comprehension data, there is a lack of research focusing on children's production of prosody that refers to internal states. Production studies, as well as studies combining both children's production and comprehension, would be particularly useful. One challenge is that testing production typically assesses children's prosodic expression of her/his *own* belief state, while testing comprehension assesses a child's comprehension of what some other individual believes. This has been pointed out by Ünal and Papafragou (2016), with respect to the study of children's acquisition of evidentiality/knowledge sources. Thus combining production and perception studies will be useful in assessing how intonation is used for conveying a child's own beliefs as well as how s/he is sensitive to the belief states of others. These differences in perspective taking could possibly explain why some studies have shown that comprehension abilities lag behind production abilities (Cutler & Swinney, 1987 amongst others; Quam & Swingley, 2012), that is to say, children could be more adept at representing their own mental states through prosody than they are at using it to comprehend the mental states of others. There is also a need for production studies that concurrently observe the production of mental state language such as verbs of cognition and modal verbs alongside the emergence of belief state prosody. Furthermore, a more holistic view of prosodic development could also prove useful for the study of internal state prosody. For instance, Ito (this volume) points out advantages of studying affective prosody along with focus prosody. Thus testing children's performance on other functions of prosody in tandem with internal state prosody, or even different aspects of internal state prosody (e.g. emotional along with belief state prosody) could prove quite useful. Understanding individual differences in children such as their general emotional development, their cognitive development (perspective-taking and belief-reasoning abilities) and individual traits (how empathetic they might be) should also be taken into account in future work.

In sum, the work discussed here points to the idea that children access different functions of prosody at different points on their developmental trajectories, based both on how relevant these functions are developmentally and how reliable specific cues are (Fernald, 1992). Work comparing more reliable versus less reliable prosodic cues would also be fruitful as we continue to understand how children develop the ability to use prosody, regardless of its specific meaning. For instance, the consistency of form-function mapping is an important component of Chen's (this volume) proposal for a typological view of the acquisition of prosodic focus across languages. It is important to remember that prosody, something that parents use for attention-maintaining, soothing or affect early in the child's life, begins to be used in more linguistic ways later on. We also know that prosodic bootstrapping occurs in infancy, so there are prosodic cues that we might say are more stable across the lifespan. But the fact that the meanings of prosody might fluctuate between infancy and late childhood reveals the very unique nature of prosodic acquisition, since no other aspects of the grammar seem to show this dynamicity. This, combined, with the sophisticated nature of understanding the internal states of others, makes the prosodic acquisition of emotional and belief states a very interesting problem indeed, and we hope this chapter can provide some impetus for further investigation into this exciting area of research.

Acknowledgements

We are thankful to the volume editors and the three reviewers for their very helpful comments, which have undoubtedly improved the quality of this manuscript.

The second author acknowledges funding by the Spanish Ministry of Science and Innovation (grant FFI2015- 66533-P "Intonational and gestural meaning in language") and a grant awarded by the Generalitat de Catalunya (2014 SGR 00925) to the Prosodic Studies Group.

References

Aguert, M., Laval, V., Lacroix, A., Gil, S., & Bigot, L. L. (2013). Inferring emotions from speech prosody: Not so easy at age five. *PLoS ONE*, 8(12). doi: 10.1371/journal.pone.0083657

Aguert, M., Laval, V., Le Bigot, L., & Bernicot, J. (2010). Understanding expressive speech acts: The role of prosody and situational context in french-speaking 5- to 9-year-olds. *Journal of Speech, Language, and Hearing Research*, 53(6), 1629–1641. doi: 10.1044/1092-4388(2010/08-0078)

Armstrong, M. E. (2014). Child comprehension of intonationally-encoded disbelief. In W. Orman & M. J. Valleau (Eds.), *Proceedings of the 38th Annual Boston University Conference on Language Development* (pp. 25–38). Somerville, MA: Cascadilla Press.

Armstrong, M. E. (2015). Accounting for intonational form and function in Puerto Rican Spanish polar questions. *Probus*. doi:10.1515/probus-2014-0016

Armstrong, M. E. (2016). Production of mental state intonation in the speech of toddlers and their caretakers. *Language Acquisition* . doi:10.1080/10489223.2016.1242588

Armstrong, M. E., Andreu, L., Esteve-Gibert, N., & Prieto, P. (2016). Children's online use of morphosyntax and prosody in overriding context-based hypotheses. *Probus*, 28(1), 57–90.

Armstrong, M. E., Esteve-Gibert, N., & Prieto, P. (2014). The acquisition of multimodal cues to disbelief. *Proceedings of Speech Prosody* – Dublin, 1134–1138.

Armstrong, M. E., Piccinini, P., & Ritchart, A. (2016b). Non-question rises in narratives produced by mothers and daughters. *Proceedings of Speech Prosody 8* – Boston, 123–127.

Armstrong, M. E., & Prieto, P. (2015). The contribution of context and contour to perceived belief in polar questions. *Journal of Pragmatics*, 81, 77–92. doi: 10.1016/j.pragma.2015.03.006

Barak, L., Fazly, A., & Stevenson, S. (2014). Gradual acquisition of mental state meaning: *A computational investigation*. *Proceedings of Cognitive Science Society* – Quebec City, 1886–1891.

Bates, E., & MacWhinney, B. (1982). Functionalist approaches to grammar. In E. Wanner & L. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 173–218). New York, NY: Cambridge University Press.

Berman, J. M. J., Chambers, C. G., & Graham, S. A. (2010). Preschoolers' appreciation of speaker vocal affect as a cue to referential intent. *Journal of Experimental Child Psychology*, 107(2), 87–99. doi: 10.1016/j.jecp.2010.04.012

Berman, J. M. J., Chambers, C. G., & Graham, S. A. (2016). Preschoolers' real-time coordination of vocal and facial emotional information. *Journal of Experimental Child Psychology*, 142, 391–399. doi: 10.1016/j.jecp.2015.09.014

Berman, J. M. J., Graham, S. A., Callaway, D., & Chambers, C. G. (2013). Preschoolers use emotion in speech to learn new words. *Child Development*, 84(5), 1791–1805. doi: 10.1111/cdev.12074

Berman, J. M. J., Graham, S. A., & Chambers, C. G. (2013). Contextual influences on children's use of vocal affect cues during referential interpretation. *Quarterly Journal of Experimental Psychology*, 66(4), 705–726. doi:10.1080/17470218.2012.713367

Booth, J. R., Hall, W. S., Robison, G. C., & Kim, S. Y. (1997). Acquisition of the mental state verb know by 2- to 5-year-old children. *Journal of Psycholinguistic Research*, 26(6), 581–603. doi:10.1023/A:1025093906884

Butcher, C., & Goldin-Meadow, S. (2000). Gesture and the transition from one- to two-word speech: When hand and mouth come together. In D. McNeill (Ed.), *Language and gesture* (pp. 235–257). New York, NY: Cambridge University Press. doi:10.1017/CBO9780511620850.015

Crespo Sendra, V., Kaland, C., Swerts, M., & Prieto, P. (2013). Perceiving incredulity: The role of intonation and facial gestures. *Journal of Pragmatics*, 47(1), 1–13. doi:10.1016/j.pragma.2012.08.008

Cruttenden, A. (1985). Intonation comprehension in ten-year-olds. *Journal of Child Language*, 12(3), 643–661. doi:10.1017/S030500090000670X

Cutler, A., & Swinney, D. A. (1987). Prosody and the development of comprehension. *Journal of Child Language*, 14, 145–167. doi:10.1017/S0305000900012782

de Villiers, J. (2005). Can language acquisition give children a point of view? In J. W. Astington, & J. A. Baird (Eds.), *Why language matters for Theory of Mind* (pp. 186–219). New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780195159912.003.0010

Esteve-Gibert, N., & Prieto, P. (2013). Prosody signals the emergence of intentional communication in the first year of life: Evidence from Catalan-babbling infants. *Journal of Child Language*, 40(5), 919–944. doi:10.1017/S0305000912000359

Esteve-Gibert, N., Prieto, P., & Liszkowski, U. (2017). Twelve-month-olds understand social intentions based on prosody and gesture shape. *Infancy*, 22(1), 108–129.

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child Development*, 60(6), 1497–1510. doi:10.2307/1130938

Fernald, A. (1992). Meaningful melodies in mothers' speech to infants. In H. Papousek, U. Jurgens, & M. Papousek (Eds.), *Nonverbal vocal communication: Comparative and developmental approaches* (pp. 262–282). Cambridge: Cambridge University Press.

Fernald, A. (1993). Approval and disapproval: Infant responsiveness to vocal affect in familiar and unfamiliar languages. *Child Development*, 64(3), 657–674. doi:10.2307/1131209

Friend, M. (2000). Developmental changes in sensitivity to vocal paralanguage. *Developmental Science*, 3(2), 148–162. doi:10.1111/1467-7687.00108

Friend, M. (2001). The transition from affective to linguistic meaning. *First Language*, 21(63), 219–243. doi:10.1177/014272370102106302

Friend, M. (2003). What should I do? Behavior regulation by language and paralanguage in early childhood. *Journal of Cognition and Development*, 1, 161–183. doi:10.1207/S15327647JCD0402_02

Friend, M., & Bryant, J. B. (2000). A developmental lexical bias in the interpretation of discrepant messages. *Merrill-Palmer Quarterly*, 46(2), 342–369.

Frota, S., Butler, J., & Vigário, M. (2014). Infants' perception of intonation: Is it a statement or a question? *Infancy*, 19(2), 194–213. doi:10.1111/infa.12037

Gil, S., Aguert, M., Bigot, L. L., Lacroix, A., & Laval, V. (2014). Children's understanding of others' emotional states: Inferences from extralinguistic or paralinguistic cues? *International Journal of Behavioral Development*, 38(6), 539–549. doi:10.1177/0165025414535123

Gravano, A., Benus, S., Hirschberg, J., German, E. S., & Ward, G. (2008). The effect of contour type and epistemic modality on the assessment of speaker certainty. *Proceedings of Speech Prosody* – Campinas, 401–404.

Gunlogson, C. (2003). *True to form: Rising and falling declaratives as questions in English*. New York, NY: Routledge.

Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511616983

Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology*, 56(4), 208–216. doi:10.1037/h0022263

Huang, Y. T., & Snedeker, J. (2009). Online interpretation of scalar quantifiers: Insight into the semantics-pragmatics interface. *Cognitive Psychology*, 58(3), 376–415. doi:10.1016/j.cogpsych.2008.09.001

Hübscher, I., Esteve-Gibert, N., Igualada, A., & Prieto, P. (2017). Intonation and gesture as bootstrapping devices in speaker uncertainty. *First Language*, 37(1), 24–41.

Hübscher, I., Wagner, L., & Prieto, P. (2016). Young children's sensitivity to polite stance expressed through audiovisual prosody in requests. *Proceedings of Speech Prosody* – Boston, 897–901.

Ito, K., Jincho, N., Minai, U., Yamane, N., & Mazuka, R. (2012). Intonation facilitates contrast resolution: Evidence from Japanese adults and 6-year olds. *Journal of Memory and Language*, 66(1), 265–284. doi:10.1016/j.jml.2011.09.002

Jun, S. A. (2010). *Prosodic typology: The phonology of intonation and phrasing*. Oxford: Oxford University Press.

Kelly, S. D. (2001). Broadening the units of analysis in communication: Speech and nonverbal behaviours in pragmatic comprehension. *Journal of Child Language*, 28(2), 325–349. doi:10.1017/S0305000901004664

Khu, M., Chambers, C., & Graham, S. A. (2017). When you're happy and I know it: Four-year-olds' emotional perspective taking during online language comprehension. *Child Development*. doi:10.1111/cdev.12855

Krahmer, E., & Swerts, M. (2005). How children and adults produce and perceive uncertainty in audiovisual speech. *Language and Speech*, 48(1), 29–53. doi:10.1177/00238309050480010201

LaBounty, J., Wellman, H. M., Olson, S., Lagattuta, K., & Liu, D. (2008). Mothers' and fathers' use of internal state talk with their young children. *Social Development*, 17(4), 757–775. doi:10.1111/j.1467-9507.2007.00450.x

Lai, C. (2010). What do you mean, you're uncertain? The interpretation of cue words and rising intonation in dialogue. *Proceedings of Interspeech* – Makuhari, 1413–1416.

Lemche, E., Kreppner, M. J., Joraschky, P., & Klann-Delius, G. (2007). Attachment organization and the early development of internal state language: A longitudinal perspective. *International Journal of Behavioral Development*, 31(3), 252–262. doi:10.1177/0165025407076438

MacWhinney, B. (1987). The competition model. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 249–308). Hillsdale, NJ: Lawrence Erlbaum Associates.

MacWhinney, B. (2013). The logic of the unified model. In S. Gass, & A. Mackey (Eds.), *Handbook of Second Language Acquisition* (pp. 211–227). New York, NY: Routledge.

Mampe, B., Friederici, A. D., Christophe, A., & Wermke, K. (2009). Newborns' cry melody is shaped by their native language. *Current Biology*, 19(23), 1994–1997. doi:10.1016/j.cub.2009.09.064

Matthews, D. (2014). *Pragmatic development in first language acquisition*. Amsterdam/Philadelphia John Benjamins.
Matsui, T. (2014). Children's understanding of linguistic expressions of certainty and evidentuality. In D. Matthews (Ed.), *Pragmatic development in first language acquisition* (pp. 295–316). Amsterdam: John Benjamins.
McNeill, D., Cassell, J., & McCullough, K. E. (1994). Communicative effects of speech-mismatched gestures. *Research on Language and Social Interaction*, 27(3), 223–237. doi:10.1207/s15327973rlsi2703_4
Moore, C., Harris, L., & Patriquin, M. (1993). Lexical and prosodic cues in the comprehension of relative certainty. *Journal of Child Language*, 20(1), 153–167. doi:10.1017/S030500090000917X
Moore, C., Pure, K., & Furrow, D. (1990). Children's understanding of the modal expression of speaker certainty and uncertainty and its relation to the development of a representational Theory of Mind. *Child Development*, 61(3), 722–730. doi:10.2307/1130957
Morton, J. B., & Trehub, S. E. (2001). Children's understanding of emotion in speech. *Child Development*, 72(3), 834–843. doi:10.1111/1467-8624.00318
Morton, J. B., Trehub, S. E., & Zelazo, P. D. (2003). Sources of inflexibility in 6-year-olds' understanding of emotion in speech. *Child Development*, 74(6), 1857–1868. doi:10.1046/j.1467-8624.2003.00642.x
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.
Nelson, N. L., & Russell, J. A. (2011). Preschoolers' use of dynamic facial, bodily, and vocal cues to emotion. *Journal of Experimental Child Psychology*, 110(1), 52–61. doi:10.1016/j.jecp.2011.03.014
Papaeliou, C. F., & Trevarthen, C. (2006). Prelinguistic pitch patterns expressing 'communication' and 'apprehension'. *Journal of Child Language*, 33(1), 163–178. doi:10.1017/S0305000905007300
Papafragou, A., & Musolino, J. (2003). Scalar implicatures: Experiments at the semantics-pragmatics interface. *Cognition*, 86(3), 253–282. doi:10.1016/S0010-0277(02)00179-8
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 and Phonetics*, 17(4–5), 345–354. doi:10.1080/0269920031000079994
Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation*. Indiana University Linguistics Club, MIT.
Pierrehumbert, J., & 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: The MIT Press.
Quam, C., & Swingley, D. (2012). Development in children's interpretation of pitch cues to emotions. *Child Development*, 83(1), 236–250. doi:10.1111/j.1467-8624.2011.01700.x
Rubio-Fernández, P., & Geurts, B. (2013). How to pass the false-belief task before your fourth birthday. *Psychological Science*, 24(1), 27–33. doi:10.1177/0956797612447819
Sakkalou, E., & Gattis, M. (2012). Infants infer intentions from prosody. *Cognitive Development*, 27(1), 1–16. doi:10.1016/j.cogdev.2011.08.003
Sauter, D. A., Panattoni, C., & Happé, F. (2013). Children's recognition of emotions from vocal cues. *British Journal of Developmental Psychology*, 31(1), 97–113. doi:10.1111/j.2044-835X.2012.02081.x

Shatz, M., Wellman, H. M., & Silber, S. (1983). The acquisition of mental verbs: A systematic investigation of the first reference to mental state. *Cognition*, 14(3), 301–321. doi:10.1016/0010-0277(83)90008-2
Ünal, E., & Papafragou, A. (2016). Production-comprehension asymmetries and the acquisition of evidential morphology. *Journal of Memory and Language*, 89, 179–199.
Vaillant-Molina, M., Bahrick, L. E., & Flom, R. (2013). Young infants match facial and vocal emotional expressions of other infants. *Infancy*, 18(SUPPL.1), E97–E111. doi:10.1111/infa.12017
Vanrell, M. M., Mascaró, I., Torres-Tamarit, F., & Prieto, P. (2013). Intonation as an encoder of speaker certainty: Information and confirmation yes-no questions in Catalan. *Language and Speech*, 56(2), 163–190. doi:10.1177/0023830912443942
Visser, M., Krahmer, E., & Swerts, M. (2014). Children's expression of uncertainty in collaborative and competitive contexts. *Language and Speech*, 57(1), 86–107. doi:10.1177/0023830913479117
Ward, G., & Hirschberg, J. (1988). Intonation and propositional attitude the pragmatics of L* + H LH%. *Proceedings of the Fifth Eastern States Conference on Linguistics* – Philadelphia, 512–522.
Warren, P. (2016). *Uptalk: The phenomenon of rising intonation*. Cambridge: Cambridge University Press. doi:10.1017/CBO9781316403570
Wells, B., Peppé, S., & Goulandris, N. (2004). Intonation development from five to thirteen. *Journal of Child Language*, 31(4), 749–778. doi:10.1017/S030500090400652X
Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13(1), 103–128. doi:10.1016/0010-0277(83)90004-5

CHAPTER 15

Get the focus right across languages

Acquisition of prosodic focus-marking in production

Aoju Chen
Utrecht University

Languages differ in both the use of prosody in focus marking and prosodic systems. In the light of recent studies of acquisition of prosodic focus marking in typologically different languages, we argue that typological differences have direct influences on acquisition of prosodic focus marking. More specifically, differences in reliance on phonetic uses of prosody and transparency of how prosody is used phonologically to encode focus can lead to cross-linguistic differences in rate of acquisition. Differences in whether the prosodic means involved are also used for lexical purposes and in interface between prosody and word order can trigger cross-linguistic differences in route of acquisition.

Introduction

Prosody has multiple functions in communication. For example, in many languages speakers vary prosody to structure the flow of information so that more important information is distinguished from less important information in a sentence and changes in the information status of referents (e.g. from 'new' to 'given') are signalled. Adult prosodic competence thus entails that children can produce and interpret prosody in a contextually appropriate way, and efficiently process prosodic information in comprehension. This chapter is concerned with how children develop their ability in using prosody to mark focus in speech production across languages.

Focus is an information structural category that refers to the predication on a topic in a sentence, and typically contains information that is new to the receiver (Lambrecht, 1994; Vallduví & Engdahl, 1996). The focal constituent can be either a syntactic constituent (e.g. a phrase, a clause, a sentence) or an element of a syntactic constituent (e.g. the noun of a noun phrase). This distinction is known as the difference between broad focus and narrow focus (Ladd, 1980). Focus can

DOI 10.1075/tilar.23.15che

also contain contrastive information, e.g. a correction or an alternative to what has been introduced previously. In both theoretical discussions and experimental studies of focus, question-answer pairs are often used to determine what is the focus and what is not the focus (or: what is the topic) in declarative sentences (Roberts, 2012). For example, who-questions and what-questions render either the subject or the object focal in SVO sentences (Examples (1) b–e); what-happens questions put the whole sentence in focus (Example (1a)). Contrastive focus can be determined by the presence of a limited number of alternatives in the discourse (Examples (1a) and (1f)–(g)). It can be either 'broad' or 'narrow' in terms of the size of the focal constituent.

(1) a. Speaker A: What happened?
Speaker B: The Johnsons bought the villa. (broad focus)
Speaker C: No, the Smiths leased the entire estate. (contrastive broad focus)

b. Speaker A: Who bought the villa?
Speaker B: The Johnsons bought the villa. (narrow focus on subject)

c. Speaker A: What distracted Mary?
Speaker B: Music distracted Mary. (narrow focus on subject)

d. Speaker A: Whom did Mary call?
Speaker B: Mary called Thomas. (narrow focus on object)

e. Speaker A: What did the Johnsons buy?
Speaker B: The Johnsons bought the villa. (narrow focus on object)

f. Speaker A: Did music distract Mary?
Speaker B: (No.) Laughter distracted Mary. (contrastive narrow focus on subject)

g. Speaker A: Did the Johnsons buy the farmhouse?
Speaker B: (No.) The Johnsons bought the villa. (contrastive narrow focus on object)

Production studies of prosodic focus marking in adults have investigated how a word is realised in different focus and non-focus conditions. These studies have shown that languages differ in the exact prosodic means involved in focus marking (Baumann & Kügler, 2015), and transparency of how prosody is used for marking focus, defined as the probability of a form being mapped onto a focus condition. Furthermore, languages differ in whether the prosodic means involved are also used for lexical purposes, and in the importance of prosodic means for focus marking relative to other linguistic means (e.g. word order, focus particle). Past research on the acquisition of non-prosodic aspects of language has shown that cross-linguistic differences in what is encoded and how it is encoded affect both the rate (i.e. how early a form-function mapping is produced in an adult-like manner) and route of acquisition (i.e. the order in which linguistic forms are acquired) (Slobin, 1982; Guo et al., 2008). Specifically, linguistic coding of cognitively complex concepts is acquired later than that of cognitively simple concepts, but linguistic coding of similar concepts can be acquired earlier in one language than in another language if the mapping between what is coded and how it is coded (hereafter form-function mapping) is more transparent, more frequent or pragmatically more useful in that language (e.g. Slobin, 1982; Bavin, 1995). The question that arises in the context of prosodic focus-marking is whether the aforementioned cross-linguistic differences will lead to differences in the rate and route of acquisition of the prosodic realisation of the same focus type between children acquiring typologically different languages. To address this issue, we will review recent studies on children acquiring languages that exhibit the above-mentioned typological differences, including Mandarin Chinese (hereafter Mandarin), Seoul Korean (hereafter Korean), Central Swedish (hereafter Swedish), Finnish, and Dutch. These studies have been conducted with children comparable in their social economic status and ages, using similar data elicitation methods, and have largely taken an autosegmental-metrical approach to prosodic analysis (e.g. Pierrehumbert, 1980; Ladd, 1996; Gussenhoven, 2004). We will refer to previous work on children's use of prosody in German and English when necessary. In the remainder of this section, we will consider in detail the relevant typological differences and how they might affect acquisition of prosodic focus marking.

First, generally speaking, focus is realised with more prosodic prominence, compared to non-focus. Contrast and narrow scope can lead to additional prosodic prominence, compared to non-contrastive focus and broad focus. In some languages, contrastive narrow focus is realised differently than non-contrastive narrow focus (Baumann, Becker, Grice, & Mücke, 2007, on German; Hanssen, Peters, & Gussenhoven, 2008, on Dutch; Frota, 2000, on European Portuguese). However, languages differ in realisation of prosodic prominence. In line with the autosegmental-metrical framework, we distinguish two ways to achieve prosodic prominence: phonological means and phonetic means. Phonological means are defined as making coarse-grained variation in prosody, including accent placement – accenting or not accenting a word (in West Germanic languages), choice of accent type, e.g., accenting a word with a falling pitch accent (H*L) or a high-level pitch accent (H*) (in West Germanic languages), adding a prominence-marking high (H) tone (hereafter focal-H tone, following Gussenhoven and Bruce, 1999) to a word or not (in Swedish), and phrasing – inserting a phrasal boundary or not before and after a word (in Korean). Phonetic means are defined as making fine-grained changes in prosodic parameters within a phonological category, also known as phonetic implementation of a phonological category, e.g. changes in the pitch span of a pitch accent or a lexical tone, and realisation of a phrasal boundary. Some languages use both phonological and phonetic means (e.g. English and Dutch) for focus marking (Sityave & House, 2003; Katz & Selkirk, 2011; Hanssen et al., 2008), whereas other languages exclusively rely on phonetic means (e.g. Mandarin) (e.g., Xu, 1999). For example, in Dutch, narrow focus is typically realised with a falling accent whereas post-focus (i.e. constituents following focus) is typically deaccented. Pre-focus (i.e. constituents preceding focus) is usually realised phonologically in a similar way to narrow focus but is phonetically different from narrow focus. Chen (2009) found that H*L-accented subject nouns are realised with a larger pitch span, a lower pitch minimum after the peak, an earlier peak alignment, and a longer word duration in subject-narrow focus than in pre-focus. Focus types are also distinguished via phonetic means in Dutch (Hanssen et al., 2008). Considering that phonetic uses of prosody may perceptually be less salient to observe and require good control of prosodic parameters in production, we may expect children to learn to use phonetic means earlier in languages that only uses phonetic means, because of extensive exposure in the input. Alternatively, children may learn phonetic prosodic focus marking at a similar age across languages because children, regardless of their native language, may need a similar amount of time to establish the form-function mapping in the input and develop sufficient control of prosodic parameters in production.

Furthermore, languages can differ in transparency of how prosody is used for focus marking, in particular regarding the use of phonological cues. For example, in Swedish, a word is realised with a focal-H tone only when it is focal. The mapping between the placement of a focal-H tone and focus is thus highly transparent, especially regarding narrow focus and contrastive focus (e.g. Bruce, 1998, 2007). Similarly, in Korean, a focal constituent initiates a new accentual phrase, which contains one or more lexical words; post-focal constituents are frequently merged into the same accentual phrase as the focal constituent (Jun, 2005). Phrasing and dephrasing are thus consistently associated with narrow focus and post-focus respectively. In contrast, as mentioned earlier, deaccentuation is typically associated with non-focus but accentuation is used for both focus and non-focus in Dutch. The relation between accentuation and focus is thus not transparent in Dutch. Consequently, children may acquire phonological marking of focus earlier in Swedish and Korean than in Dutch.

Moreover, languages differ in whether the prosodic parameters involved in focus marking are also used for lexical purposes. Pitch is used to distinguish words in tone languages (e.g. Mandarin) and to a lesser degree, in lexical pitch accent languages (e.g. Swedish), but not in intonation languages (e.g. English, German, Dutch). For example, Mandarin has four lexical tones, i.e., a high level tone (Tone 1), a rising tone (Tone 2), a low tone (Tone 3), and a falling tone (Tone 4), which are primarily identified by pitch movements (e.g., Chao, 1965). The syllable 'ma' can mean 'mother' in Tone 1, 'hemp' in Tone 2, 'horse' in Tone 3, and 'to scold' in Tone 4. Speakers need to maintain the shape of the pitch contour in a word for the sake of the identity of the lexical tone, leaving limited acoustic space for pitch variation for focus-marking purposes. It may thus take Mandarin speaking-children more time to establish the use of pitch for focus-marking purposes in the input and require more precise control of pitch production to use pitch appropriately in focus marking than it would in the case of duration. Consequently, they may learn to use duration earlier than pitch for focus-marking purposes. On the other hand, Mandarin-speaking children produce lexical tones with considerable accuracy as early as three years of age (e.g., Wong, 2012). This implies that they may have developed considerable sensitivity to pitch variation and skills in pitch control in production, and consequently learn to use pitch earlier than duration. For children acquiring an intonation language, phonetic uses of pitch are expected to be learned earlier than duration on the assumption that phonetic implementation of a pitch accent does not affect the lexical meaning of a word and may allow more room for variation for focus marking purposes, whereas substantial variation in duration can affect rhythmic properties (e.g., Arvaniti, 2009).

Additionally, languages differ in importance of prosodic means for focus marking relative to non-prosodic means. For example, focus is mostly realised via prosody but can be realised via word order in combination with prosody in some cases in both Dutch and German. In Dutch, a focal constituent can be moved to the position immediately preceding the finite verb (i.e. the prefield), regardless of whether it is contrastive or not. The OVS word order can thus be used when the object is in narrow focus or contrastive focus. However, fronting an object to the prefield is not common in Dutch (Bouma, 2008). In German, the prefield can be filled by both a contrastive focal object or an object in broad focus in OVS sentences. Fronting a focal object to the prefield is relatively common in German (Frey, 2006; Féry, 2006), rendering possibly a less strong reliance on prosody in German than in Dutch. Word order plays a much bigger role in focus marking in Finnish than in German and Dutch (see Arnhold, Chen, & Järvikivi, 2016, for review). Unmarked SVO word order can be used in different focus conditions but marked word orders (OVS, OSV, SOV, VSO, VOS) convey differences in the position of focus and type of focus (e.g. OVS for non-contrastive narrow focus on the subject; VSO for contrastive focus on the verb). Prosody is also used for focus-marking. For example, a broad-focus sentence is typically spoken with a fall-rise pitch accent in all content words but finite verbs. But a focal word is spoken with a larger pitch span, a longer duration, and a higher intensity in narrow focus than in broad focus. Such differences in the role of prosody and word order may lead to differences in children's use of prosody in unmarked and marked word orders. We expect to observe more extensive use of prosody in marked word order OVS in Finnish than in German and Dutch.

In what follows, we will first briefly describe the methods used in the studies to be reviewed. Then we will review findings from these studies and discuss evidence for and against the aforementioned predictions on how typological differences may influence acquisition of prosodic focus marking. We will integrate the patterns emerging from our review into a cross-linguistic theory of acquisition of prosodic focus. Finally, we will suggest directions for future research.

An overview of methods

Researchers have used both corpus-based and experimental methods to sample children's production in different information structural contexts. In a corpus-based study, play sessions between the child and his mother are recorded in home settings at a regular interval for a period of time (e.g. Wieman, 1976; Behrens & Gut, 2005; Chen & Fikkert, 2007). This method is typically used with toddlers and usually yields a small number of question-answer dialogues in which reliable analysis on the information structure of the declarative sentences can be done. In an experimental study, researchers can execute better control of the information structural context, choice of words and syntactic construction, and therefore can elicit a larger amount of usable data from children of a large age range. Three tasks have been used to elicit declarative sentences from children aged three to eleven (Table 1) in the experimental studies to be discussed: the picture/video-description task, the picture-matching game, and the answer-reconstruction task. We will review these tasks below and give our analysis of the information structure of the sentence elicited when it is different from the original analysis. Since only contrastive narrow focus was embedded in these tasks, we will hereafter use the term 'contrastive focus' to refer to contrastive narrow focus.

The picture/video-description task

Hornby and Hass (1970) asked English-speaking three- to four-year-olds to describe pairs of pictures that differed by one feature (subject, verb or object) in SVO sentences (e.g. A girl rides a bike vs. A boy rides a bike). The children were prompted with the question 'What's happening in this picture?' each time before they described a picture. The use of such a question would render the

Table 1. An overview of children's ages in previous experimental studies

authors	language	3–5 years		7–8 years		10–11 years	
		mean age	age range	mean age	age range	mean age	age range
Chen	Dutch	5; 2	4; 5 ~ 5;6	7; 11	7; 5 ~ 8;10		
Romøren	Dutch	5; 2	4; 4 ~ 5; 8	8; 0	7; 1 ~ 8;11	10; 7	10; 4 ~ 12; 0
Chen & Höhle	Dutch	5; 1	4; 4 ~ 5; 5				
Hornby & Hass	English	4; 0	3; 8 ~ 4; 6				
Wonnacott & Watson	English		3; 5 ~ 4; 9				
Müller et al.	German	5;01	4;01 ~ 5;10				
Sauermann et al.	German	4; 5	4; 1 ~ 4; 9				
Arnhold et al.	Finnish	5; 1	4; 6 ~ 5; 6				
Yang; Yang et al.	Korean	5; 3	4;10 ~ 5;10	8;0	7; 4 ~ 8;10	10;10	10; 3 ~ 11;11
Yang; Yang & Chen,	Mandarin	5; 2	4; 6 ~ 5;10	7;10	7; 2 ~ 8; 3	10; 9	10; 1 – 11; 5
Romøren; Romøren & Chen	Swedish	5; 0	4; 3 ~ 5; 6	8; 3	7; 6 ~ 8; 8	10; 6	10; 0 ~ 11; 0

whole sentence focal. But considering the straightforward contrast between the two pictures in each pair, it would be more appropriate to consider the constituent carrying contrastive information (e.g. the subject in 'A boy rides a bike') as being in contrastive focus. Wonnacott and Watson (2008) used a video-clip version of Hornby and Hass' picture-description task to elicit SVO sentences from English-speaking three- to four-year olds in different information structural conditions, including contrastive focus on the subject, contrastive focus on the object, and contrastive focus on both the subject and object.

The picture-matching game

Chen (2009, 2011a) elicited SVO sentences from Dutch-speaking four- to eight-year-olds and adults by means of a picture-matching game. In this game, the experimenter told the participant that she had two sets of pictures, and one picture from one set went together with a picture from the other set. However, the pictures were mixed up; the participant's help was needed to sort out the pictures. The experimenter showed the participant one picture each time, drew the participant's attention to the entity in the picture and asked a wh-question about it (e.g., Look! A beet. Who is eating the beet?). The participant then received an answer from a virtual robot (displayed on a computer screen) via a headphone set. The robot's sentences were constructed from words recorded in a randomised word list such that they contained no sentence-level prosody (i.e. intonation and rhythmic properties). The participant was instructed to use the robot's words to answer the experimenter's question but speak normally, instead of sounding like the robot. The experimenter then looked for the matching picture and formed the complete set. Two focus conditions were included, narrow focus on the subject and narrow focus on the object.

This game has recently been adapted and used in studies of children acquiring Mandarin (Yang, 2017; Yang & Chen, 2014a, 2014b, 2018), Korean (Yang, 2017; Yang, Cho, Kim, & Chen, 2015) and Swedish and Dutch (Romøren, 2016; Romøren & Chen, 2015a, 2015b). In the adapted version, the robot was taken out to encourage more direct interaction between the experimenter and the participant. The participant was given direct access to the information that the experimenter needed and could respond directly to the experimenter's queries. This was achieved by providing the participant with his own set of pictures, each of which depicted a complete event including an agent, a patient and an action. In addition to the two focus conditions included in the earlier version of the game, three more focus conditions were embedded in the game, i.e. narrow focus on the sentence-medial constituent, contrastive focus on the sentence-medial constituent, and broad focus over the whole sentence, as illustrated in (2). Additional experimental factors were embedded in the game in different languages, such as lexical tones in the Mandarin version of the game, lexical accents in the Swedish version of the game, and number of syllables in the Korean version of the game. To ensure that the participants would use the intended words, a picture-naming task was carried out prior to the picture-matching game. In this task, each participant named each animal, personage, object and action present in the game. The entities in the pictures were thus referentially given to the participant at the start of the picture-matching game, rendering the use of a definite article in reference in the game appropriate.

(2) a. Experimenter: Look! The carrot. It looks like someone is drawing the carrot. Who is drawing the carrot?
Participant: The girl is drawing the carrot. (narrow focus on subject)

b. Experimenter: Look! The spoon. And there is also the cook. It looks like the cook is doing something to the spoon. What is the cook doing to the spoon?
Participant: The cook is drawing the spoon. (narrow focus on verb)

c. Experimenter: Look! The dog. It looks like the dog is drawing something. What is the dog drawing?
Participant: The dog is drawing a boot. (narrow focus on object)

d. Experimenter: Look! The carrot! And there is also the cat. It looks like the cat is doing something to the boot. I'll guess: the cat is licking the boot.
Participant: The cat is drawing the boot. (contrastive focus on verb)

e. Experimenter: Look, this is very blurry picture. I cannot see what's in the picture. What's happening here?
Participant: Participant: The cat is drawing the spoon. (broad focus)

The answer-reconstruction game

In an answer-reconstruction game, participants focus on reconstructing answers to certain questions by a non-human interlocutor who either cannot hear well or speak the language well. Müller, Höhle, Schmitz, and Weissenborn (2006) used an answer-reconstruction game to elicit SVO and OVS sentences with narrow focus on the subject and the object from German-speaking four- to five-year-olds and adults. In this game, the participants watched a number of three-picture comic strips together with a hand puppet. Each comic strip was accompanied by a pre-recorded narrative consisting of a three-sentence story, a wh-question, and an answer to the wh-question. Unlike the other sentences in the narrative, the answer sentence was made up of words recorded in isolation and lacked sentence-level prosody, similar to the sentences used in Chen (2009, 2011a). The participants were requested by the hand puppet, who could not hear well, to repeat the answer in each story. As the events described before and after the wh-question differed only in the subject and object referents, we analyse the narrow focus as contrastive focus.

Sauermann, Höhle, Chen, and Järvikivi, (2011), Arnhold et al. (2016), and Chen and Höhle (2018) used an answer-reconstruction game to elicit SVO and OVS sentences in different focus types, i.e. broad focus, narrow focus on the subject or the object and contrastive focus on the subject or the object from four- to five-year-olds speaking German, Finnish and Dutch. In this game, the participants watched a robot answering the experimenter's wh-questions on a number of pictures. The robot's answers lacked sentence-level prosody and had either SVO or OVS word order, which might or might not be contextually appropriate. The participants were asked to reconstruct the robot's answers in a way that they found acceptable.

Acquisition of prosodic focus marking across languages

Effects of reliance on phonetic uses of prosody

As mentioned earlier, Mandarin relies exclusively on phonetic means for focus marking. West Germanic languages use phonetic means to distinguish narrow focus from non-focus only when phonological means do not suffice, different from Mandarin, but exclusively rely on phonetic means to distinguish focus types, similar to Mandarin. Comparing Mandarin-speaking children to children speaking a West Germanic language, we can find out whether a stronger reliance on phonetic means for focus marking in a language can lead to earlier acquisition of phonetic uses of prosody in that language than in other languages.

First, we review comparable studies of prosodic realisation of narrow focus in Mandarin- and Dutch-speaking children. Chen (2009) analysed subject nouns accented with H*L in focus and pre-focus produced by Dutch-speaking four- to eleven-year-olds. She found that the Dutch-speaking children used neither pitch-related cues nor duration to distinguish the subject nouns in focus from the same nouns in pre-focus at age four or five. They were adult-like only in the use of pitch span and related parameters (i.e. pitch minimum and pitch maximum) at age seven or eight: They realised the subject nouns with a larger pitch span in the focus condition than in the pre-focus condition by only lowering the pitch minimum more in the focus condition. In contrast, Yang and Chen (2018) found that Mandarin-speaking children were fully adult-like at age four or five in using duration to distinguish focal verbs from pre-focal verbs, and showed adult-like use of pitch minimum (i.e. using a lower pitch minimum in focal verbs) in some lexical tones. They began to show adult-like use of pitch span in all lexical tones (i.e. larger pitch span in focus), pitch maximum in some lexical tones (i.e. higher pitch maximum in focus), and pitch minimum in most of the lexical tones by the age of eleven. These findings suggest that Mandarin-speaking children are faster in acquiring phonetic marking of narrow focus, even though their use of pitch may still be in development after the age of eleven.

Notably, Müller et al. (2006) showed that German-speaking four- to five-year-olds, like the adult controls, uttered subject nouns and object nouns in SVO sentences with a higher mean pitch when the nouns were in contrastive focus than when they were not focused. But as Müller et al. (2006) did not provide information on accent placement in the nouns, it is not possible to tell whether the difference in mean pitch was of a phonetic nature or driven by a phonological difference (e.g. frequent accenting in the focal nouns vs. frequent deaccenting in their non-focal renditions).

Second, we review studies of prosodic realisation of focus types in Mandarin- and English-speaking children. Yang and Chen (2018) found that like adults, Mandarin-speaking children were adult-like at age four or five in their use of duration to distinguish focus types. They realised the verbs in SVO sentences with a longer duration in the narrow focus condition than in the broad focus condition and made no difference between the narrow focus condition and the contrastive focus condition. They became adult-like in not using pitch-related cues for this purpose at age seven or eight. But they exhibited usage of pitch span in distinguishing narrow focus from broad focus at age ten or eleven, similar to adults in read speech (e.g., Xu, 1999), suggesting careful manner of speaking in this age group (Yang & Chen, 2018). Regarding English-speaking children, Hornby and Hass (1970) found that English-speaking three- to four-year-olds frequently used 'emphatic stress', probably referring to sentence-level accent with more prominence, in contrastive focus but rarely in broad focus. Using similar methods, MacWhinney and Bates (1978) found that the use of emphatic stress to mark contrast still increased between three and six years of age in English-speaking children. Recent research has provided acoustic evidence for English-speaking children's ability to encode contrastive focus with more prosodic prominence than broad focus. Wonnacott and Watson (2008) found that English-speaking three- to four-year-olds accented the subject nouns in SVO sentences regardless of information structure but, like adults, they realised the subject nouns with a higher pitch-maximum and intensity in contrastive focus than in broad focus. Unlike adults, however, they did not use duration for this purpose. Although a direct comparison between Mandarin-speaking children and English-speaking children is not possible due to methodological differences, the findings seem to suggest that Mandarin-speaking children are more developed in their use of phonetic means to distinguish focus types than English-speaking children in a similar age range (3 ~ 5 years). Mandarin-speaking children do not only use pitch-related cues but also duration, and they distinguish narrow focus from contrastive focus and broad focus. English-speaking children rely on the use of pitch-related cues in distinguishing contrastive focus from broad focus. Hence, stronger reliance on phonetic means for focus marking also appears to facilitate the use of phonetic cues for realisation of focus types, in spite of the fact that both Mandarin and English rely on phonetic means to distinguish focus types.

Effects of transparency in phonological uses of prosody

Earlier corpus-based work has yielded no conclusive evidence that children are adult-like in the use of accentuation to mark narrow focus at the age of two in West Germanic languages (see Chen, 2011b, for review). Phonological focus marking in older children (aged 4 to 11 years) has recently been examined in Korean (Yang, 2017; Yang et al., 2015), Swedish (Rømoren, 2016; Romøren & Chen, 2015b) and Dutch (Chen, 2011a; Romøren, 2016). The use of phonological means was examined in sentence-medial position in all three of these languages, in sentence-final position in Swedish and Dutch, and in sentence-medial position in Korean and Dutch.

In sentence-medial position, the Swedish-speaking children were adult-like at the age of four or five in pronouncing the verbs with the focal-H more frequently in the focus condition than in the pre- and post-focus conditions. A very small but statistically significant difference was, however, found between the adults and the four- and five-year-olds in that the latter used the focal-H slightly less frequently in focus than the adults. Such a difference was not observed in the seven- to eight-year-olds and ten- to eleven-year-olds. The Korean-speaking children were adult-like at four or five in more frequently dephrasing the object noun phrases in the pre- and post-focus conditions than in the focus condition. In contrast, the Dutch-speaking children (Romøren, 2016) were not adult-like in their use of accentuation at four or five: they failed to accent more frequently in the focus condition than in the post-focus condition. They became adult-like only at age ten or eleven. But unlike the adults, the Dutch-speaking ten- to eleven-year-olds accented the focal verbs more frequently than the pre-focal verbs, possibly a consequence of being self-conscious about their production (Romøren, 2016). Furthermore, the four- to five-year-olds were not adult-like in choice of accent type. Descriptive statistics showed that the adults had a preference for H*L in the focus condition and a preference for !H*L in the post-focus condition, whereas the four- to five-year-olds showed a similar preference for H*L and !H*L in the focus condition and a similar preference for !H*L and !H* in the post-focus condition (Rømoren, 2016).

In sentence-final position, the Swedish-speaking children were adult-like at age four or five in adding the focal-H to the object nouns primarily in the focus condition and only occasionally in the post-focus condition (Rømoren, 2016). The Dutch-speaking children were adult-like in accenting the object nouns more frequently in the focus condition than in the post-focus condition at age four or five (Rømoren, 2016). They appeared to be also adult-like at age four or five in showing a preference for !H*L in focus and deaccentuation in post-focus, followed by !H*L. However, Chen (2011a) found that the Dutch-speaking children did not develop an adult-like preference for H*L in focus until the age of seven or eight. The difference in the results between the two studies of Dutch-speaking children may be related to a difference in the data elicitation method (Romøren, 2016). Chen (2011a) used the 'robot' version of the picture game; Romøren (2016) used the 'robot-free' version of the game, which might have encouraged more informal speech in the adults and children and consequently more uses of !H*L than H*L.

In sentence-initial position, Chen (2011a) found that the Dutch-speaking four- to five-year-olds were like the adults in accenting the subject nouns in both focus and non-focus conditions and primarily with H*L, although the children accented the subject nouns slightly less frequently than adults in the focus condition. In Korean, Yang (2017) found that the children were adult-like in their use of dephrasing at age four or five, although they dephrased less frequently than adults.

Together, the findings reviewed above show that the more transparent form-function mapping between phonological means and narrow focus in Swedish and

Korean leads to an earlier mastery of phonological focus marking in these two languages than in Dutch. However, Dutch-speaking children's production is closer to adults' production in the use of accent type in sentence-initial and final positions than in sentence-medial position. Considering that adults accent the subject nouns nearly always with H*L independent of focus conditions, it is possible that children notice this regularity in Dutch from an early age. How can we account for the earlier mastery of phonological focus marking in the final position than in the medial position? It has been suggested that final focus is less effectively realised than non-final focus because prosody in the sentence-final focal word is in the first place important for the distinction between questions and statements (Liu & Xu, 2005). This would imply slower acquisition in sentence-final position. However, verbs may be a special kind of non-final constituent. Röher, Baumann, and Grice (2015) have found a clear relation between an increase in informational givenness and deaccentuation in nouns but to a much smaller degree in verbs in read-out German sentences. This finding is in line with Romøren's (2016) finding for verbs in Dutch. That is, although adults accented the verbs more frequently in the focal condition than in the non-focal conditions, the verbs were accented rather frequently in both focal and non-focal conditions (> 60%). This is clearly not the case for the sentence-final object nouns, which were accented in 82% of the cases in focus and 37% of the cases in post-focus. The mapping between accentuation and focus is thus less transparent in sentence-medial verbs than in sentence-final object nouns, and this in turn can explain the slower acquisition of prosodic focus marking in sentence-medial position in Dutch. Thus, transparency in the form-meaning mapping can also affect the rate of acquisition in different sentence-positions within the same language.

Effects of lexical use of pitch

Our review of children's phonetic focus marking has shown that, comparing the use of pitch and duration within a language, Dutch-speaking children learn to use pitch-related cues before the duration cue (Chen 2009), whereas Mandarin-speaking children acquire the use of duration earlier than the use of pitch-related cues (Yang & Chen, 2018). These findings support the prediction that the use of pitch for lexical purposes in Mandarin makes it more difficult for children to vary pitch within a lexical tonal category for focus-marking purposes.

Effects of relative importance of prosody and word order in focus marking

Arnhold et al. (2016), Sauermann et al. (2011), and Chen and Höhle (2018) investigated how four- to five-year-olds used prosody in different focus types, i.e. broad focus, narrow focus and contrastive focus, in both SVO and OVS sentences in

Finnish, German and Dutch. Arnhold et al. (2016) found that the Finnish-speaking children realised the object nouns with a longer mean syllable duration in object-narrow focus than in subject-narrow focus in both SVO and OVS sentences. Further, they produced a noun with a larger pitch span when it was in contrastive focus than when it was in narrow focus and broad focus in OVS sentences. These findings thus suggest that Finnish four- to five-year-olds use both duration and pitch span to some degree to distinguish contrastive focus and narrow focus from non-focus or broad focus in both the SVO and OVS. German-speaking children use pitch and duration to a lesser degree than Finnish-speaking children. Sauermann et al. (2011) found that the German-speaking children used only pitch in the subject nouns to distinguish focus types. They realised the subject nouns with a higher pitch-maximum and wider pitch span in subject-contrastive focus and subject-narrow focus than in broad focus in SVO sentences and with a wider pitch span in subject-contrastive focus and subject-narrow focus than in broad focus in OVS sentences. Dutch-speaking children use pitch and duration to an even more limited degree than German-speaking children. Chen and Höhle (2018) reported that the Dutch-speaking children only realised the subject nouns with a higher pitch-maximum and a longer duration in subject-contrastive focus than in broad focus and with a higher pitch-maximum in subject-contrastive focus than in subject-narrow-focus in SVO sentences. These findings thus show that four- to five-year-olds' use of prosody to distinguish focus types reflects the relative importance between word order and prosody in a language. A more equal role of word order and prosody can encourage more extensive use of prosody in marked word order, whereas a more dominant role of prosody can restrict the use of prosody in unmarked word order.

Conclusions

The overarching question of this article is how typological differences in prosodic focus marking and prosodic systems affect acquisition of prosodic focus marking. To address this question, we have reviewed the use of phonological and phonetic means for realising narrow focus and distinguishing focus types in children aged two to eleven years acquiring a range of languages including Mandarin, Korean, Swedish, Finnish, English, German and Dutch.

In the light of the findings reviewed above, we propose a cross-linguistic theory of the acquisition of prosodic focus marking. The core of this theory is that acquisition of prosodic focus marking is influenced by (at least) four types of typological differences: reliance on phonetic means, transparency of form-function mapping between phonological cues and focus, whether the prosodic means involved are also used for lexical purposes, and the importance of prosodic means relative to non-prosodic means for focus marking. The first two types of typological differences affect the rate of acquisition; the latter two affect the route of acquisition, as illustrated in Figure 1.

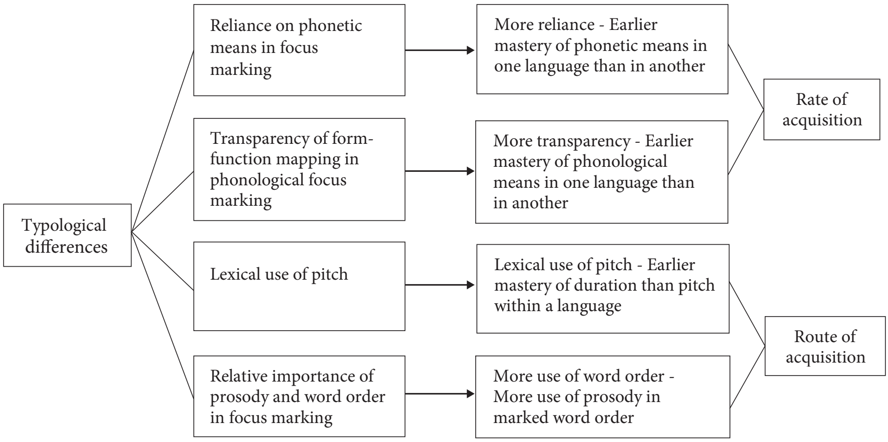


Figure 1. A cross-linguistic model of the acquisition of prosodic focus marking

Specifically, children acquiring languages that exclusively rely on phonetic means for focus marking acquire the use of phonetic means – both in distinguishing narrow focus from non-focus and in distinguishing focus types – at an earlier age than children acquiring languages that use both phonological means and phonetic means for focus marking (e.g. Mandarin vs. English and Dutch). Furthermore, children acquire phonological encoding of narrow focus at an earlier age in languages with more transparent mapping between phonological means and focus conditions (e.g. Swedish and Korean vs. Dutch). The effect of transparency of form-function mapping is also present in phonological marking of focus in different sentence positions within the same language. That is, children acquire phonological focus marking earlier in the sentence position where the form-function mapping is transparent than in the sentence position where the form-function mapping is blurry (e.g. narrow focus in sentence- final position vs. sentence-medial position in Dutch). Moreover, children learn to use pitch-related cues for focus-marking purposes later than duration cues if pitch is also used for lexical purposes (e.g. Mandarin vs. Dutch). Finally, the relative importance between prosody and word order for focus marking has an effect on children's use of phonetic means in distinguishing focus types in SVO and OVS. Four- to five-year-olds acquiring languages with a more common use of word order next to prosody in focus marking use prosody more extensively and are less restricted by the word order of the sentences (e.g. Finnish) than children acquiring languages with prosody playing a primary role in focus marking (e.g. German and Dutch).

Future research

The existing work has generated substantial insight into the acquisition of prosodic focus marking in typologically different languages. However, we are just at the beginning of understanding how children learn to use prosody to encode focus. Future research is needed in several directions. First, the majority of the earlier work is concerned with children aged four to eleven years. Children are adult-like in various aspects of prosodic focus marking already at age four or five across languages. This indicates that substantial development has taken place before children turn four. Given that most children start to produce two-word utterances at two and encoding focus is very useful pragmatically in multi-word utterances, we suggest that more research should be done on toddlers' use of prosody in different focus conditions in different languages (cf. Romøren, 2016). Corpus-based methods can be used to obtain production data from toddlers. Controlled elicitation methods have been used on three-year-olds. For example, Chen and Fikkert (2017b) (2007, September) reported the use of a picture-description task to elicit Adjective+Noun constructions with focus on the adjective (e.g. What kind of bear is it? A cute bear). This task can be adopted to elicit adjective+noun phrases with focus on the noun and other two-word constructions.

Second, past work is primarily concerned with children acquiring languages that predominantly use prosody for information structural purposes. More studies of children acquiring free word order languages, e.g. Turkish (İşsever, 2003), where prosody is used in tandem with word order, similar to Finnish, and Greek, where prosody has a primary role and word order a secondary role (Keller & Alexopoulou, 2001), similar to German, are needed to advance our understanding of the effect of relative importance between word order and prosody on the acquisition of prosodic focus marking over time.

Third, researchers have so far only studied children's prosodic focus-marking in sentences uttered as statements, typically in a friendly manner. But how do children encode focus prosodically in questions, where prosody is also used to express interrogativity and the information structure can be less straightforward than in statements? How do children learn to express affect and encode focus at the same time? Ito (this volume) has pointed out that research on acquisition of 'affect prosody' has been separated from reach on acquisition of 'focus prosody' in comprehension and called for a 'holistic' approach to study the interaction between affect and focus prosody. Such separation also applies for research on acquisition of prosodic realisation of illocutionary acts (e.g. questions vs. statements) and affect on the one hand and acquisition of prosodic realisation of focus on the other hand. We suggest that a holistic approach is needed – whereby prosodic realisation of focus in speech production is studied in interaction with prosodic realisation of illocutionary acts and affect – in order to obtain a more comprehensive picture of children's prosodic encoding of focus in natural conversations. Relatedly, new data elicitation methods are needed to study children's prosodic focus marking in questions and in different affect conditions.

Finally, this review has shown the usefulness of using comparable data-elicitation methods and a common theoretical framework in comparisons between children acquiring different languages. Recent research has yielded first evidence that children's vocabulary size and musical ability are related to their ability to use prosody to encode focus at certain ages and in certain focus conditions (Chen, 2016, May). This finding suggests that it may be helpful to control for children's vocabulary size and musical ability in cross-linguistic comparisons on prosodic focus marking. However, a lack of standardised tests in the languages of interest prevents researchers from collecting comparable data on development in other areas. Also, our understanding of how prosodic development is connected to development in other areas, linguistic or non-linguistic, is still rather limited. This poses a difficulty to researchers in determining what should be taken into account when making cross-linguistic comparisons on prosodic focus marking. In response to these difficulties, we suggest that adopting a common data elicitation method and theoretical framework for prosodic analysis and information structure should be the first step towards interpretable cross-linguistic studies of prosodic focus marking.

Funding

The article was written with the support of a VIDI grant (276-89-001) from the NWO (Netherlands Organisation for Scientific Research).

Acknowledgements

I thank Karlijn Blommers for editing the references, the editors, and an anonymous reviewer for their comments and Shari Speer for comments and suggesting example sentences.

References

Arnhold, A., Chen, A., & Järvikivi, J. (2016). Acquiring complex focus-marking: Finnish four- to five-year-olds use prosody and word order in interaction. *Frontiers in Psychology*, 7, 1886.

Arvaniti, A. (2009). Rhythm, timing and the timing of rhythm. *Phonetica*, 66(108), 46–63. doi:10.1159/000208930

Behrens, H., & Gut, U. (2005). The relationship between prosodic and syntactic organization in early multiword speech. *Journal of Child Language*, 32(1), 1–34. doi:10.1017/S0305000904006592

Bavin, E. L. (1995). Language acquisition in crosslinguistic perspective. *Annual Review of Anthropology*, 24(1), 373–396. doi:10.1146/annurev.an.24.100195.002105

Baumann, S., Becker, J., Grice, M., & Mücke, D. (2007). Tonal and articulatory marking of focus in German. *Proceedings of the 16th International Congress of Phonetic Sciences* – Saarbrücken, 1029–1032.

Baumann, S., & Kügler, F. (2015). Prosody and information status in typological perspective – Introduction to the special issue. *Lingua*, 165, 179–182. doi:10.1016/j.lingua.2015.08.001

Bouma, G. (2008). *Starting a sentence in Dutch: A corpus study of subject- and object-fronting*. Unpublished PhD dissertation. University of Groningen.

Bruce, G. (1998). *Allmän och svensk prosodi*. Lunds Universitet: Institutionen för Lingvistik.

Bruce, G. (2007). Components of a prosodic typology of Swedish intonation. In T. Riad, & C. Gussenhoven (Eds.), *Tones and tunes. Typological and comparative studies in word and sentence prosody* (Vol. 1, pp. 113–146). Berlin: Mouton de Gruyter. doi:10.1515/9783110207569.113

Chao, Y. R. (1965). *A grammar of spoken Chinese*. Berkeley, CA: University of California Press.

Chen, A. (2009). The phonetics of sentence-initial topic and focus in adult and child Dutch. In M. C. Vigario, S. Frota, & M. J. Freitas (Eds.), *Phonetics and phonology: Interactions and interrelations* (pp. 91–106). Amsterdam: John Benjamins. doi:10.1075/cilt.306.05che

Chen, A. (2011a). Tuning information packaging: Intonational realization of topic and focus in child Dutch. *Journal of Child Language*, 38(5), 1055–1083. doi:10.1017/S0305000910000541

Chen, A. (2011b). The developmental path to phonological encoding of focus in Dutch. In S. Frota, P, Prieto, & G. Elordieta (Eds.) *Prosodic production, perception and comprehension* (pp. 93–109). Berlin: Springer. doi:10.1007/978-94-007-0137-3_5

Chen, A. (2015). Children's use of intonation in reference and the role of input. In L. Serratrice, & S. Allen (Eds.), *The acquisition of reference* (pp. 83–104). Amsterdam: John Benjamins. doi:10.1075/tilar.15.04che

Chen, A. (2016, May). *Different children different prosody: Inpidual differences in prosodic development*. Keynote lecture at Speech Prosody 2016, Boston, USA.

Chen, A., & Fikkert, P. (2007, September). Dutch 3-year-olds' use of intonation in marking topic and focus. *Poster presented at Generative Approaches to Language Acquisition 2007*, Barcelona, Spain.

Chen, A. & Fikkert, P. (2007). Intonation of early two-word utterances in Dutch. *Proceedings of the 16th International Congress of Phonetic Sciences* – Saarbrücken, 315–320.

Chen, A., & Höhle, B. (2018). *Four- to five-year-olds' use of word order and prosody in focus marking in Dutch*. *Linguistics Vanguard*.

Frey, W. (2006). Contrast and movement to the German prefield. In V. Molnár, & S. Winkler (Eds.), *The architecture of focus* (pp. 235–264). Berlin: De Gruyter. doi:10.1515/9783110922011.235

Frota, S. (2000). *Prosody and focus in European Portuguese: Phonological phrasing and intonation*. New York, NY: Garland.

Guo, J., Lieven, E., Budwig, N., Ervin-Tripp, S., Nakamura, K., & Özçalişkan, Ş. (eds.) (2008). *Crosslinguistic approaches to the psychology of language: Research in the tradition of Dan Isaac Slobin*. New York, NY: Psychology Press.

Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511616983

Gussenhoven, C., & Bruce, G. (1999). Word prosody and intonation. In H. van der Hulst (Ed.), *Word prosodic systems of the languages of Europe* (pp. 233–271). Berlin: Mouton de Gruyter.

Hanssen, J., Peters, J., & Gussenhoven, C. (2008). Prosodic effects of focus in Dutch declaratives. *Proceedings of the 4th Conference on Speech Prosody* – Campinas, 609–612.

Hornby, P. A., & Hass, W. A. (1970). Use of contrastive stress by preschool children. *Journal of Speech Language and Hearing Research*, 13(2), 395–399. doi:10.1044/jshr.1302.395

İşsever, S. (2003). Information structure in Turkish: The word order-prosody interface. *Lingua*, 113, 1025–1053. doi:10.1016/S0024-3841(03)00012-3

Jun, S. A. (2005). Korean intonational phonology and prosodic transcription. In S.-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 201–229). Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199249633.003.0008

Katz, J., & Selkirk, E. (2011). Contrastive focus vs. discourse-new: Evidence from phonetic prominence in English. *Language*, 87(4), 771–816. doi:10.1353/lan.2011.0076

Keller, F., & Alexopoulou, T. (2001). Phonology competes with syntax: Experimental evidence for the interaction of word order and accent placement in the realization of Information Structure. *Cognition*, 79(3), 301–372. doi:10.1016/S0010-0277(00)00131-1

Ladd, D. R. (1980). *The structure of intonational meaning*. Bloomington, IN: Indiana University Press.

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

Lambrecht, K. (1994). *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511620607

Liu, F., & Xu, Y. (2005). Parallel encoding of focus and interrogative meaning in Mandarin intonation. *Phonetica*, 62, 70–87. doi:10.1159/000090090

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(5), 539–558. doi:10.1016/S0022-5371(78)90326-2

Müller, A., Höhle, B., Schmitz, M., & Weissenborn, J. (2006). Focus-to-stress alignment in 4 to 5-year-old German-learning children. *Proceedings of the Generative Approaches to Language Acquisition Conference 2005* – Siena, 393–407.

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

Roberts, C. (2012). Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics*, 5(6), 1–69.

Röher, C., Baumann, S., & Grice, M. (2015). The effect of verbs on the prosodic focus marking of information status: Production and perception in German. *Proceedings of the 18th International Congress of Phonetic Sciences* – Glasgow, 1–5.

Romøren, A. S. H. (2016). *Hunting Highs and Lows: The acquisition of prosodic focus marking in Swedish and Dutch*. Unpublished PhD dissertation. Utrecht University.

Romøren, A. S. H., & Chen, A. (2015a). Quiet is the new loud: Pausing and focus in child and adult Dutch. *Language and Speech*, 58(1), 8–23. doi:10.1177/0023830914563589

Rømoren, A. S. H., & Chen, A. (2015b). The acquisition of prosodic focus marking in Central Swedish: Sorting out lexical and post-lexical tones. *Proceedings of the 18th International Congress of Phonetic Sciences* – Glasgow, 318–330.

Sauermann, A., Höhle, B., Chen, A., & Järvikivi, J. (2011). Intonational marking of focus in different word orders in German children. *Proceedings of the 28th West Coast Conference on Formal Linguistics* – Los Angeles, 313–322.

Slobin, D. I. (1982). Universal and particular in language acquisition. In E. Wanner, & L. R. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 128–172). Cambridge: Cambridge University Press.

Sityaev, D., & House, J. (2003). Phonetic and phonological correlates of broad, narrow and contrastive focus in English. *Proceedings of the 15th International Congress of Phonetic Sciences* – Barcelona, 1819–1822.

Vallduví, E., & Engdahl, E. (1996). The linguistic realization of information packaging. *Linguistics*, 34(3), 459–519. doi:10.1515/ling.1996.34.3.459

Wieman, L. A. (1976). Stress patterns of early child language. *Journal of Child Language*, 3(2), 283–286. doi:10.1017/S0305000900001501

Wong, P. (2012). Acoustic characteristics of three-year-olds' correct and incorrect monosyllabic Mandarin lexical tone productions. *Journal of Phonetics*, 40(1), 141–151. doi:10.1016/j.wocn.2011.10.005

Wonnacott, E., & Watson, D. G. (2008). Acoustic emphasis in four year olds. *Cognition*, 107(3), 1093–1101. doi:10.1016/j.cognition.2007.10.005

Xu, Y. (1999). Effects of tone and focus on the formation and alignment of pitchcontours. *Journal of Phonetics*, 27(1), 55–105. doi:10.1006/jpho.1999.0086

Yang, A. (2017). Prosodic focus-marking in Mandarin Chinese-speaking and Seoul Korean-speaking children. PhD dissertation. LOT: Utrecht.

Yang, A., & Chen, A. (2014a). Prosodic focus marking in Chinese four- and eight-year-olds. *Proceedings of 7th International Conference on Speech Prosody* – Dublin, 713–717.

Yang, A., & Chen, A. (2014b). Prosodic focus marking in child and adult Mandarin Chinese. *Proceedings of the 4th International Symposium on Tonal Aspects of Languages* – Nijmegen, 54–58.

Yang, A., & Chen, A. (2018). The developmental path to adult-like prosodic focus-marking in Mandarin Chinese-speaking children. *First Language*, 38(1), 26–46. doi:10.1177/0142723717733920

Yang, A., Cho, T., Kim, S., & Chen, A. (2015). Phonetic focus-marking in Korean-speaking 7- to 8-year-olds and adults. *Proceedings of the 18th International Congress of Phonetic Sciences* – Glasgow, 1–5.

PART 4

Prosody in bilingualism and in specific populations

CHAPTER 16

Bilingual children's prosodic development

Conxita Lleó
University of Hamburg

This chapter focuses on various domains of prosodic development – syllables, metrical feet, prosodic words, stress patterns, phonological phrases, intonation and rhythm – as they are acquired by bilingual children, exposed to two languages from birth: a societal or majority language and a minority or heritage language. The studies reviewed in this chapter discuss the bilingual acquisition of these prosodic domains and the influence of one language onto the other when children develop surrounded by two languages in contact. The discussion deals with the various aspects of the acquisition of prosody and prosodic structure described in the main section of the chapter, and reports on the outcomes of cross-language interaction: acceleration, delay, transfer, order of acquisition, and fusion.

Introduction

There are great differences between acquiring Spanish in Madrid, in Barcelona or in Hamburg, as a child. The reasons for such differences are widely known: Madrid is considered a monolingual city, Spanish being the official language there; Barcelona is bilingual, Spanish being in contact with Catalan, as both languages are official in Catalonia; Hamburg is monolingual like Madrid, but the official language is German, Spanish being only used privately in some families. There is a high probability that a child growing up in Madrid will acquire one single L1, namely Spanish, whereas a child growing up in Barcelona will probably acquire 2L1, Catalan and Spanish. There are many other possible situations, e.g. a child growing up in a Spanish speaking family in Hamburg may acquire 2L1 as well, German as the societal or majority language, and Spanish as the minority or heritage language. Such a situation is one of simultaneous bilingualism (Meisel, 1990; Montrul, 2011; Oh & Nash, 2014), whereas in case the child grows up monolingual at first, and learns a second language (L2) after a certain age, this is considered sequential bilingualism. The focus of this chapter is on simultaneous bilingualism, or 2L1.

DOI 10.1075/tilar.23.16lle

Phonology and prosody

The phonology of any human language is comprised of segments and prosody, also referred to as segmental and suprasegmental categories, respectively (Goldsmith, 1995, among others). Initial research on phonological acquisition concentrated on segments, i.e. phones and phonemes. However, non-linear phonology, with its two levels of representation – skeleton and melody – opened a new perspective that brought prosody to the forefront and it soon became another important focus of research. Especially the prosodic hierarchy, proposed in Nespor and Vogel (1986), was welcomed by researchers in the field of phonological acquisition because it gave principled support to the prosodic domain. The prosodic hierarchy made clear that segments relate to one another by means of prosody: they may be shortened or lengthened, produced with a high or a low tone, and with more or less intensity.

According to Nespor and Vogel (1986), the prosodic hierarchy places the prosodic constituents on several levels – syllable, metrical foot (Ft), prosodic word (PW), phonological phrase (PPh), and intonation phrase (IPh).[1] A distinction is generally made between prosody – stress, intonation and rhythm – and prosodic structure – syllable structure, foot structure, word structure, and phrasing (Kehoe, 2013).

In the 1970s and 1980s, when research on bilingual phonology began, one central discussion focused on whether the child's bilingual input is stored in two systems (dual system hypothesis*)* or in one single system (unitary system hypothesis*)* (Fabiano & Goldstein, 2005; Keshavarz & Ingram, 2002; Redlinger & Park, 1980). Initially, even researchers who assumed two systems posed the question as to whether the two grammars are separate from the beginning or whether there is a point in time at which they separate (Volterra & Taeschner, 1978). Nowadays most professionals doing research on the bilingual acquisition of adult phonology would agree that bilingual speakers of a heritage and a majority language possess two grammars and two lexicons, one for each language. However, it has been proposed that there is a stage in which languages are mixed and so are grammars, as well (Leopold, 1953; Vogel, 1975, for phonology).

Towards the end of the century, the view that the languages are separated from the beginning of phonological acquisition was gaining adherents (Fabiano & Goldstein, 2005; Ingram, 1981/1982; Johnson & Lancaster, 1998). Evidence favors

1. There are two more levels to the prosodic hierarchy: the mora, time unit smaller than the syllable, and the utterance, larger than the IPh, comprised of two or more IPhs; they are not included in this description, because their relevance to the data considered in this chapter is very limited.

the dual system hypothesis, which does not imply that acquisition is independent for the two languages. On the contrary, there is much interaction between the two languages of the bilingual child, which leads to the question of how one language influences onto the other, and to the predictability degree of the outcome of simultaneous bilingual acquisition.

A different view is that of Vihman (2002), who proposes that at first the child has no grammatical system. According to this position, the question whether there is one system or two cannot be posed until the child has acquired a certain mass of language. That is, if monolingual children do not have a system of grammar, how could some children (the bilinguals) have two systems? This view converges with certain proposals from Optimality Theory (OT), which posit a stage at which constraints are not active at first, but only after some exposure to language has taken place (Lleó, 2017b; see Section "Brief excursus on optimality theory and phonological acquisition" for a cursory introduction to Optimality Theory). In this chapter I adopt Optimality Theory (hereafter called "OT") as the theoretical framework from which I will discuss the findings on bilingual acquisition of prosody.

Brief excursus on optimality theory and phonological acquisition

Towards the end of last century, there was much theoretical activity in phonology, especially visible through the birth of non-linear phonology. Autosegmental phonology, metrical phonology and prosodic phonology focused on relevant aspects of phonological acquisition. The so-called Principles and Parameters Theory, formulated within Government and Binding Theory (Chomsky, 1981), was successfully applied by some scholars to the phonological acquisition data (e.g. Fikkert, 1994). However, this theory was based on rules, and evidence against rules was growing among researchers. Thus, as soon as OT appeared (Prince & Smolensky, 1993/2004), it attracted many scholars working on phonological acquisition (e.g. Pater, 1997, and many others).

In OT, constraints are the crucial operational concepts, not rules (Kager, 1999; Prince & Smolensky, 1993/2004). There are basically two types of constraints: markedness and faithfulness. An example for markedness is the NoCoda constraint, which penalizes the presence of a coda in a syllable. Faithfulness constraints control the distance between input and output, limiting the types of modifications that lexical forms may admit. An example is MaxIO, penalizing deletions. It has been argued that at the initial stages of acquisition markedness constraints are outranking (Gnanadesikan, 1995). A common situation is, thus, one in which a non-dominant constraint is violated in order to satisfy an outranking constraint. For instance, the child who produces a target word ending in a coda, e.g. *lápiz* ['la.piθ] 'pencil', probably will begin by omitting the coda, and thus violating MaxIO, given the omission of a segment (the consonant in coda position). Later on, each time that the target word is produced with the coda, MaxIO is fulfilled, but the NoCoda constraint, banning codas, is violated. For illustration, see (1) to (4).

(1) NoCoda: lexical items should not have codas
(2) MaxIO: All elements appearing in the input should be maintained in the output

Constraints are organized in hierarchies, and each hierarchy corresponds to a grammar. For example, the initial stage, with NoCoda outranking MaxIO, is represented as in (3). At a later stage, the hierarchy of constraints will be inverted as in (4), with MaxIO outranking NoCoda. According to this latter hierarchy, each time that a segment in coda position is produced, NoCoda is violated.

(3) NoCoda >> MaxIO
(4) MaxIO >> NoCoda

This entails many violations of the markedness constraint (1), in case there are many codas in the target language, which leads to constraint demotion, i.e. a constraint that undergoes many violations must be taken away from its outranking position (Pater, 1997; Pater & Barlow, 2003). This is the fate of markedness constraints, which are demoted and must find their way down, interspersed among faithfulness constraints, in agreement with Tesar and Smolensky's (1998) Error Driven Constraint Demotion Algorithm. According to this algorithm, a constraint that is often violated must be demoted. For example, CVC (or closed) syllables are marked compared to CV (or open) syllables, because each production of a coda violates the NoCoda constraint. Thus, the many violations of the NoCoda constraint led soon to its demotion in German, and in Spanish, as well, under the pressure from German.[2]

Aims and structure of the present chapter

This chapter describes the acquisition of prosody by bilingual children, focusing on prosodic structure, i.e. beginning with syllable structure and stress patterns, and going into prosody in a more general sense: intonation and rhythm. All descriptions are based on literature about the acquisition of European languages (beginning in the 1980s) and focus on simultaneous bilingual acquisition of prosody (monolingual research on these same domains is discussed in Demuth,

2. The great frequency of codas in German leads to the early mastery of codas by the bilingual children. Note that the OT explanation is based on frequency and markedness (see Lleó 2017b).

this volume, Kehoe, this volume, and Post & Payne, this volume, among others). After considering the development of these domains in childhood, the Discussion summarizes the results of prosodic acquisition found in this chapter, and examines the cross-language interactions and the outcomes of such interaction in early bilingual phonology.

Bilingual acquisition of prosody

Bilingual acquisition of the syllable

Children produce their first words around the time of their first birthday. This milestone is preceded by several months, in which the child produces what is generally known as canonical babbling (Vihman, 1996). Studies on bilingual babbling are rare. A study of three simultaneous Portuguese-Swedish bilingual siblings, exposed to English as L3 at the moment of schooling, constitutes an exception (Cruz-Ferreira, 1999). The author finds what she calls *prosodic mixes*: "the intrusion of prosodic patterns of one language into another". Although the study focuses on the analysis of words, the author briefly mentions that "from around 0;4 the children stopped babbling non-Swedish and non-Portuguese sounds, and started producing definite gibberish in each language" (Cruz-Ferreira, 1999: 3). Depending on the interlocutor, gibberish dialogues sounded more like Portuguese or more like Swedish, as e.g. when talking to the Portuguese-speaking parent they replicated vowel reduction and penultimate stress, and produced the palatal lateral (nonexistent in Swedish). When talking to the Swedish-speaking parent they produced glottal fricatives (nonexistent in Portuguese). Moreover, if the children were asked to produce a word in Swedish, which they only knew in Portuguese, "they would hum a Swedish double tone on a schwa-like vowel" (Cruz-Ferreira, 1999: 3–4).

Although no words are recognizable in babbling, the child produces syllables, comprised of C(onsonant) and V(owel). The CV syllable is the simplest one, with a non-branching onset and a non-branching rhyme, i.e. onset and rhyme correspond to one single node each. A syllable with a single node in the rhyme (a non-branching rhyme) is an open syllable, whereas a syllable with a rhyme comprised of a V and a coda ((C)VC) is more complex, it is a closed syllable, heavy or Strong. The CV syllable in a language like German or English with vowel length distinction can be of two types: with a short vowel, the nucleus is non-branching and the syllable is W(eak); and with a long vowel, it has a branching nucleus and the syllable is S(trong). Note that such languages are quantity sensitive: they require that stressed syllables be S (i.e. the rhyme be comprised of a long vowel or of a (short or long) vowel plus a consonant (VC). Syllables and syllabic structures of monolinguals are discussed in Demuth (this volume) and Kehoe (this volume).

Bilingual acquisition of closed syllables: Codas

German has about 67% of closed syllables (Meinhold & Stock, 1980), whereas Spanish has about 26.5% (Delattre, 1965: 42). In German, many child words are monosyllabic, and the coda of monosyllables is then required to fulfill binarity at the syllabic level, i.e. in case the rhyme is otherwise only comprised of a short vowel. Lleó, Kuchenbrandt, Kehoe, and Trujillo (2003) focused on the acquisition of codas by monolingual and bilingual children between the ages of 1;1 to 2;4. Results show that monolingual German speaking children produce more than 80% of codas after 1;11, but monolingual Spanish do not yet produce 50% of codas at 2;4. Extrapolating from these results, two outcomes of cross-language interaction in relation to codas were hypothesized: either (a) more codas would be produced by the bilinguals in Spanish as compared to those produced by Spanish monolinguals, or (b) bilinguals in German would produce fewer codas than German monolinguals. The first hypothesis turned out to be correct: bilinguals produced more than 50% of codas in Spanish after 1;9. Thus, German-Spanish bilinguals acquire syllable codas faster in Spanish, compared to Spanish monolinguals.

In another study of a Portuguese-French bilingual child, Almeida, Rose and Freitas (2012) showed that word medial codas emerged at the same time (2;4) in both languages. The bilingual child only produced fricatives in coda position and she began to produce liquid codas in both languages at 2;9. Also, results showed that plosives (which only appear in French in coda position) were the last codas to be produced in both languages (3;1). Here, French received some influence from Portuguese, with a delaying effect in the development of word-medial codas. This conclusion on coda acquisition is based on the fact that French monolinguals are reported to produce all types of consonants in word-medial codas (Dos Santos, 2007; Rose, 2000).

Bilingual acquisition of complex syllables: Consonant clusters

Almeida et al. (2012) also analyzed the consonant clusters produced by the Portuguese-French bilingual child, whose societal language was Portuguese and heritage language was French. Their results confirmed the order of acquisition obtained for consonant clusters in previous studies (Kehoe, Hilaire-Debove, Demuth, & Lleó, 2008; Núñez-Cedeño, 2008), namely, that on the one hand, obstruent+lateral (Cl) was produced before obstruent+rhotic (Cɾ) in Portuguese, converging to the monolingual development of French (and Spanish, for that matter). On the other hand, lack of vowel epenthesis shown by this bilingual child illustrates an acceleration effect of branching onsets in Portuguese. In general, when children are acquiring consonant clusters, they go through a stage in which they delete one of the consonants of the cluster. However, children acquiring Portuguese maintain both consonants of the cluster, but epenthetic vowels are produced between the consonants, which converts complex or branching onsets into simple or non-branching ones. Thus, Almeida et al. (2012) argue that the influence shown by the development of bilinguals may go in one direction, from the societal language to the heritage language, or vice versa, from the weak onto the strong language. Clearly, these results greatly differ from those of Lleó et al. (2003), which showed acceleration of coda production under the influence of the societal language (German). In Almeida et al. (2012) cross-language interaction goes in both directions: from the societal (Portuguese) to the heritage language (French) with delay of medial coda production in French, and from the heritage language to the societal language with acceleration of branching onsets (consonant clusters) in Portuguese.

Bilingual acquisition of metrical feet and stress patterns

In the prosodic hierarchy the next prosodic constituent over the syllable is the metrical foot. According to Hayes (1995: 40) the foot is the "minimal bracketed unit" of Metrical Theory. Feet are generally comprised of two syllables. In the prosody of natural languages, two types of feet are considered: trochees (SW, as e.g. *spider* [ˈspaj.dəɹ]), and iambs (WS, as e.g. *giraffe* [ʒəˈɹaf]). Allen and Hawkins (1980) formulated the trochaic bias hypothesis, by which children at the initial stage of acquisition would prefer trochees, both in perception and production. Since then, there has been much discussion as to whether the trochaic bias is a universal principle or a phenomenon conditioned by the target language. This question is difficult to answer, because there is a majority of trochees in most target languages studied (such as English, German, Italian or Spanish).

The first studies on this topic seemed to favor the universality of the trochee. Lleó and Arias (2006) analyzed data produced by a German-Spanish simultaneous bilingual child from 1;9 to 2;6. The data contained mainly trochees and some iambs. The goal of the analysis was to find possible stress deviations from the Spanish adult trochaic pattern. Their findings show that this pattern is already acquired at the age of 1;8 as the errors found involve only a few iambs produced as trochees. In the literature on Metrical Phonology and on its acquisition, it is widely assumed that the acquisition of stress in languages like Spanish, Catalan, English or German takes place by means of some algorithm, rather than on a lexical basis (Fikkert, 1994; Klein, 1984; Prieto, 2006), i.e. whether stress is "memorized" word by word by the child, or whether it is assigned in an automatic way, by means of some sort of algorithm. The occurrence of stress errors in child language should provide evidence in support of an algorithm-based stress theory (Hochberg, 1988).

Children acquiring languages like Spanish and Catalan produce trochees before they produce iambs, in agreement with the target languages (Prieto & Vanrell, 2007). Thus, in such trochee-dominant languages, truncating the initial syllable of WS and WSW is expected. In the case of WSW, a perfect (syllabic) trochee remains after truncation of the unfooted syllable (see next section), namely SW (Piñeros, 2000); and in the case of WS, truncation leads to the so-called moraic trochee, S, i.e. a syllable with a heavy or strong rhyme, comprised of a long vowel, or a (short or long) vowel plus a consonantal coda.

The main difference between the Catalan-speaking and the Spanish-speaking children is that Catalan-speaking omit the unstressed initial syllable in WS structures for a longer time than Spanish-speaking children, although Catalan contains more disyllabic WS forms than Spanish. Prieto (2006) hypothesizes that Catalan-speaking children are more sensitive to the foot structures than to the PW structures (see Demuth, this volume, for a discussion of prosodic words). Even for a language like Hebrew, with stress on the word-final syllable, a preference for trochees has been reported (Ben-David, 2012). As mentioned by Kehoe (this volume) and Prieto (2006), the young Hebrew children's behavior was comparable to that of Catalan and Spanish children, as WS and WSW were produced with truncation of the initial weak syllable. Such a result might seem to converge to the claim of a universal trochaic bias.

Recently, Bijeljac-Babic, Höhle and Nazzi (2016) by means of the head-turn preference procedure (Jusczyk, Mandel, Myers, Turk, & Gerken, 1995) have shown that French-German 6-month-old bilinguals have a trochaic bias in perception, evidencing a preference to listen to a trochaic pattern like 'gaba versus an iambic pattern like ga'ba. This result in the perception domain converges with the results of monolingual German 6-month-olds, who also showed a preference for trochees. However, these results diverge from those of French 6-month-old monolinguals, who did not show any preferences (Höhle, 2009).

The "trochaic bias" has often been questioned, as even children exposed to English seemed not to respond to it (Vihman, de Paolis, & Davis, 1998), and showed a tendency towards iambs. This was accounted for by the frequent usage of "phrasal" iambs (i.e. phonological phrases comprised of an article or another unstressed functional word followed by a monosyllabic lexical item, e.g. *the man* [ðə'mæn]) produced by the English speaking children. Such iambic phrases have a prosodic structure comparable to the structure of an iambic foot, with stress on the final syllable: *the man* [ðə'mæn] is comparable to an iambic-shaped prosodic word *Giraffe* [ʒə'ɹaf]. More recently, on the basis of a study of an English-French bilingual child, Rose and Champdoizeau (2008) have proposed to give up the trochaic bias hypothesis and to adopt what is generally referred to as the neutral start hypothesis. According to the latter, children would not start the acquisition of prosody with a preference for trochees over iambs, but would converge towards the feet of their target-language, based on a criterion of frequency. However, the participants of that study were 2- to 3-year-old children, which is an age definitely beyond the initial state, to which the trochaic bias was referred.

The claim that iambs in English are primarily "phrasal" (Vihman et al., 1998) takes into consideration the frequency criterion. However, on the one hand, [ʒəˈɹaf] vs. [ðəˈmæn] are comparable only superficially, as their prosodic structures are in fact different, because *giraffe* is a single PW and *the man* is a PW preceded by a syllable with the status of a functional word: $[\text{ʒəˈɹaf}]_{PW}$ vs. $[[\text{ðə}]_{\sigma}\ [\text{ˈmæn}]_{PW}]_{PPh}$. On the other hand, if there were a universal trochaic bias, why would the production of a determiner plus noun be the structural basis for disyllabic lexical items? Arias and Lleó (2009) adopt a solution that makes the universality of the trochee unnecessary: Iambic-shaped words have traditionally been a challenge for theories of stress within the generative framework, especially for the metrical theory of stress, as advocated by Hayes (1995). His iambic-trochaic law (trochees may be quantity sensitive or quantity insensitive, whereas iambs must always be quantity sensitive) has inspired much work in the field, but it is controversial, as researchers have found languages that seem to base their stress rules on iambic feet in spite of not being quantity sensitive (i.e. not differentiating lexically between short and long vowels; Altshuler, 2006).[3]

Spanish has the syllabic trochee as its favorite foot. However, young children hear many iambs. Especially in child language, adverbs like *aquí* 'here' and *allí* 'there', or infinitives like *a comer* 'let's eat', *a dormir* 'let's sleep', *a jugar* 'let's play', used as imperatives, are relatively frequent. Iambs with a light last syllable (e.g. Spanish *allí*) cannot be handled according to the iambic-trochaic law, because the presence of iambs in a language presupposes quantity sensitivity, and Spanish is not such a language.[4] Children acquiring other languages like Greek or Portuguese do not seem to display any strong metrical preferences at the beginning, or else, they rather show a preference for iambs. Moreover, early words in European Portuguese are not constrained by minimality requirements, in the sense of Section "Bilingual acquisition of prosodic word structures" (Vigário, Freitas, & Frota, 2006).

3. A formulation of the law close to the original one is that of Hay and Diehl (2007): "elements contrasting in duration naturally form rhythmic groupings with final prominence, whereas elements contrasting in intensity form groupings with initial prominence". The former refer to iambs, the latter to trochees. Obviously, the tumble stone of Hayes' Metrical Theory are iambs in a quantity insensitive language. Once this constraint is demoted (i.e. looses power), the theory is able to handle such cases satisfactorily, as in Altshuler (2006).

4. The view that Spanish is a quantity-insensitive language, as I am assuming here, is not accepted by all phonologists, as e.g. Harris (1983). However, the evidence held in favor of the quantity sensitivity of Spanish is sketchy, only representing a residue of the quantity-sensitivity of Latin, the parent language.

Unfooted syllables: Their role in the expansion of prosody

Researchers have proposed that when young children begin to produce meaningful language, they are constrained to producing minimal words, comprised of one foot only (Demuth, 1996; Demuth & Fee, 1995). That is, if the target word intended by the child is comprised of a pre-tonic syllable preceding the metrical foot (e.g. the initial syllable of words of type *pelota* $[\text{pe}[\text{'lo.ta}]_{Ft}]_{PW}$ 'ball' in Spanish), such a syllable is unfooted, because it does not have a foot to which it could belong to. The fate of such a syllable tends to be truncation (i.e. syllable deletion), if the child still obeys the minimal word constraint (Lleó, 1997, 2002). According to Prieto (2006) this constraint seems to be active in languages that are quantity sensitive (Section "Bilingual acquisition of the syllable"). Lleó (2002) analyzed the unfooted syllables intended by three Spanish monolinguals, three German monolinguals, as well as by three German-Spanish bilinguals in relation to their truncation. It was found that monolingual Spanish children produced the unfooted syllable of words of type *pelota* 'ball' very soon – between 1;3 and 1;6 – whereas German-Spanish bilinguals produced them in Spanish a bit later – between 1;8 and 1;10. These results show that the presence of another language that is being acquired along with L1 may exert some specific influence on the acquisition of L1, i.e. it may cause delay in the process of acquiring a certain target-language category.

Results regarding unfooted syllables in the bilingual condition, as compared with results on codas (Section "Bilingual acquisition of closed syllables: Codas") show a crucial difference: codas increased their production in the bilingual condition, while the number of unfooted syllables was reduced. The explanation for these different results lies in frequency: codas are extremely frequent in German, which leads to acceleration in Spanish, given the frequent violations of the NoCoda constraint in German (Section "Brief excursus on Optimality Theory and phonological acquisition"). Unfooted syllables are rather infrequent in both languages, which leads to the reduction of their production by the bilinguals in Spanish.

The minimal word constraint must be soon overcome if the child wants to communicate. That is, the child prosodic system must expand. It has been shown that after producing a single foot, children acquiring English or German go into producing words comprised of two feet (Demuth & Fee, 1995), whereas children acquiring a Romance language like Spanish or Italian go from producing one foot to producing a foot preceded by an unfooted syllable (Lleó, 2002) (see Demuth, this volume, and Kehoe, this volume, for further information on this order of acquisition).

In the case of German-Spanish bilinguals, Lleó (2002) hypothesized that either (a) more unfooted syllables would be produced by the bilinguals in German, as compared to the monolinguals, under the influence of Spanish, or

(b) less unfooted syllables would be produced by the bilinguals in Spanish, under the influence of German, as compared to the unfooted syllables produced by the Spanish monolinguals. The second hypothesis was correct, as bilinguals truncated the unfooted syllable of Spanish words for a longer period of time, as long as the monolinguals truncate it in German. In the case of the societal language, German, truncation applied at the same rate as in monolingual acquisition. Figure 1 shows group results of percentages of unfooted syllable truncation, the curve with circles showing the group percentages of monolingual German and the one with triangles the group percentages of monolingual Spanish. Figure 2 shows such percentages for individual German-Spanish bilingual children in Spanish.

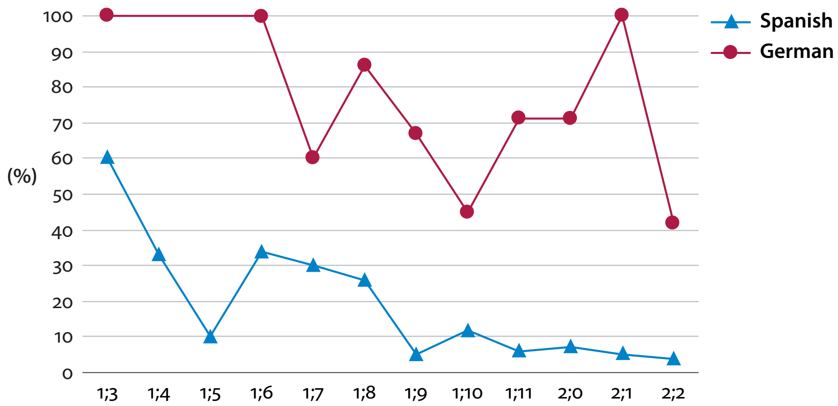


Figure 1. Percentages of unfooted syllable truncation by Spanish and German monolinguals (adapted from Lleó, 2002: 298).

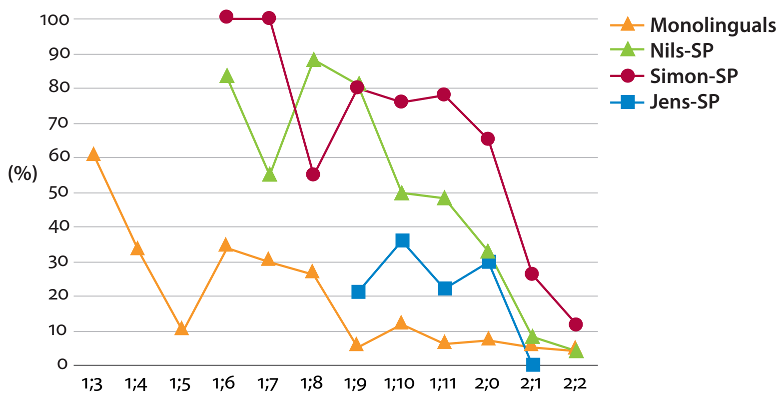


Figure 2. Percentages of unfooted syllable truncation by Spanish monolinguals and three bilinguals in Spanish (adapted from Lleó, 2002: 305).

Bilingual acquisition of prosodic word structures

Lleó (2006) analyzed the development of prosodic word structures as produced by three Spanish monolingual children, three German monolingual children, and three German-Spanish bilingual children between the ages of 1;1 and 2;2. Results showed that in the case of Spanish monolinguals, the evolution of word structures goes from disyllables to trisyllables, and later on to monosyllables, whereas bilinguals in Spanish also begin with disyllables, but from there they go to producing monosyllables and later on trisyllables. The order followed by the German-Spanish bilinguals in Spanish is comparable to the order of German monolinguals, who begin producing disyllables and monosyllables and finally produce trisyllables. This order difference is expected in the bilingual condition, given the presence of another language, which may exert influence on certain aspects of L1 acquisition.

Whereas bilingual acquisition of the syllable was clear-cut in the German-Spanish bilinguals (with acceleration in the bilingual condition) and in the Portuguese-French bilingual child (Almeida et al., 2012) (i.e. codas were delayed in the heritage language, French, and branching onsets were accelerated in the societal language, Portuguese), acquiring metrical feet seemed a more difficult endeavor to predict, especially because of the many variables to consider.

Bilingual acquisition of intonation

The earliest target-language specific characteristics of intonation have been recently discovered in the prosody of infants' cry, which is dependent on the language spoken in the newborn's surroundings. Mampe, Friederici, Christophe, and Wermke (2009) claim that newborns' cry melody is shaped by their native language. They analyzed the crying patterns of 30 French and 30 German newborns. The French infants produced rising contours, whereas the German infants preferred falling contours, similarly to the tendencies in the target languages. Production data by bilinguals are not available at such young age. However, some target-like productions by as young as 12-month-old monolingual infants have been studied in Catalan and in Spanish by Prieto, Estrella, Thorson, and Vanrell (2011). The study was based on productions by four Catalan speaking children and two Spanish speaking children between 0;11 and 2;4. Prieto et al. (2011) have witnessed some intonation contours that develop before two-word combinations are ready to appear. The authors claim that intonation acquisition goes ahead of the acquisition of syntax and is independent of grammatical development.

In target Spanish there is a crucial intonation difference between the contour of a final or nuclear intonational phrase (IPh) and a non-final or pre-nuclear phonological phrase (PPh): In broad focus declarative sentences the final phrase has a falling contour, whereas the non-final one has a rising contour and a delayed peak

(Face, 1999; Navarro Tomás, 1944). This is not the case, though, in a language like German, where both types of phrases, final and non-final, have a falling contour (Féry, 1993). The alignment of pitch peaks by bilingual children was studied further in Lleó, Rakow, and Kehoe (2004). The authors analyzed the contours of pre-final and final phrases of broad focus declaratives. The study was conducted with two German monolinguals, one Spanish monolingual and two German-Spanish bilinguals; all children were between 2;10 and 3;1 years old. Lleó et al. (2004) found that German-Spanish bilinguals tend to avoid the rising contour with delayed peak (L+>H* in ToBI terms) and they replace it with a falling pitch accent, typical of phrase-final positions.

These and further studies of children acquiring Spanish (e.g. Lleó et al., 2004) found that monolingual children produce the rising contour of non-final phrases shortly after their second birthday, whereas German-Spanish bilinguals do not produce the rising non-final pitch accent before age 3;0 in Spanish (Lleó et al., 2004). Lleó and Rakow (2011) compared declaratives and yes/no-interrogatives in Spanish and in German monolingual children, and German-Spanish bilingual children between the ages of 2;0 and 3;0. The analysis of declaratives in comparison to yes/no-interrogatives showed that both monolinguals and bilinguals, 2- and 3-year-olds have command of the pitch accents of declaratives and interrogatives. However, bilinguals tend to produce similar values in both languages, and do not reach command of all scaling and alignment differences before age 3;0 (Face, 2004) (see Frota & Butler, this volume, Esteve-Gibert & Prieto, this volume, and Armstrong & Hübscher, this volume, for more information on intonation development in monolinguals).

In the case of declaratives and yes/no-interrogatives, the alignment differences of the initial and final contours as well as the scaling differences of the first peak are delayed for the bilinguals, who have not yet acquired these differences at age 3;0, while monolinguals had already acquired these characteristics at the age of 2;0 (Lleó & Rakow, 2011). Overall, results on bilinguals show a delay of several months in the acquisition of alignment and scaling. Lleó (2016) analyzed wh-questions produced by three monolingual German, two monolingual Spanish and two German-Spanish bilinguals, all of them at about 3;0 years of age, with one adult participating as control in each language. Although both in German and Spanish wh-questions tend to be produced with a falling final contour by the adults, the German monolinguals and the bilinguals produced more rising than falling contours.

Queen (2001) analyzed the intonation patterns of six German-Turkish 10-to-12-year-old bilingual children. Recordings had been done over a nine-month period. The results of this study show that bilingual children use two rising contours with stress on the low tone both in German and Turkish, emerging from the interaction between the two languages: L*HH% (similar to a German contour), used for discourse cohesion and L*H% (similar to a Turkish contour), used to signaling continuation.

In many Spanish varieties, declaratives and yes/no-questions can have similar word order but they differ only in intonation: declaratives have a falling final contour and yes/no-questions a rising one (Table 1). The same is true for American English. However, in Cuban Spanish yes/no-questions have a falling pattern. In fact, the Cuban variety spoken in Miami can have the falling contour in both declaratives and yes/no-questions (the latter also showing a rising pattern). Alvord (2010) carried out a study focusing on the varieties spoken in Miami in order to better understand the distribution of the intonation patterns of absolute interrogatives there. Participants were 25 Cuban women, living in Miami, divided into three groups. Group 1 was born in Cuba and emigrated to the US after becoming 11 years old; group 2 was born in the US or arrived there before the age of 6; and group 3 was born in the US to at least one parent of group 2. Results showed that the first group preferred the Cuban falling contour, whereas the second group had a strong preference for the rising pattern and thus showed a stronger influence of American English. Unexpectedly, the third group hardly produced any questions with rising intonation, although this group corresponds to simultaneous bilinguals (2L1), who are proficient in both languages, and even show a preference for English. The author postulates that this result, with its dominant falling pattern, may be related to affirming their Cuban identity vs. the American rising pattern.

Table 1. Comparison of intonation contours in Spanish and German declaratives and interrogatives (adapted from Lleó & Rakow, 2011: 267).

Spanish	
Declaratives	Interrogatives
1. Lower initial F0 peak at post-tonic	1. Higher initial F0 peak
2. Medial F0 rise	2. Medial F0 fall
3. Final stressed syllable high F0	3. Final stressed syllable low F0
4. F0 fall at the end (BT)	4. F0 rise at the end
German	
Declaratives	Interrogatives
1. Initial F0 peak at stressed syllable	1. Higher initial F0 peak
2. Medial F0 rise	2. Medial F0 fall
3. Final stressed syllable high F0	3. Stressed syllable low F0
4. F0 fall at the end	4. F0 rise after last stress

Bilingual acquisition of rhythm

In the recent history of prosody, there have been several approaches to rhythm, especially to rhythm metrics. One of the most applied measurements is based on consonantal intervals (PVI-C) and on normalized differences in vocalic intervals (nPVI-V) (Grabe & Low, 2002; Low, Grabe, & Nolan, 2001). Other approaches are based on different metrics, as for example percentages of vowels (%V), VarcoV, which also incorporates normalization of speech rate, and is thus comparable to nPVI-V (Post & Payne, this volume, and White & Mattys, 2007, discuss acquisition of rhythm in monolinguals).

The analysis of rhythm in the bilingual condition generally involves some rhythm metrics out of those just mentioned. Spanish is considered to be syllable-timed, while German and English are said to be stress-timed, showing more variable intervals (Abercrombie, 1967; Pike, 1945; Ramus, Nespor, & Mehler, 1999). Figure 3 shows that the PVIs of Spanish child utterances appear closely grouped, while the PVIs of German child utterances occupy a larger extension in the PVI coordinates (see Figure 3).

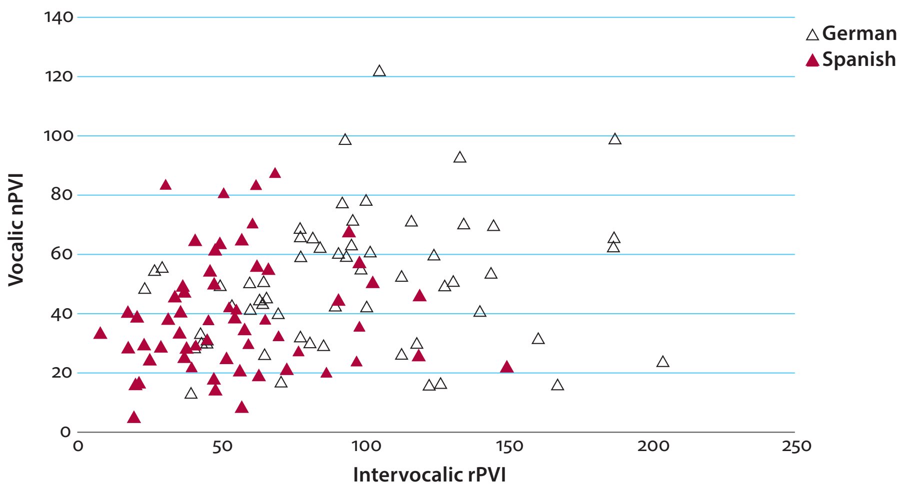


Figure 3. PVI scores for the monolingual German and Spanish children (adapted from Kehoe, Lleó, & Rakow, 2011: 339).

Bunta and Ingram (2007) investigated the acquisition of rhythm of English-Spanish bilingual children at 4 and 5 years old, and found significant language differences both for the vocalic and intervocalic PVIs: Intervals were concentrated on the low values in Spanish, and on higher values in English. There also was a preference for syllable timing over stress timing, and the values of intervals in stress timing tended to become similar to those of syllable timing. In the case of German-Spanish bilinguals (with German as the majority language and Spanish as the heritage language), Kehoe et al. (2011) found that bilinguals displayed similar rhythmic patterns in both languages. Variability between vocalic intervals was reduced in German, i.e. resembling syllable timing, but consonantal variability increased, compared to monolinguals. Moreover, the values of syllable timing in Spanish became more stress-timed, which was interpreted as influence of the majority language onto the heritage language.

Thus, several results in bilingual production led to compromise values and consequent lack of differentiation between the two languages of the bilinguals, who needed more time than the monolinguals to differentiate the languages through their rhythm features (Bunta & Ingram, 2007; Gut, 2000; Kehoe et al., 2011; Lleó, Rakow, & Kehoe, 2007; Mok, 2011; Schmidt & Post, 2015). According to Mok (2011), and Kehoe et al. (2011), bilinguals have a certain disadvantage. However, when the ambient language and its effects were shut off, cross-language interaction in bilinguals exhibited some advantage, represented in faster mastery of consonantal intervals (Schmidt & Post, 2015; cf. Kehoe et al., 2011). The alleged advantage represented by consonantal intervals is not as clear as the disadvantage represented by vocalic intervals (Bunta & Ingram, 2007; Kehoe et al., 2011), because vocalic intervals are considered better measures for stress timing than consonantal intervals.

Discussion

Summary of results

Research on bilingual prosody has led to some clear-cut results especially in the domain of the acquisition of syllabic structure, revealing that children begin acquisition with the production of open syllables, and then go over to closed syllables, and that bilingual children show a fast development of closed syllables if they are frequent in at least one of the languages to be acquired. Lleó et al. (2003) found acceleration in the acquisition of codas by bilinguals in Spanish, as compared to the Spanish monolinguals. The finding on codas was unexpected (Section "Bilingual acquisition of closed syllables: Codas"), because closed syllables are marked, and marked categories are expected to be acquired later than unmarked ones. However, as briefly discussed in Section "Brief excursus on optimality theory and phonological acquisition", the many violations against a markedness constraint like the NoCoda constraint will lead soon to its acquisition, expressed in terms of demotion.

However, Almeida et al. (2012) in their study of a Portuguese-French bilingual child found a delaying influence of Portuguese on French in the production of codas.

French has more types of consonants in the coda than Portuguese, which created the expectation that the Portuguese-French bilingual child would produce more codas than the Portuguese monolinguals. However, French has only 24% of closed syllables (Delattre and Olsen, 1969), and thus the number of codas in Portuguese and French is not as high as the number of codas in German and Spanish, and NOCODA is not demoted as soon as in the German-Spanish bilingual case. On the basis of this finding, the authors consider that the explanation provided for the German-Spanish bilinguals (Lleó et al., 2003) is not able to account for the Portuguese-French case. In order to have acceleration of coda production, the presence of the relevant category (here codas) must have a frequency crossing beyond a certain threshold; otherwise, the presence of codas is not strong enough to stimulate coda production, and concomitantly accelerate NOCODA demotion. At least in one of the languages the category in question must reach a frequency of more than 50% of occurrence (e.g. in relation to codas, German has about 67% and Spanish 26.5% (although the definition of such a threshold requires further research).

Regarding the acquisition of consonant clusters, they all go through a stage in which they are reduced to one single consonant (Lleó & Prinz, 1996; Pater & Barlow, 2003; Vivar, 2012). Thus, the syllabic level seems to be a strong, steady level of structure, well organized and responding to systematic criteria like simplicity, and constraint demotion as a consequence of constraint violation.

Research on the bilingual acquisition of stress patterns led to less clear results, most of the studies investigating the presence or absence of a "trochaic bias". Bijeljac-Babic et al. (2016) investigated the early sensitivity of stress patterns in German-French bilingual infants at 6 months of age. The authors argue that "the development of a trochaic bias in simultaneous bilinguals is not delayed compared to monolingual German learning infants." They thus conclude that infants at this age seem to have a trochaic bias, because they prefer trochees even if they are exposed continuously to the iambic input of French. If French and Hebrew monolinguals do not show any bias (neither trochaic nor iambic) (Ben-David, 2012), this could mean that the trochaic bias is not strong enough to manifest itself under the impact of iambic input, but it is strong enough not to be overruled. The difficulty with French is that many researchers have come to the conclusion that in this language words are not trochaic nor iambic, because French does not have word stress (Verluyten, 1982). Some words receive phrasal stress, displayed by prominence of the last syllable of a phrase (phonetically manifested as lengthening), which makes them appear as iambic (stressed on the final syllable).

Intonation, especially its differences across languages, seems to develop very soon, as it has recently been shown that German and French newborns' cries can already be differentiated by means of their falling or rising contours. However, in the bilingual condition some of the intonation differences of monolinguals are blurred, bilinguals often substituting the contours of their societal or strongest language for the heritage language contours, as e.g. Spanish children substitute final contours for pre-final ones. At the beginning of word production falling contours seem to be preferred over rising contours. However, it has been found that German-Spanish bilingual children at 3;0 prefer to produce wh-questions by means of rising contours (Lleó, 2016), in spite of the adult falling model for questions (Lleó & Rakow, 2011). The child seems to operate with a redefinition of question intonation requiring rising contours, which is probably inspired by the high frequency of yes/no-questions, and they generally show rising intonation.

Findings on rhythm are also indeterminate, because bilinguals use compromise values with a consequent lack of differentiation between the two languages. They needed more time than the monolinguals to differentiate the languages through their rhythmic features. In perception, it is not clear-cut either whether children at first prefer syllable-timed or stress-timed rhythm.

Various outcomes of bilingual prosodic acquisition

In bilingual acquisition, where two languages are in close contact with one another, the question is what are the cross-linguistic outcomes of this interaction. Paradis and Genesee (1996) proposed three possible results of cross-language interaction: delay, acceleration and transfer. Further outcomes have been added to these: order of acquisition (Lleó, 2006), and fusion (Queen, 2001; see Lleó & Cortés, 2013, as well). In the following paragraphs we discuss these possible outcomes in relation to the findings reported in Section "Summary of results" of this chapter.

Lleó (2006) proposes a different order of acquisition under bilingualism. Two categories that are acquired in the order A then B by monolinguals may be acquired as first B then A by bilinguals in that same language. An example is the development of prosodic word structures. Lleó (2006) showed that monosyllabic words are acquired faster by German-Spanish bilinguals than by Spanish monolinguals, due to the influence of German, which has more monosyllabic lexical items. A case of acceleration is found in the bilingual acquisition of coda consonants. Lleó et al. (2003) found that German-Spanish bilingual children produce more codas in Spanish than Spanish monolinguals, and this could be explained as a frequency effect: codas are extremely frequent in German, which leads to accelerating the demotion of the NoCoda constraint.

However, a delay effect is observed when focusing on the bilingual acquisition of unfooted syllables. It has been shown that monolingual Spanish children produce the unfooted syllable of words of type *pelota* 'ball' sooner (between 1;3 and 1;6) than German-Spanish bilinguals produce them in Spanish (between 1;8 and 1;10) (Lleó, 2002). Unfooted syllables are rather infrequent in both languages, which leads to the reduction of their production by the bilinguals in Spanish, as they do not reach the threshold necessary to demote the relevant constraint. This does not show that bilingual acquisition necessarily leads to generalized delay. It rather shows that the presence of another language that is being acquired along with L1 may exert some specific influence on the acquisition of L1, i.e. it causes delay in the process of acquiring a certain target language category. The acquisition of declarative and yes/no-interrogative intonation in bilinguals is also delayed, the alignment differences of the initial and final contours as well as the scaling differences of the first peak are delayed for the bilinguals.

Another example of delay is that of the bilingual acquisition of rhythm. Kehoe et al. (2011) found that the German of German-Spanish bilinguals tends to be less stress-timed (it shows less vocalic interval variability) than the German of German monolinguals, and their Spanish is less syllable-timed than that of Spanish monolinguals (it shows more vocalic interval variability). However, children in this study were 2 to 3 years old, and at that early stage it is not clear yet whether children (even in L1 acquisition) have reached the adult-like patterns (Post & Payne, this volume).

Transfer consists in incorporating a grammatical property of one of the languages into the other language. Mennen (2004) and Olson (2013) found evidence of transfer in bilingual acquisition of intonation, particularly in terms of the development of alignment and scaling. Rising intonation with delayed peak (L+>H* in ToBI terms) is produced in pre-nuclear (or non-final) phrases in Spanish (Face, 1999). The substitution of final pitch accent for the rising L + > H* with delayed peak can be characterized as transfer, too (see Lleó, 2017b).

Queen (2001) discusses the notion of fusion in relation to intonation, which implies a two-way influence between two languages. Her findings on German-Turkish bilinguals indicate that bilingual children use the same two rising contours both in German and Turkish: the rising L*HH% pattern, which is similar to a German contour, is used for discourse cohesion and the rising L*H% pattern, which is similar to a Turkish contour, is used to signaling continuation. Clearly, the two contours are extended from one language into the other, i.e. they are a product of the interaction between the two languages, which has also been dubbed *blending* (Kehoe, 2013).

Conclusions

This chapter presented the acquisition of prosody by simultaneous bilingual children, i.e. children exposed to two languages from birth. Often one of the languages is the strong or societal language, whereas the other is the weak or heritage language. With the goal to unifying the various topics related to prosody, Prosodic Phonology (with its prosodic hierarchy) provided a solid basis to bring out the notions of prosody and prosodic structure. The analysis considered the prosodic constituents starting with the smallest one, the syllable (except for the mora), and proceeding to the larger ones, the metrical foot, the prosodic word, the phonological phrase, and the intonation phrase. Results presented in the chapter have also been briefly discussed with respect to OT, which has enabled us to frame results theoretically. (See Lleó 2017b for a more detailed OT treatment.)

From all constituents, the syllable is the most established and best understood one, followed by the foot, which is crucial in relation to stress patterns. Phonological words represent lexical items as they exit the lexicon. Finally, intonation offers a variety of prosodic positions, on which phonological segments are located. The analysis of rhythm involves e.g. vocalic and consonantal intervals and their mutual relations. The last two domains (intonation and rhythm) tend towards compromise values in the productions of bilingual children, somehow approximating both languages to one another. And they pose principled questions like whether syllable-timed rhythm is universally less marked than stress-timed rhythm. These questions are difficult to answer, because in the bilingual condition there appears some influence in both directions, from syllable- to stress-timing and conversely, from stress- to syllable-timing.

Three further issues have been discussed. First, we have discussed the trochaic bias hypothesis in relation to the acquisition of prosody, and whether this bias is innate and universal. Bilingual infants can be a good test for such an issue, especially if they have to acquire two languages with different rhythmic structures. The first prosodic structures produced and preferred by children seem to be trochaic. However, since many of the target languages investigated are mainly trochaic, this is the structure to be expected at the initial state (Bijeljac-Babic et al., 2016). Analyzing the production of utterances by bilingual children, especially if they acquire a trochaic and an iambic language simultaneously, would provide an ideal situation to try and begin to answer this question in a fundamental way. However, there often are complicating factors. For instance, French is sometimes used as the iambic language that can be acquired simultaneously with a trochaic one. However, since French is often considered a language without lexical stress, its relevance in such studies is doubtful (see Martin, 2015; Verluyten, 1982).

Second, we have addressed whether prosodic information is organized in one or in two phonological systems by the bilingual child, an issue often discussed in the recent past, which now seems to have settled down. During the 1970s and 1980s, many studies of L1 phonological acquisition advocated the unitary model, whereas at the turn of the century, more researchers spoke for the dual hypotheses model (Ingram, 1981/82; Keshavarz & Ingram, 2002). Nowadays, it appears to be proven that the bilingual speaker organizes language knowledge in two lexica, and two grammars, which does not preclude much interaction between the two grammars of the bilingual child. However, what precedes this duality is still an open question, as some studies have concluded that initially there is no system (Vihman, 2002). And from the perspective of OT, there seems to be a stage in which neither faithfulness nor markedness constraints are yet at work (Fikkert & Levelt, 2008; Lleó, 2017a, 2017b; White, 2011).

Finally, the chapter has discussed the cross-language outcomes in early bilingual prosody. Paradis and Genesee (1996) predicted three outcomes of bilingualism – delay, acceleration and transfer – which they nonetheless did not find in the syntax. However, such outcomes have been abundantly found in phonology, especially in prosody (Mennen, 2004; Olson, 2013). Two further outcomes, order of acquisition and fusion, were added to the list, following the findings in Lleó (2017b) and Queen (2001). Our studies of simultaneous bilingual acquisition have found delay in the acquisition of unfooted syllables (Section "Unfooted syllables: their role in the expansion of prosody"), alignment and scaling of intonation (Section "Bilingual acquisition of intonation"), and rhythm (Section "Bilingual acquisition of rhythm"). At the same time, Lleó et al. (2003) describe acceleration of coda production in Spanish by German-Spanish bilinguals, due to the influence of German (Section "Bilingual acquisition of closed syllables: Codas"). Transfer of intonation has also been found, as well as fusion of intonation (Section "Bilingual acquisition of intonation").

Many cases of cross-language interaction were found. Nevertheless, much more research is necessary to draw a general view of the outcomes of languages in contact, especially in the prosodic domain. In Lleó and Cortés (2013), internal factors guiding the outcomes of cross-linguistic phonological interaction were analyzed, such as frequency of occurrence of some category, like syllabic codas; markedness of a certain category, like unfooted syllables; uniformity, because invariable forms are preferred over variable ones (allomorphs and allophones). External factors, such as age of acquisition, the language spoken at home, the societal language, the language of the peer-group, as well as typological factors, are less understood and require more research to have a complete picture of how bilingual children acquire the prosodic features of their target languages.

References

Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh: Edinburgh University Press.

Allen, G., & Hawkins, S. (1980). Phonological rhythm: Definition and development. In G. Yeni-Komshian, J. Kavanagh, & C. Ferguson (Eds.), *Child phonology: Production* (Vol. 1, pp. 227–256). New York, NY: Academic Press.

Almeida, L., Rose, Y., & Freitas, J. (2012). Prosodic influence in bilingual phonological development: Evidence from a Portuguese French first language learner. In A. K. Biller, E. Y. Chung, & A. E. Kimball (Eds.), *Proceedings of the 36th Annual Boston University Conference on Language Development* (pp. 42–52). Somerville, MA: Cascadilla Press.

Altshuler, D. (2006). Osage fills the gap: The quantity insensitive iamb and the typology of feet. *Rutgers Optimality Archive Nr.* 870.

Alvord, S. M. (2010). Variation in Miami Cuban Spanish interrogative intonation, *Hispania*, 93, 235–255.

Arias, J., & Lleó, C. (2009). Comparing the representation of iambs by monolingual German, monolingual Spanish and bilingual German-Spanish children. In K. Braunmüller, & J. House (Eds.), *Convergence and divergence in language contact situations* (pp. 205–234). Amsterdam: John Benjamins. doi:10.1075/hsm.8.09ari

Ben-David, A. (2012). The development of prosodic structure: Evidence from typical longitudinal data. *Brill's Annual of Afroasiatic Languages and Linguistics* 4, 55–79.

Bijeljac-Babic, R., Höhle, B., & Nazzi, T. (2016). Early prosodic acquisition in bilingual infants: The case of the perceptual trochaic bias. *Frontiers in Psychology*, 7, 210.

Bunta, F., & Ingram, D. (2007). The acquisition of speech rhythm by bilingual Spanish- and English-speaking 4- and 5-year-old children. *Journal of Speech, Language, and Hearing Research*, 50, 999–1014. doi:10.1044/1092-4388(2007/070)

Chomsky, N. (1981). *Lectures on Government and Binding, The Pisa Lectures*. Hawthorne, NY: Foris Publications.

Cruz-Ferreira, M. (1999). Prosodic mixes: Strategies in multilingual language acquisition. *International Journal of Bilingualism*, 3, 1–21. doi:10.1177/13670069990030010101

Delattre, P. (1965). *Comparing the phonetic features of English, German, Spanish and French*. Heidelberg: Julius Groos.

Demuth, K. (1996). The prosodic structure of early words. In J. L. Morgan, & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 171–184). Mahwah, NJ: Lawrence Erlbaum Associates.

Demuth, K., & Fee, E. J. (1995). *Minimal words in early phonological development*. Unpublished manuscript. Brown University and Dalhousie University.

Dos Santos, C. (2007). *Développement phonologique en Français langue maternelle: Une étude de cas*. Unpublished PhD dissertation. Université Lumière, Lyon 2.

Fabiano, L., & Goldstein, B. (2005). Phonological cross-linguistic effects in bilingual Spanish-English speaking children. *Journal of Multilingual Communication Disorders*, 3, 56–63. doi:10.1080/14769670400027316

Face, T. (1999). A phonological analysis of rising pitch in Castilian Spanish. *Hispanic Linguistics*, 11, 37–49.

Face, T. (2004). The intonation of absolute interrogatives in Castilian Spanish. *Southwest Journal of Linguistics*, 23, 65–79.

Féry, C. (1993). *German intonational patterns*. Tübingen: Niemeyer. doi:10.1515/9783111677606

Fikkert, P. (1994). *On the acquisition of prosodic structure*. Dordrecht: ICG Printing.

Fikkert, P., & Levelt, C. (2008). How does place fall into place? The lexicon and emergent constraints in the developing phonological grammar. In P. Avery, B.E. Dresher, & K. Rice (Eds.), *Contrast in Phonology: Perception and acquisition* (pp. 219–256). Berlin: Mouton.

Gnanadesikan, A. E. (1995). *Markedness and faithfulness constraints in child phonology*. Unpublished manuscript. University of Massachusetts at Amherst.

Goldsmith, J. A. (ed.) (1995). *The handbook of phonological theory*. Cambridge, MA: Blackwell.

Grabe, E., & Low, E. (2002). Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven, & N. Warner (Eds.), *Papers in laboratory phonology* 7 (pp. 515–546). Cambridge: Cambridge University Press. doi:10.1515/9783110197105.515

Gut, U. (2000). Cross-linguistic structures in the acquisition of intonational phonology by German-English bilingual children. In S. Döpke (Ed.), *Cross-linguistic structures in simultaneous language acquisition* (pp. 201–225). Amsterdam: John Benjamins.

Harris, J. W. (1983). *Syllable structure and stress in Spanish: A nonlinear analysis*. Cambridge, MA: The MIT Press.

Hay, J. S. F., & Diehl, R. L. (2007). Perception of rhythmic grouping: Testing the iambic/trochaic law. *Perception and Psychophysics*, 69, 113–122. doi:10.3758/BF03194458

Hayes, B. (1995). *Metrical stress theory*. Chicago, IL: The University of Chicago Press.

Hochberg, J. G. (1988). Learning Spanish stress: Developmental and theoretical perspectives. *Language*, 64, 683–706. doi:10.2307/414564

Höhle, B. (2009). Crosslinguistic perspectives on segmentation and categorization in early language acquisition. In E. Bavin (Ed.), *The Cambridge handbook of child language*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511576164.008

Ingram, D. (1981/82)The emerging phonological system of an Italian-English bilingual child. *Journal of Italian Linguistics*, 2, 95–113.

Johnson, C. E., & Lancaster, P. (1998). The development of more than one phonology: A case study of a Norwegian-English bilingual child. *The International Journal of Bilingualism*, 2, 265–300.

Kager, R. (1999). *Optimality Theory*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511812408

Kehoe, M. (2013). *The development of prosody and prosodic structure*. New York, NY: Nova Publishers.

Kehoe, M., Hilaire-Debove, G., Demuth, K., & Lleó, C. (2008). The structure of branching onsets and rising diphthongs: Evidence from the acquisition of French and Spanish. *Language Acquisition*, 15, 5–57. doi:10.1080/10489220701774229

Kehoe, M., Lleó, C., & Rakow, M. (2011). Speech rhythm in the pronunciation of German and Spanish monolingual and German-Spanish bilingual 3-year-olds. *Linguistische Berichte*, 227, 323–351.

Kemler Nelson, D. G., Jusczyk, P., Mandel, D. R., Myers, J., Turk, A., & Gerken, L. A. (1995). Brief report: The head-turn preference procedure for testing auditory perception. *Infant Behavior and Development*, 18, 111–116. doi:10.1016/0163-6383(95)90012-8

Keshavarz, H., & Ingram, D. (2002). The early phonological development of a Farsi-English bilingual child. *International Journal of Bilingualism*, 6, 255–270. doi:10.1177/13670069020060030301

Klein, H. B. (1984). Learning to stress: A case study, *Journal of Child Language*, 11, 375–390. doi:10.1017/S0305000900005821

Leopold, W. (1953). Patterning in children's language learning. *Language Learning*, 5, 1–14. doi:10.1111/j.1467-1770.1953.tb01398.x

Lleó, C. (1997). Filler syllables, Proto-articles and early prosodic constraints in Spanish and German. In A. Sorace, S. Heycock, & C. Shillcock (Eds.), *Language acquisition: Knowledge, representation and processing.* (pp. 251–256). Bingley: Emerald.

Lleó, C. (2002). The role of markedness in the acquisition of complex prosodic structures by German-Spanish bilinguals. *International Journal of Bilingualism*, 6, 291–313. doi:10.1177/13670069020060030501

Lleó, C. (2006). The acquisition of prosodic word structures in Spanish by monolingual and Spanish-German bilingual children. *Language and Speech*, 49, 205–229. doi:10.1177/00238309060490020401

Lleó, C. (2016). A preliminary study of *wh*-questions in German and Spanish child language. In Meghan E. Armstrong, Nicholas Henriksen & Maria del Mar Vanrell (Eds.), *Intonational grammar in Ibero-Romance: Approaches across linguistic subfields* (pp. 325–349). Amsterdam: John Benjamins

Lleó, C. (2017a). Aspects of the phonology of Spanish as a heritage language: From incomplete acquisition to transfer. *Bilingualism: Language and Cognition*, 1–16. doi:10.1017/S1366728917000165

Lleó, C. (2017b). German-Spanish bilinguals' phonological grammars: Permeable or resilient? In E. Babatsouli, & D. Ingram (Eds.), *Protolanguage and interlanguage*, Sheffield: Equinox.

Lleó, C., & Arias, J. (2006). Foot, word and phrase constraints in first language acquisition of Spanish stress. In F. Martínez-Gil, & S. Colina (Eds.), *Optimality-theoretic studies in Spanish phonology* (pp. 479–496). Amsterdam: John Benjamins.

Lleó, C., & Cortés, S. (2013). Modelling the outcome of language contact in the speech of Spanish-German and Spanish-Catalan bilingual children. *International Journal of the Sociology of Language*, 221, 101–125.

Lleó, C., & Prinz, M. (1996). Consonant clusters in child phonology and the directionality of syllable structure assignment. *Journal of Child Language*, 23, 31–56. doi:10.1017/S0305000900010084

Lleó, C., & Rakow, M. (2011). Intonation targets of yes/no questions by Spanish and German monolingual and bilingual children. In E. Rinke, & T. Kupisch (Eds.), *The development of grammar: Language acquisition and diachronic change* (pp. 263–286). Amsterdam: John Benjamins. doi:10.1075/hsm.11.14lle

Lleó, C., Kuchenbrandt, I., Kehoe, M., & Trujillo, C. (2003). Syllable final consonants in Spanish and German monolingual and bilingual acquisition. In N. Müller (Ed.), *(Non)Vulnerable domains in bilingualism* (pp. 191–220). Amsterdam: John Benjamins. doi:10.1075/hsm.1.08lle

Lleó, C., Rakow, M., & Kehoe, M. (2004). Acquisition of language-specific pitch accent by Spanish and German monolingual and bilingual children. In T. Face (Ed.), *Laboratory approaches to Spanish phonology* (pp. 3–27). Berlin: Mouton de Gruyter.

Lleó, C., Rakow, M., & Kehoe, M. (2007). Acquiring rhythmically different languages in a bilingual context. *Proceedings of the 16th International Congress of Phonetic Sciences* – Saarbücken, 1545–1548.

Low, E., Grabe, E., & Nolan, F. (2001). Quantitative characteristics of speech rhythm: Syllable-timing in Singapore English. *Language and Speech*, 43, 377–401.

Mampe, B., Friederici, A. D., Christophe, A., & Wermke, K. (2009). Newborn's cry melody is shaped by their native language. *Current Biology*, 19, 1944–1947. doi:10.1016/j.cub.2009.09.064

Martin, P. (2015). Lexical stress in Romance languages. In S-A. Jun (Ed.), *The structure of spoken language: Intonation in Romance* (pp. 59–119). Cambridge: Cambridge University Press. doi:10.1017/CBO9781139566391.007

Meinhold, G., & Stock, E. (1980). *Phonologie der deutschen Gegenwartssprache*. Leipzig: Bibliographisches Institut.

Meisel, J. M. (ed.) (1990). *Two first languages: Early grammatical development in bilingual children*. Dordrecht: Foris. doi:10.1515/9783110846065

Mennen, I. (2004). Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics*, 32, 543–563. doi:10.1016/j.wocn.2004.02.002

Mok, P. (2011). The acquisition of speech rhythm by three-year-old bilingual and monolingual children: Cantonese and English. *Bilingualism: Language and Cognition*, 14, 458–472. doi:10.1017/S1366728910000453

Montrul, S. (2011). Introduction: Spanish heritage speakers: Bridging formal linguistics, psycholinguistics and pedagogy. *Heritage Language Journal*, 8, i–vi.

Navarro Tomás, T. (1944). *Manual de entonación española*. New York, NY: Hispanic Institute in the United States.

Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris.

Núñez-Cedeño, R. 2008. On the acquisition of Spanish onsets: A case study. *Southwest Journal of Linguistics*, 27, 77–106.

Oh, J. S., & Nash, B. A. (2014). Attitudes and motivations of adult Spanish language learners: Comparison of heritage language learners and second language learners. *Heritage Language Journal*, 11, 29–44.

Olson, D. J. (2013). Bilingual language switching and selection at the phonetic level: Asymmetrical transfer in VOT production. *Journal of Phonetics*, 41, 407–420. doi:10.1016/j.wocn.2013.07.005

Paradis, J., & Genesee, F. (1996). Syntactic acquisition in bilingual children: Autonomous or interdependent? *Studies in Second Language Acquisition*, 18, 1–25. doi:10.1017/S0272263100014662

Pater, J. (1997). Minimal violation and phonological development. *Language Acquisition*, 6, 201–253. doi:10.1207/s15327817la0603_2

Pater, J., & Barlow, J. (2003). Constraint conflict in cluster reduction. *Journal of Child Language*, 30, 487–526. doi:10.1017/S0305000903005658

Pike, K. (1945). *The intonation of American English*. Ann Arbor, MI: University of Michigan Press.

Piñeros, C. E. (2000). Foot-sensitive word minimization in Spanish. *Probus*, 12, 291–324. doi:10.1515/prbs.2000.12.2.291

Prieto, P. (2006). The relevance of metrical information in early prosodic word acquisition: A comparison of Catalan and Spanish. *Language and Speech*, 49, 231–259. doi:10.1177/00238309060490020501

Prieto, P., & Vanrell, M. (2007). Early intonational development in Catalan. *Proceedings of the XVIth International Congress of Phonetic Sciences* – Saarbrücken, 309–314.

Prieto, P., Estrella, A., Thorson, J., & Vanrell, M. (2011). Is prosodic development correlated with grammatical and lexical development? Evidence from emerging intonation in Catalan and Spanish. *Journal of Child Language*, 38, 1–37.

Prince, A. & Smolensky, P. (1993/2004). *Optimality Theory: Constraint interaction in Generative Grammar*. Malden, MA: Wiley-Blackwell.

Queen, R. M. (2001). Bilingual intonation patterns: Evidence of language change from Turkish-German bilingual children. *Language in Society*, 30, 55–80. doi:10.1017/S0047404501001038

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73, 265–292. doi:10.1016/S0010-0277(99)00058-X

Redlinger, W., & Park, T. Z. (1980). Language mixing in young bilingual children. *Journal of Child Language*, 7, 337–352. doi:10.1017/S030500090000266X

Rose, Y. (2000). *Headedness and prosodic licensing in the L1 acquisition of phonology*. Unpublished PhD dissertation. McGill University.

Rose, Y., & Champdoizeau, C. (2008). There is no innate trochaic bias: Acoustic evidence in favor of the neutral start hypothesis. In A. Gavarró, & M. J. Freitas (Eds.), *Language acqui-*

sition and development. Proceedings of GALA 2007 (pp. 359–369). Newcastle upon Tyne: Cambridge Scholars.

Schmidt, E., & Post, B. (2015). The development of prosodic features and their contribution to rhythm production in simultaneous bilinguals. *Language and Speech*, 58, 24–47. doi:10.1177/0023830914565809

Tesar, B., & Smolensky, P. (1998). Learnability in Optimality Theory. *Linguistic Inquiry*, 29, 229–268. doi:10.1162/002438998553734

Verluyten, S. P. (1982). *Investigation on French prosodics and metrics*. Unpublished PhD dissertation. Universiteit Antwerpen.

Vigário, M., Freitas, M. J., & Frota, S. (2006). Grammar and frequency effects in the acquisition of prosodic words. *Language and Speech*, 49, 175–203. doi:10.1177/00238309060490020301

Vihman, M. M. (1996). *Phonological development: The origins of language in the child*. Oxford: Basil Blackwell.

Vihman, M. M. (2002). Getting started without a system: From phonetics to phonology in bilingual development. *International Journal of Bilingualism*, 6, 239–254. doi:10.1177/13670069020060030201

Vihman, M. M., de Paolis, R., & Davis, B. (1998). Is there a "trochaic bias" in early word learning? Evidence from infant production in English and French. *Child Development*, 69, 935–949. doi:10.2307/1132354

Vivar, P. (2012). *Adquisición de los ataques complejos en español en una muestra de niños entre 1;6 a 2;8 años: Análisis desde la Teoría de la Optimidad*. PhD dissertation. Universidad de Concepción.

Vogel, I. (1975). One system or two: An analysis of a two year-old Romanian-English bilingual's phonology. *Papers and Reports on Child Language Development*, 9, 43–62.

Volterra, V., & Taeschner, T. (1978). The acquisition and development of language by bilingual children. *Journal of Child Language*, 5, 311–326. doi:10.1017/S0305000900007492

White, L. (2011). Second language acquisition at the interfaces, *Lingua*, 121, 577–590. doi:10.1016/j.lingua.2010.05.005

White, L., & Mattys, S. L. (2007). Rhythmic typology and variation in first and second languages. In P. Prieto, J. Mascaró, & M. J. Solé (Eds.), *Segmental and prosodic issues in Romance phonology* (pp. 237–257). Amsterdam: John Benjamins. doi:10.1075/cilt.282.16whi

CHAPTER 17

Prosodic development in atypical populations

Sue Peppé
Queen Margaret University College

This chapter concerns various atypical conditions in children and how the development of prosody may be affected in them. Methods for evaluating prosodic ability are considered, taking into account receptive and expressive ability. Brief summaries of research into several conditions in which prosodic development is affected are given, including hearing impairment, language disorder, Down syndrome, childhood onset fluency disorder, Williams syndrome and a more detailed treatment of autism spectrum disorder. The chapter concludes with a brief summary of current strategies for rehabilitation.

Introduction

Chapters in this book so far have mainly been concerned with what happens in the typical acquisition of prosody, although some have touched on prosody in communication disorders. Clearly, some children with typical development (TD) will acquire prosody in different ways; the timing of acquisition will vary, for instance. In this chapter we will be considering atypical variation, i.e., children with atypical conditions whose prosody is viewed as disordered.

Prosody is atypical and has been investigated in many childhood disorders, not all of which can be considered here for reasons of space. Those that have been selected include language impairment (formerly known as specific language impairment, SLI) for the likelihood that atypical prosodic processing may be a big contributor to the condition; hearing impairment, to look at the effect of a purely receptive disorder; Down syndrome (DS), as a condition where speech motor skills are known to be affected; childhood fluency disorder (formerly known as stuttering) as a condition where the origin of speech difficulties may be mixed; and two conditions with atypical neurology from birth, Williams syndrome (WS) and autistic spectrum disorder (ASD). Particular attention has been paid to prosody in this last condition, and research into it will be examined in some detail.

DOI 10.1075/tilar.23.17pep

Summaries of research will be brief and selective, by no means exhaustive, but will include consideration of both receptive – where available – and expressive prosody. Methods of assessment are considered, and some intervention strategies with reports of their effectiveness will be discussed.

The communicative effects achieved by prosody are mostly achieved in conjunction with other means. These may be linguistic (lexical choice, word order, grammatical structures, as indicated in several previous chapters in this volume), but may also involve gestures (Rusiewicz & Esteve-Gibert, this volume); vocal quality (Campbell & Mokhtari, 2003) facial expressions (Davidson, Scherer, & Hill Goldsmith, 2009), and the practical constraints of context, all of which can have an influence on how prosody is interpreted. As Ito (this volume) points out, children may only gradually integrate semantic and pragmatic knowledge with prosodic structure. These factors must be borne in mind when considering prosody in atypical conditions, and the possibilities of intervention.

Understanding whether atypical prosody is a primary or secondary disorder is important for the planning of intervention. Earlier chapters in this volume (such as that by Teixidó, François, Bosch and Männel) have shown that prosody is fundamental to speech segmentation and the learning of language. If prosody processing in the infant is atypical, this can be a contributory cause of language impairment. On the other hand, there are many examples of disordered prosody which are the result of an atypical condition. Diminished hearing will affect how prosody is perceived, which in turn determines how it is produced. Atypical orofacial structure, as in DS, or generalised physical difficulties, as in CP, will affect the motor skills involved in prosody. Similarly, disfluency causes prosodic disturbance by impinging directly on speech-rhythm. In these cases the reasons for disorder in prosody is fairly straightforward, but the etiology of prosodic disorder is not always clear-cut. Conditions may affect children's faculties at a global, neurological or fundamental level, as in Williams syndrome (WS) and autism spectrum disorder (ASD): for instance, if children with these conditions have atypical affective prosody, it may be because their emotional reactions are atypical. At all events, the effect on prosody of atypical conditions is complex.

What is atypical prosody?

Probably the most familiar form of atypical prosody is in the speech of non-native speakers of English, which can vary greatly. The most usual observations are that the rhythm and stress-patterns are non-English, that speech is singsong or monotonous, or that stress is misplaced. Articulatory irregularities, occurring in children with motor speech disorder, interact with prosody; for example, /vʌbɒl/

may be heard as a realisation of 'rubble' until the possibility is considered of the stress being on the second syllable; then it can be interpreted as two words: 'the ball'. Additionally, segmental errors can impede the effective production of speech across multisyllabic sequences (Kent & Vorperian, 2013); and epenthetic vowels are sometimes introduced into clusters that are hard to pronounce, thereby adding syllables which disrupt speech-rhythm.

The characteristics of prosody in atypical conditions vary greatly and will be considered in more detail below, but generally involve misplaced stress or atypical realisation of stress; disruption of speech-rhythm, often involving atypical relative syllable-duration; unusual speech-melody or intonation-patterns; and disordered phrasing. Overall rate, pitch and loudness of speech may also be disordered. The different elements of prosody (length, loudness, pitch, silence) have, however, different roles to play in prosodic functions. For example, as well as having a role in stress, syllable-lengthening operates in prosodic boundary to indicate the phrasing (or 'chunking') of speech; it also affects speech-rate. There is thus a danger for intervention in taking the approach that one prosodic element, such as syllable length, needs attention: first it is necessary to determine the need for lengthening or shortening in different communicative situations.

Researchers tend to devise ad hoc procedures for their experiments. Clinical assessments of acquired disorders where prosody is impaired sometimes include protocols for testing prosody: for example, the Frenchay Dysarthria Assessment (Enderby, 1983). Such protocols are usually tailored towards the needs of the experiment or the disorder, and are not available publicly. In general terms, analysis of spontaneous speech samples, narratives, reading, and imitation tasks are used to assess expressive prosody, with disambiguation a common technique for assessing receptive skills. There is however, wide variation in the methodology used for examining prosody, making it sometimes hard to compare the results of different investigations. Since some of the results of the research into disordered prosody refer to tasks in assessment procedures, a review follows of some of the available assessment procedures, restricted to those available for practical or clinical use. For a further review of the procedures described here, see also Diehl & Paul (2009).

Assessment procedures

Generally, it is a perception that the child's speech delivery is unusual that triggers the need to examine prosody. The feedback guiding prosodic expression might be affected by atypical hearing, perception or understanding of prosody, suggesting that reception skills should be investigated first, but historically this step has not often been taken and assessment instruments tend to focus on expressive skills.

Crystal's Profiling of Prosody, or 'PROP' (Crystal, 1982) was one of the earliest formalised procedures for examining expressive speech prosody. It involved recording a sample of spontaneous speech from the child, transcribing the sample both for words and prosody, and on this evidence making judgments as to whether particular aspects of delivery were normal or atypical.

Shriberg's Prosody-Voice Screening Profile, or PVSP (Shriberg, Kwiatkowski, & Rasmussen, 1990) works on a similar principle: a sample of spontaneous conversational speech is recorded, and examined for the appropriateness or otherwise of three aspects of prosody (phrasing, speech-rate, stress) and four aspects of voice (pitch, loudness, laryngeal quality and resonance). Both of these procedures require researchers/clinicians to be able to transcribe prosody or (in the case of the PVSP) to learn codes to be assigned to sections of the samples. The advantage of this approach is that its material is natural unscripted speech. The downside is that the speaker's intention is not known but only inferred, and the content of utterances can vary greatly, making acoustic analysis complex. The PROP has been used very little for research or clinical investigation; the PVSP more so.

The Profiling Elements of Prosody in Speech-Communication assessment procedure (PEPS-C; Peppé & McCann, 2003) investigates receptive and expressive prosody skills in parallel tasks. Formal and functional levels of ability are also assessed. The functional tasks consist of perceiving or expressing meanings as conveyed by prosody by means of minimal pairs, without reference to formal designations of prosody or transcription. For example, the word 'carrots' can be said either as a question or as a statement; the two utterances are differentiated only by the intonation (in this case, rising or falling). The expressive functional tasks do not use spontaneous utterances; instead the words to be used are prescribed and uniform (and similar to those in the receptive tasks). This has the advantage that individuals' test responses can be directly compared in acoustic analysis, made simpler because they are automatically recorded. Acoustic analysis enables identification of which prosodic exponents are deviating and in what way, thus explaining how they can give rise to misleading or atypical prosody. Functions other than the question/statement distinction that are tested include affect (i.e., the use of prosody to distinguish between opinions (liking and disliking something); boundary or 'chunking' (are words grouped together in a phrase or not, as in "fruit-salad and milk" versus "fruit, salad and milk"); and focus (where stress placement indicates which word is the subject of focus or emphasis in an utterance). At a formal level, auditory discrimination tasks indicate whether testees perceive at least basic prosodic differences; and imitation tasks evaluate a testee's ability to produce the forms of prosody. The minimal pair approach ensures that utterance content is comparable across testees, and also that speaker intention is verifiable. Performance is judged by reference to that of control groups in individual experiments;

very little normative data is available. A further disadvantage is that there are no levels of difficulty that allow for maturation; task items are the same for all ages.

Van Santen, Prud'hommeaux, and Black (2009) devised an automated assessment of prosody production. They stress however that their tools were not such as could be easily adapted for clinical use or diagnosis.

Conditions with atypical prosody

Research into the atypical development of prosody presents an incomplete and patchy picture. Some conditions occur more frequently than others and so there is more reason to investigate them: more demand, more clinical need, and so more research. Other reasons are the problems that can affect all research: small sample sizes, problems with the selection of participants and diagnostic criteria, and the lack of an extensive and accurate description of participants' characteristics.

A number of conditions in which prosody is affected (childhood apraxia of speech, dyslexia, and dysarthria) have not been considered in this chapter, but a few suggestions are given here as to research concerning prosody in them. For childhood apraxia of speech (CAS): Alcock, Passingham,Watkins, & Vargha-Khadem (2000); Odell & Shriberg (2001); Boutsen & Christman (2002). For dyslexia: Goswami, Gerson, & Astruc (2010); Marshall, Harcourt-Brown, Ramus, & van der Lely (2009). For dysarthria, as in cerebral palsy (CP) or traumatic brain injury (TBI): Patel, Hustad, Connaghan, & Furr (2012); Lowit, Kuschmann, MacLeod, Schaeffler, & Mennen (2010); Whitehill, Patel, & Lai (2008).

Language impairment

Language disorder is the replacement term in the DSM-5 (American Psychiatric Association, 2013) for expressive and mixed receptive-expressive language problems, including the condition formerly known as Specific Language Impairment (SLI), a term used often in previous research. According to DSM-5, the disorder involves persistent difficulties in the comprehension or production of spoken, written, sign language, or other forms of language. Fisher (2007) and Graham and Fisher (2015) suggest that this kind of language disorder is probably genetically determined. With regard to prosody, the question arises as to whether prosodic deficit is caused by language disorder or is itself a contributory cause of language impairment.

On the expressive side, McGregor and Leonard (1994) observed that in language impairment many function words, which generally carry little stress, are lost (possibly as a result of atypical auditory processing). The effect is that instead of stressed syllables being separated by unstressed syllables, stressed syllables are often adjacent, making for speech-rhythm that sounds disordered. Snow (1998) found that children with SLI were not impaired in using syllable lengthening and pitch change to indicate finality of phrase or conversational turn, and a further study (Snow, 2001) found they could imitate rising and falling intonation contours as well as children with TD. This suggests that while children with this kind of language disorder may have atypical speech-rhythm, their use of pitch and relative syllable duration appears to be intact.

Much research into language impairment is understandably concerned with receptive abilities. Stark and Tallal (1981) proposed that auditory processing is a core deficit for children with language problems. Corriveau, Pasquini, and Goswami (2007), examining the auditory cues of amplitude rise time and duration in children with SLI, find results that suggest strongly that infants' difficulties in the accurate processing of these prosodic cues affect the acquisition of language via impaired word segmentation and the development of degraded phonological representations. Holliman, Wood, and Sheehy (2010) similarly suggest, as has been mentioned in previous chapters, that rhythmic awareness and sensitivity may be prerequisite skills for speech segmentation and for establishing phonemic representations. Poor rhythmic skills could therefore hinder normal language development. Research by Van der Meulen Janssen, and den Os (1997) found, however, that children with language impairment were not different from controls with TD at identifying emotion from intonation, suggesting that some aspects of prosodic processing are spared.

Few studies examine both receptive and expressive prosodic ability in children's language disorders to find if there is any relationship between them. One study where this might be possible, because of the comparability of the PEPS-C expressive and receptive tasks, is Wells and Peppé (2003), but the authors do not report on correlations between receptive and expressive prosodic ability.

Hearing impairment

About 2 to 3 out of every 1,000 children in the United States are born with a detectable level of hearing loss in one or both ears (Vohr, 2003). It is measured in terms of loss of decibels (dB) rather than loss of frequencies, but information about ability to hear frequencies will help determine what prosodic information is available to a person with hearing impairment. Pitch variation in speech is carried on the lower speech frequencies (<500Hz), whereas segmental information is mainly carried on frequencies >500Hz.

There are relatively few studies of receptive prosodic ability in people with unaided hearing loss. Some researchers (Jackson & Kelly, 1986; Osberger & McGarr, 1982) reported difficulties in perceiving stress patterns. Rubin-Spitz and

McGarr (1986) found that participants with severe and profound hearing loss succeeded in perceiving stress patterns only in stimuli where the stressed and unstressed syllables differed in amplitude. In contrast, Most (2000) reported an 80.3% success rate in perceiving syllable stress, attributed to the utilization of an appropriate and natural task, comprising meaningful sentences containing minimal pairs of target words that differed in their stress pattern.

Children with hearing loss produce atypical prosody. Calvert and Silverman (1983) suggest that a 90dB or poorer hearing loss is generally equated with irregular speech rhythm caused by problems controlling varying loudness, pitch and duration. These authors also suggest that changes in pitch may be perceived as overall changes in loudness; that rising pitch, for example, may be perceived as a decrease in loudness. Most (2000) found that children with hearing loss showed higher fundamental frequency, longer duration of syllables across stressed and unstressed syllables, and less success in conveying stress information. Speech rate in hearing-impaired people is often considered slow (Boone, 1966; Stathopoulous, Duchan, Sonnenmeier, & Bruce, 1986).

Cochlear implants (CIs) have improved the hearing of many individuals with hearing impairment enormously. Several studies have assessed their effect on prosodic processing (e.g., Lenden & Flipsen, 2007; Peng, Tomblin, Spencer, & Hurtig, 2007; Peng, Tomblin, & Turner, 2008). The general finding is that both perception and production of prosody in CI-implanted children are less good than in normally-hearing children, but even if cochlear implants do not result in fully normal speech, they provide significant benefits in terms of improving prosody to children with hearing impairment.

With regard to the relationship between receptive and expressive prosody, Most and Frank (1991) concluded that their data suggested that there is correlation between perception and production. The study by Klieve and Jeanes (2001) also suggests that intervention focusing on receptive ability can further improve the prosodic production in CI-implanted children (see below, Intervention).

Down syndrome (DS)

DS is a genetic disorder associated with mild to moderate cognitive impairment and characteristic facial features. The enlarged tongue, relatively small mid-face bones and reduced muscle tone in lips and cheeks (Pilcher, 1998) affect speech production. People with DS are known to have articulation deficits (Shriberg & Widder, 1990) and these, when combined with poor prosodic production, play a significant role in the poor intelligibility of people with DS.

Pettinato and Verhoeven (2009) show that perception of weak syllables is problematic in adolescents with Down syndrome; limitations of phonological processing may be responsible for prosodic abnormalities. Stojanovik (2010) found that receptive prosodic ability in children with DS was significantly better than prosodic production, in line with the DS language profile of strengths in language comprehension and weaknesses in language production. Areas of particular difficulty were the expression of affect and of pre-final narrow focus, as well as the discrimination of prosodic differences and imitation of prosodic contours.

A review of speech impairment in DS by Kent and Vorperian (2013) includes a comprehensive section on fluency and prosody. In general it is thought that individuals with DS have limitations in the perception, imitation and spontaneous production of prosodic features.

Childhood-onset fluency disorder (formerly stuttering)

This involves recurrent prolongations, reverberations, or blocks of sounds, syllables, phrases or words (Maguire, Yeh, & Ito, 2012). During these unintentional breaks in speech, the individual is not able to make sounds. The condition appears to be highly heritable (Dworzynski, Remington, Rijksdijk, Howell, & Plomin, 2007). Two examples of research, one on expressive prosody and one on receptive, are given here.

Arbisi-Kelm, Hollister, Zebrowski, and Gupta (2014) found that children who stuttered produced a narrower F0 range across utterance types and a greater degree of pre-boundary lengthening preceding relative clauses in syntactically complex sentences, as well as higher F0 variability at these juncture points, when compared with non-stuttering children. Utterance duration did not differ significantly between groups.

Research by Weber-Fox, Hampton Wray, and Arnold (2013) suggests that prosodic processing is atypical in children who stutter. In their study, evoked potentials (stimulated by a cartoon video and heard naturally spoken sentences that were either correct or contained semantic or syntactic – phrase structure – violations in the children who stuttered were characterised by longer N400 peak latencies and greater negative amplitudes.

Williams syndrome

Williams syndrome (WS) is a neurodevelopmental disorder of genetic origin. Along with mild to severe intellectual disability, individuals with WS present with relative proficiency in language in the face of serious impairments in non-verbal domains (e.g., Bellugi, Lichtenberger, Jones, Lai, & George, 2000).

An area of particular interest in WS prosody is the expression of affect. Reilly, Klima, and Bellugi (1990) found that people with WS appeared to use affective prosodic cues effectively to engage listeners and keep their attention, but that their use of affective prosodic devices was exaggerated and inappropriate for the context.

Plesa-Skwerer, Schofield, Verbalis, Faja Tager-Flusberg (2007) came to a similar conclusion. Martínez-Castilla, Sotillo, and Campos, 2011, using the Spanish version of PEPS-C (Martínez-Castilla & Peppé, 2008), found that in the expression of affect many participants with WS used the opposite intonation pattern for the affective function they wanted to express. They found, however, no evidence of poor understanding of the tasks.

Concerning other functions of prosody, Martínez-Castilla et al. (2011) found deficits in the ability to produce prosodic boundary, but Stojanovik, Setter, and van Ewijk (2007), also using the PEPS-C, found that in the production of questioning versus declarative intonation, the WS group scored better than the language-matched controls, suggesting that differentiating between questions and statements, important in the prosodic management of turn-taking (a pragmatic skill) is a relative strength in the WS profile.

Concerning other functions of prosody, Plesa-Skwerer, Faja, Schofield, Verbalis, & Tager-Flusberg (2006) found that children with WS were less able to distinguish negative emotion than controls with TD. In their 2007 study, Plesa-Skwerer et al. found that while individuals with WS were relatively good at distinguishing affective utterances without segmental content, they showed deficits on a task of distinguishing lexical stress. The authors suggest that there is perhaps dissociation between affective and linguistic prosody in WS. A study by Nazzi and Ramus (2003) suggested that development of prosodic word segmentation is seriously delayed in WS, which may help to explain late lexical onset in this population. Martínez-Castilla et al. (2011) found deficits in the ability to perceive prosodic boundary.

These findings suggest difficulty with regard to the use of prosody for linguistic understanding, and difficulty in expressing their own emotions through prosodic means: it seems that although they can produce the forms, they often misuse them in communication.

Autism spectrum disorder (ASD)

Autism is a neurodevelopmental disorder characterized by deficits in social interaction, verbal and non-verbal communication, and restricted and repetitive patterns of behaviour, interests and activities, often accompanied by sensorineural abnormalities. Kanner (1943) produced the first description of the condition. Symptoms generally appear by the age of three. In the DSM-5, Asperger's syndrome (Asperger, 1944), a term once used to designate people with high-functioning autism, is not mentioned but the term has been used in research. Prosody is not mentioned specifically as a diagnostic feature, nor is it mentioned that people with autism are sometimes non-verbal, but indicators of ASD include 'failure of normal back-and-forth conversation', 'deficits in using communication for social purposes, such as greeting and sharing information, in a manner that is appropriate for the social context', 'impairment of the ability to change communication to match context or the needs of the listener'; 'difficulties…taking turns in conversation, rephrasing when misunderstood, and knowing how to use verbal and non-verbal signals to regulate interaction'; and 'difficulties understanding what is not explicitly stated (e.g., making inferences) and non-literal or ambiguous meanings of language (e.g., idioms, humour, metaphors, multiple meanings that depend on the context for interpretation).' Many of these communication skills depend at least in part on prosody: for example, it can signal that a turn has ended and what type of response is required. Similarly, language for social purposes includes the use of prosody for conveying attitude; and prosody can be essential for indicating that language is non-literal or ambiguous.

McCann and Peppé (2003) summarise research into prosody in ASD up to 2002, but there are many aspects to be considered and the following review is divided into different topics: at the form level, studies of acoustic measures of expressive prosody (pitch and duration), which provide the basis for studies of listeners' perceptions of ASD prosody; studies of imitated prosody; studies of how expressive prosody relates to meaning (i.e., functionally); and receptive prosodic ability, i.e., how children with ASD process prosodic cues and understand functional uses of prosody.

Acoustic measures of prosody: As far as duration is concerned, Diehl and Paul (2012) found that, on one imitation subtest of the PEPS-C, the ASD group showed significantly longer duration of utterances. Diehl and Paul (2009) had previously found differences in duration of utterances for imitative and spontaneous speech in individuals with ASD. In a sentence-completion task, Grossman, Bemis, Plesa Skwerer, and Tager-Flusberg (2010) found increased duration of stressed syllables in individuals with ASD, leading to longer total duration of words than in peers with TD. Shriberg, Paul, Black, and van Santen (2011) did not find such ASD-TD differences in younger children, but it is possible that children with TD decrease the duration of syllables as their motor control increases, while children with ASD fail to achieve this decrease. Van Santen, Prud'hommeaux, Black, and Mitchell (2010) noted that duration was one of the most important acoustic features differentiating individuals with ASD from their peers with TD.

Regarding pitch, Diehl, Watson, Bennetto, McDonough, & Gunlogson (2009) found that while average speaking pitch was not significantly different, the ASD group had wider pitch range in their narrative; wider pitch range was also found by Bonneh, Levanon, Dean-Pardo, Lossos, & Adini (2010) and by Sharda et al. (2010). Diehl and Paul (2012) did not, however, find differences between ASD and groups with TD in either pitch range or other acoustic categories (intensity, average pitch, pitch range). The studies do, however, suggest that on the whole children with ASD have wide pitch-range, and this would explain descriptions such as 'exaggerated', but not 'monotonous'. The latter term, however, may refer more to speech-rhythm than to pitch, implying that there is little reduction of duration in unstressed syllables; the measurements suggest fairly consistently that duration is longer in the speech of children with ASD.

Listener's perceptions of prosody in ASD: Tager-Flusberg, Paul, and Lord (2005) describe odd intonation as 'one of the most immediately recognizable clinical signs of the [autism] disorder' (p. 348). Fay and Schuler (1980) said the speech of even highly verbal individuals with autism can be 'bizarre', while Baron-Cohen and Staunton (1994) noted that children with ASD tend not to conform to the 'regional' accent of their peers, perhaps due to imperviousness to social pressures. Hargrove (1997) describes the speech of people with autism as often characterized by poor inflection and excessive or misassigned stress. Other descriptions include: robotic, stilted, exaggerated or monotonous intonation, slow syllable-timed speech, and a fast rate of speech (Baltaxe & Simmons, 1985; Baron-Cohen & Staunton, 1994; Fay & Schuler, 1980). From this it can be seen that descriptions are often contradictory: speech can sound exaggerated *and* monotonous, slow *and* fast. Diehl et al. (2009) also noted that there was significant overlap between the ASD and control groups in pitch variation, suggesting the possibility that not all individuals with high-functioning autism (HFA) exhibit intonation patterns that are different from their peers.

A study by Nadig and Shaw (2012) suggests that while acoustic measurements of conversation and structured tasks showed increased pitch range in children with ASD, this was not a contributory factor in listeners' perceptions of their prosody as atypical. Olejarczuk and Redford (2013) studied the relative contributions of rhythm, intonation and lexical information to listeners' perceptions of autistic prosody, and conclude that rhythm cues and lexical information contribute more to perceived disorder than intonation.

Imitation in ASD: This is an area that has attracted much attention, since children with ASD show imitation deficits very early in development and because atypical mirror neurons have been thought to be influential in the development of the autism condition (Rogers and Williams, 2006). Studies using the PEPS-C imitation tasks (Järvinen-Pasley, Peppé, King-Smith, & Heaton, 2008; Peppé, McCann, Gibbon, O'Hare, & Rutherford, 2007) suggest deficits, with ASD groups scoring significantly worse than controls. Filipe, Frota, Castro, and Vicente (2014), using the PEPS-C translated into Portuguese, found that children aged 5–9 in their study had difficulty in imitating prosodic patterns.

Expressive functions of prosody in ASD: Paul, Augustyn, Klin, and Volkmar (2005) found that children with ASD had difficulty with tasks involving stress, grammatical or lexical as well as pragmatic/affective use of stress. Using the PVSP, McAlpine, Plexico, Plumb, and Cleary (2014) found that children as young as 2–6 years old produced misplaced or reduced stress. Peppé et al. (2007), using the PEPS-C, also found misplaced stress; the children with ASD tended to place stress early in utterances (in line with findings by Baltaxe & Simmons, 1985, and Baltaxe & Guthrie, 1987). Filipe et al. (2014), using the Portuguese PEPS-C, found that children with ASD aged 5–9 had difficulty effectively producing turn-end and affective intonation. Peppé et al. (2007) also found that children with ASD scored lower than children with TD on the turn-end task: they tended to sound questioning when naming or producing a statement, but their questions did not sound like statements. The greatest between-group difference in Peppé et al. (2007) was in the use of affective prosody, where their intonational expression of food likes and dislikes were mainly ambiguous, expressing no clear opinion.

Van Santen, Prud'hommeaux, Black, and Mitchell (2010) used their instrumental methods for evaluating prosody production in children with ASD and TD, and revealed that the acoustic features that predict auditory-perceptual judgment are not the same as those that differentiate between groups. Specifically, a key difference between the groups appeared to be a difference in the balance between the various prosodic cues, such as pitch, amplitude, and duration, and not necessarily a difference in the strength or clarity with which prosodic contrasts are expressed.

Shriberg et al. (2011) investigated the hypothesis that disordered expressive prosody in autism might be caused by the kind of motor speech impairments that characterise childhood apraxia of speech (CAS). Using the PVSP, they compared findings from children aged 4–7 with speech errors, speech delay, ASD, CAS and TD, but found no evidence to support this hypothesis. Instead, they posit the possibility of a problem with the speech attunement framework (Shriberg, 1975, 2010), which suggests that a child 'tunes in' to the oral communications of the ambient community and 'tunes up' the phonological and phonetic behaviours subserving intelligible and socially appropriate speech, prosody, and voice production. It is indeed likely that children with ASD would not constrain their speech by such pragmatic, social and affective considerations.

Prosodic processing and functionally receptive prosody in ASD: Receptive skills have been comparatively little researched: McCann and Peppé (2003) found 5 studies in 22 years. There has been more investigation of this aspect since.

Several fMRI studies have found that the specific regions in the brain activated by prosodic variation differ according to whether there is neurotypicality or ASD (Wang, Dapretto, Hariri, Sigman, & Brookheimer, 2001; Hesling et al., 2010). This suggests that prosody is processed differently in ASD.

Communicative functions of prosody studied include the placing of stress for information focus, prosody used for affective purposes, linguistic indications such as phrasing (boundary or chunking), and pragmatic uses such as indicating end of turn and turn-end type. Paul et al. (2005) found significant differences on their task assessing perception of place of stress. Performance was generally weaker in items where stress was a factor, and in pragmatic/affective utterances as opposed to lexical distinctions. Diehl, Bennetto, Watson, Gunlogson, & McDonough (2008) found that their group with ASD was significantly less proficient than controls when prosody alone disambiguated syntax, although they scored comparably to controls when syntax alone or both prosody and syntax indicated the correct response. In studies using PEPS-C, Peppé et al. (2007) found that while children with ASD scored similarly to controls on the focus task (perceiving place of stress), they made significantly more errors when the accent was pre-final rather than final. This ASD group scored low on understanding turn-end type, as did the children in the study by Filipe et al. (2014); Peppé et al. (2007) found that the children with ASD were likely to hear questions as statements (as in Järvinen-Pasley et al., 2008). On discrimination tasks in Peppé et al. (2007), children with ASD tended to perceive difference in items where there was none. These studies show prosodic deficit in a number of areas in ASD: in prosodic processing, in understanding some uses of prosody, in the expression of affect, the placing and realisation of stress, imitation tasks, and acoustic differences, suggesting prosody is a major area of difficulty for children with ASD.

Intervention: Some programs

With regard to the aims of intervention, it is clearly desirable for children to 'sound normal', rather as it is for synthetic speech to sound natural. Another aim is for children's use of prosody to reduce ambiguity and clarify the meaning – linguistic, pragmatic and social – of what they say. It should also be considered that children need to convey the meanings they wish to convey with their prosody, i.e., that reference should be made to the speaker's intention.

For reasons of space, interventions are merely listed here, but critical reviews can be found in Hargrove, Anderson, and Jones (2009) and Hargrove (2013). Some interventions have been available for a long time, such as Melodic Intonation Therapy (Sparks & Holland, 1976), and the Lee Silverman Voice Treatment, to be found at lsvtglobal.com. There is also the IBM SpeechViewer (e.g., Thomas-Stonell, McClean, & Dolman, 1991) which provides visual models to be matched in production activities, and visual feedback for correct production, and can thus be adapted for prosodic intervention programs.

Hargrove et al. (2009) reviewed interventions used in studies 1988–2008, with comments on their reported effectiveness. They note that while there were disappointingly few studies that met their criteria for inclusion, and the indications for effectiveness suffered from contradictions, nevertheless it was clear that prosody could be altered using intervention. Hargrove (2013) provides valuable reviews of some of the evidence-based intervention techniques publicly available, with a set of seven questions as to their nature and effectiveness. Included are: the LSVT, targeting loudness, which has been used extensively by Ramig and her colleagues (see for example Ramig, Countryman, Thompson, & Horii, 1995); Rosenbek's six steps for the treatment of affective prosody (Rosenbek et al., 2004); Samuelsson's (2011) explicit intervention approach, treating various prosodic measures; and a protocol by Ballard, Robin, McCabe, and McDonald (2010) targeting lexical stress. Hargrove concludes with some concerns: the need for clearer descriptions of procedures, research on initiation of intervention, more monitoring for trade-offs between prosody and other aspects of communication, more evidence-based measures, and more collaboration between clinicians, for which she recommends the Clinical Prosody blog (clinicalprosody.wordpress.com).

These interventions address expressive prosody, and few receptive approaches are employed. Exceptions are a study by Bellon-Harn (2011) in which receptive prosody is targeted, and one by Matsuda and Yamamoto (2013) on increasing the comprehension of affective prosody in children with ASD. Rothstein (2013) offers a Prosody Treatment Program with receptive and expressive systematic cues adapted from the Dynamic Temporal and Tactile Cueing for Speech Motor Learning (Strand, Stoeckel, & Baas, 2006). Rothstein's program has the advantage that it is available for two levels: preschool (ages 3–5) and school age (5–18). It also targets both receptive and expressive prosody skills.

Klieve and Jeanes (2001) describe a ten-week intervention programme focused on improving perception and understanding of prosodic cues in linguistic contexts. In the six children with CI implants in their study, prosody perception improved substantially and impacted on the associated skills of prosodic production. This appears to be an approach that could be investigated more.

Conclusion

In this chapter we have looked at some of the research into a selection of conditions in which children's prosody is atypical, and considered whether prosodic disorder can be thought of as primary or secondary in each. We have tried to relate the theoretical issues to the practical procedures available for assessing prosodic ability, and looked at some of the research using these and other procedures to explore prosody in the selected conditions. We have tried to establish the characteristics of the atypical prosody associated with each disorder, finding that it varies widely in its realisation and in its effect for communication. We have considered some intervention programs seeking to remedy atypical prosody.

One area that has appeared to be under-researched is the relationship between expressive and receptive prosody. Correlations between these two modes have been noted in some studies (e.g., Peppé et al., 2007), and the study by Klieve and Jeanes (2001) suggests that the transfer of prosodic awareness might improve expressive prosody. This is perhaps an area for the focus of future research into conditions in children where prosody is affected.

Acknowledgements

My grateful thanks to past colleagues who provided me with helpful comments and/or copies of relevant papers: Josh Diehl, Pastora Martinez-Castilla, Melissa Redford, Anja Lowit, Larry Shriberg, Vesna Stojanovik, and Jan van Santen; also to the anonymous reviewers and editors of this chapter; and to all those who have been participants in the various studies supporting this research.

References

Alcock, K. J., Passingham, R. E., Watkins, K., & Vargha-Khadem, F. (2000). Pitch and timing abilities in inherited speech and language impairment. *Brain and Language*, 75, 34–46. doi:10.1006/brln.2000.2323

American Psychiatric Association (2013). *Diagnostic and statistical manual for mental disorders* (5th ed., DSM-5). Washington, DC: American Psychiatric Publishing.

Arbisi-Kelm, T., Hollister, J., Zebrowski, P., & Julia Gupta, J. (2014). Prosodic variability in the speech of children who stutter. *Journal of the Acoustic Society of America*, 135, 2389. doi:10.1121/1.4877900

Asperger, H. (1944). Die 'Autistichen Psychopathen' im Kindesalter. *Archiv fur Psychiatrie und Nervenkrankheiten*, 117, 76–136. doi:10.1007/BF01837709

Ballard, K. J., Robin, D. A., McCabe, P., & McDonald, J. (2010). A treatment for dysprosody in childhood apraxia of speech. *Journal of Speech, Language, and Hearing Research*, 53(5), 1227–1245. doi:10.1044/1092-4388(2010/09-0130)

Baltaxe, C. A. M., & Simmons, J. Q. (1985). Prosodic development in normal and autistic children. In E. Schopler & G. B. Mesibov (Eds.), *Communication problems in autism* (pp. 95–125). New York, NY: Plenum Press. doi:10.1007/978-1-4757-4806-2_7

Baltaxe, C., & Guthrie, D. (1987). The use of primary sentence stress by normal, aphasic and autistic children. *Journal of Autism and Developmental Disorders*, 17, 255–271. doi:10.1007/BF01495060

Baron-Cohen, S., & Staunton, R. (1994). Do children with autism acquire the phonology of their peers? An examination of group identification through the window of bilingualism. *First Language*, 14, 241–248.

Bellon-Harn, M. L. (2011). Targeting prosody: A case study of an adolescent. *Communication Disorders Quarterly*, 2, 109–117. doi:10.1177/1525740109348793

Bellugi, U., Lichtenberger, L., Jones, W., Lai, Z., & George, M. (2000). The neurocognitive profile of Williams syndrome: A complex pattern of strengths and weaknesses. *Journal of Cognitive Neuroscience*, 12, 7–29. doi:10.1162/089892900561959

Bonneh, Y. S., Levanon Y., Dean-Pardo, O., Lossos, L., & Adini, Y. (2010). Abnormal speech spectrum and increased pitch variability in young autistic children. *Frontiers in Human Neuroscience*, 4, 237.

Boone, D. R. (1966). Modification of the voices of deaf children. *Volta Review*, 68, 686–692.

Boutsen, F. R., & Christman, S. S. (2002). Prosody in apraxia of speech. *Seminars in Speech and Language*, 23, 245–255. doi:10.1055/s-2002-35799

Calvert, D. R., & Silverman, S. R. (1983). *Speech and deafness*. Washington, DC: Alexander Graham Bell Association for the Deaf.

Campbell, N., & Mokhtari, P. (2003). Voice quality: The 4th prosodic dimension. *Proceedings of 15th International Congress of Phonetic Sciences* – Barcelona, 2417–2420.

Corriveau, K., Pasquini, E., & Goswami, U. (2007). Basic auditory processing skills and specific language impairment: A new look at an old hypothesis. *Journal of Speech, Language, and Hearing Research*, 50(3), 647–666. doi:10.1044/1092-4388(2007/046)

Crystal, D. (1982). *Profiling linguistic disability*. London: Arnold.

Davidson, R. J., Scherer, K. R., & Hill Goldsmith, H. (2009). *Handbook of affective sciences*. Oxford: Oxford University Press.

Diehl, J. J., & Paul, R. (2009). The assessment and treatment of prosodic disorders and neurological theories of prosody. *International Journal of Speech-Language Pathology*, 11(4), 287–292. doi:10.1080/17549500902971887

Diehl, J. J., Bennetto, L., Watson, D., Gunlogson C., & McDonough, J. (2008). Resolving ambiguity: A psycholinguistic approach to understanding prosody processing in high-functioning autism. *Brain & Language*, 106, 144–152. doi:10.1016/j.bandl.2008.04.002

Diehl, J. J., Watson, D., Bennetto, L., McDonough, J., & Gunlogson, C. (2009). An acoustic analysis of prosody in high-functioning autism. *Applied Psycholinguistics*, 30(3), 385–404. doi:10.1017/S0142716409090201

Diehl, J. J., & Paul, R. (2012). Acoustic differences in the imitation of prosodic patterns in children with autism spectrum disorders. *Research in Autism Spectrum Disorders*, 6, 123–134. doi:10.1016/j.rasd.2011.03.012

Dworzynski, K., Remington, A., Rijksdijk, F., Howell, P., & Plomin, R. (2007). Genetic etiology in cases of recovered and persistent stuttering in an unselected longitudinal sample of young twins. *American Journal of Speech-Language Pathology*, 16, 169–178. doi:10.1044/1058-0360(2007/021)

Enderby, P. (1983). *Frenchay dysarthria assessment*. Austin, TX: Pro-Ed.

Fay, W., & Schuler, A. L. (1980). *Emerging language in autistic children*. London: Edward Arnold.

Filipe, M. G., Frota, S., Castro, S. L., & Vicente, S. G. (2014). Atypical prosody in Asperger syndrome: Perceptual and acoustic measurements. *Journal of Autism and Developmental Disorders*, 44(8), 1972–1981. doi:10.1007/s10803-014-2073-2

Fisher, S. E. (2007). Molecular windows into speech and language disorders. *Folia Phoniatrica et Logopaedica*, 59(3), 130–140. doi:10.1159/000101771

Goswami, U., Gerson, D., & Astruc, L. (2010). Amplitude envelope perception, phonology and prosodic sensitivity in children with developmental dyslexia. *Reading & Writing*, 23, 995–1019. doi:10.1007/s11145-009-9186-6

Graham, S. A., & Fisher, S. E. (2015). Understanding language from a genomic perspective. *Annual Review of Genetics*, 49, 131–160. doi:10.1146/annurev-genet-120213-092236

Grossman, R.B., Bemis, R.H., Plesa Skwerer, D., & Tager-Flusberg, H. (2010). Lexical and affective prosody in children with high-functioning autism. *Journal of Speech, Language and Hearing Research*, 53(3), 778–793. doi:10.1044/1092-4388(2009/08-0127)

Hargrove, P. (2013). Pursuing prosody interventions. Clinical Linguistic and Phonetics, 27(8), 647–660. doi:10.3109/02699206.2013.777121

Hargrove, P., Anderson, A., & Jones, J. (2009). A critical review of interventions targeting prosody. *International Journal of Speech-Language Pathology*, 11(4), 298–304. doi:10.1080/17549500902969477

Hargrove, P. (1997). Prosodic aspects of language impairment in children. *Topics in Language Disorders*, 17, 76–83. doi:10.1097/00011363-199708000-00009

Hesling, I., Dilharreguy, B., Peppé, S., Amirault, M., Bouvard, M., & Allard, M. (2010). The integration of prosodic speech in high functioning autism: A preliminary fMRI study. *PLoS ONE*, 5(7), e11571. doi:10.1371/journal.pone.0011571

Holliman, A. J., Wood, C., & Sheehy, K. (2010). The contribution of sensitivity to speech rhythm and non-speech rhythm to early reading development. *Educational Psychology*, 30(3), 247–267. doi:10.1080/01443410903560922

Jackson, P., & Kelly, D. (1986). The relationship between word and stress pattern recognition ability and hearing level in hearing impaired young adults. *Volta Review*, 88, 279–288.

Järvinen-Pasley, A., Peppé, S., King-Smith, G., & Heaton, P. (2008). The relationship between form and function level receptive prosodic abilities in autism.*Journal of Autism and Developmental Disorders*, 38(7), 1328–1340. doi:10.1007/s10803-007-0520-z

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child*, 2, 217–250.

Kent, R. D., & Vorperian, H. K. (2013). Speech impairment in Down syndrome: A review. *Journal of Speech Language and Hearing Research*, 56, 178–210. doi:10.1044/1092-4388(2012/12-0148)

Klieve, S., & Jeanes, R. C. (2001). Perception of prosodic features by children with cochlear implants: Is it sufficient for understanding meaning differences in language? *Deafness and Education International*, 3(1), 15–37. doi:10.1179/146431501790561061

Lenden, J. M., & Flipsen, P. Jr. (2007). Prosody and voice characteristics of children with cochlear implants. *Journal of Communication Disorders*, 40(1), 66–81. doi:10.1016/j.jcomdis.2006.04.004

Lowit, A., Kuschmann, A., MacLeod, J., Schaeffler, F., & Mennen, I. (2010). Sentence stress in ataxic dysarthria: A perceptual and acoustic study. *Journal of Medical Speech-Language Pathology*, 18, 77–82.

Maguire, G., Yeh, C., & Ito, B. (2012). Overview of the diagnosis and treatment of stuttering. *Journal of Experimental and Clinical Medicine*, 4(2), 92–97. doi:10.1016/j.jecm.2012.02.001

Marshall, C. R., Harcourt-Brown, S., Ramus, F., & van der Lely, H. K. J. (2009). The link between prosody and language skills in children with specific language impairment (SLI) and/or dyslexia. *International Journal of Language and Communication Disorders*, 44(4), 466–488. doi:10.1080/13682820802591643

Martínez-Castilla, P., & Peppé, S. (2008). Developing a test of prosodic ability for speakers of Iberian Spanish. *Speech Communication*, 50, 900–915. doi:10.1016/j.specom.2008.03.002

Martínez-Castilla, P., Sotillo, M., & Campos, R. (2011). Prosodic abilities of Spanish-speaking adolescents and adults with Williams syndrome. *Language and Cognitive Processes*, 26, 1055–1082. doi:10.1080/01690965.2010.504058

Matsuda, S., & Yamamoto, J. (2013). Intervention for increasing the comprehension of affective prosody in children with autism spectrum disorders. *Research in Autism Spectrum Disorders*, 7(8), 938–946. doi:10.1016/j.rasd.2013.04.001

McAlpine, A., Plexico, L. W., Plumb, A. M., & Cleary, J. (2014). Prosody in young verbal children with autism spectrum disorder. *Contemporary Issues in Communication Science and Disorders*, 41, 120–132.

McGregor, K.K. & Leonard, L.B. (1994). Subject pronoun and article omissions in the speech of children with specific language impairment: A phonological interpretation. *Journal of Speech and Hearing Research*, 37(1), 171–181. doi:10.1044/jshr.3701.171

McCann, J., & Peppé, S. (2003). Prosody in autistic spectrum disorders: A critical review. *International Journal of Language and Communication Disorders*, 38, 325–350. doi:10.1080/1368282031000154204

Most T. (2000). Production and perception of syllable stress by children with normal hearing and children with hearing impairment. *Volta Review*, 101(2), 51–70.

Most T., & Frank Y. (1991). The relationship between the perception and the production of intonation by hearing impaired children. *Volta Review*, 93, 301–309.

Nadig, A., & Shaw, H. (2012). Acoustic and perceptual measurement of expressive prosody in high-functioning autism: Increased pitch range and what it means to listeners. *Journal of Autism and Developmental Disorders*, 42(4), 499–451. doi:10.1007/s10803-011-1264-3

Nazzi, T., & Ramus, F. (2003). Perception and acquisition of linguistic rhythm by infants. *Speech Communication*, 41, 233–243. doi:10.1016/S0167-6393(02)00106-1

Odell, K. H., & Shriberg, L. D. (2001). Prosody-voice characteristics of children and adults with apraxia of speech. *Clinical Linguistics & Phonetics*, 15, 275–307. doi:10.1080/02699200010021800

Olejarczuk, P., & Redford, M. (2013). The relative contribution of rhythm, intonation and lexical information to the perception of prosodic disorder. *Proceedings of Meetings on Acoustics*, 19(1), 060154.

Osberger, M. J., & McGarr, N. (1982). Speech production characteristics of the hearing impaired. In N. Lass (Ed.), *Speech and language advances in basic research and practice* (Vol. 8, pp. 221–283). New York, NY: Academic Press

Patel, R., Hustad, K. C., Connaghan, K. P., & Furr, W. (2012). Relationship between prosody and intelligibility in children with dysarthria. *Journal of Medical Speech Language Pathology*, 20(4), 1–5.

Paul, R., Augustyn, A., Klin, A., & Volkmar, F. (2005). Perception and production of prosody by speakers with autistic spectrum disorders. *Journal of Autism and Developmental Disorders*, 35, 205–220. doi:10.1007/s10803-004-1999-1

Peng, S.-C., Tomblin, J. B., Spencer, L. J., Hurtig, R. R. (2007). Imitative production of rising speech intonation in pediatric cochlear implant recipients. *Journal of Speech Language and Hearing Research*, 50, 1210–1227. doi:10.1044/1092-4388(2007/085)

Peng, S.-C., Tomblin, J. B., Turner, C. W. (2008). Production and perception of speech intonation in pediatric cochlear implant recipients and individuals with normal hearing. *Ear and Hearing*, 29, 336–351. doi:10.1097/AUD.0b013e318168d94d

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 and Phonetics*, 17, 345–354. doi:10.1080/0269920031000079994

Peppé, S., McCann, J., Gibbon, F., O'Hare, A., & Rutherford, M. (2007). Receptive and expressive prosodic ability in children with high-functioning autism. *Journal of Speech, Language & Hearing Research*, 50, 1015–1028. doi:10.1044/1092-4388(2007/071)

Pettinato, M., & Verhoeven J. (2009). Production and perception of word stress in children and adolescents with Down syndrome. *Down Syndrome Research and Practice*, 13(1), 48–61.

Pilcher, E. S. (1998). Dental care for the patient with Down syndrome. *Down Syndrome Research and Practice*, 5(3), 111–116 doi:10.3104/reviews.84

Plesa-Skwerer, D., Faja, S., Schofield, C., Verbalis, A., & Tager-Flusberg, H. (2006). Perceiving facial and vocal expressions of emotion in Williams syndrome. *American Journal on Mental Retardation*, 111(1), 15–26. doi:10.1352/0895-8017(2006)111[15:PFAVEO]2.0.CO;2

Plesa-Skwerer, D., Schofield, C., Verbalis, A., Faja, S., & Tager-Flusberg, H. (2007). Receptive prosody in adolescents and adults with Williams syndrome. *Language and Cognitive Processes*, 22, 247–271. doi:10.1080/01690960600632671

Ramig, L. O., Countryman, S., Thompson, L. L., & Horii, Y. (1995). Comparison of two forms of intensive speech treatment for Parkinson disease. *Journal of Speech and Hearing Research*, 38, 1232–1251. doi:10.1044/jshr.3806.1232

Reilly, J., Klima, E., & Bellugi, U. (1990). Once more with feeling: Affect and language in atypical populations. *Development and Psychopathology*, 2, 367–391. doi:10.1017/S0954579400005782

Rogers, S. J., & Williams, J. H. G. (2006). *Imitation and the social mind: Autism and typical development*. New York, NY: Guilford Press.

Rosenbek, J. C., Crucian, G. P., Leon, S., Heiber, B., Rodriguez, A. M., Holiway, B., Ketterson, T. U., Ciampitti, M., Singletary, F., Heilman, K. M., & Gonzalez Riothi, L. J. (2004). Novel treatment for expressive aprosodia: A phase 1 investigation of cognitive and imitative interventions. *Journal of the International Neuropsychological Society*, 10, 786–793. doi:10.1017/S135561770410502X

Rothstein, J. A. (2013). Prosody treatment program. Retrieved from http://linguisystems.com/index/home.

Rubin-Spitz, J., & McGarr, N. (1986). Perception of stress contrasts by the hearing impaired. *Journal of the Acoustic Society of America*, 79(1), 10S. doi:10.1121/1.2023065

Samuelsson, C. (2011). Prosody intervention: A single subject study of a Swedish boy with prosodic problems. *Child Language Teaching and Therapy*, 27(1), 56–67. doi:10.1177/0265659010372185

Sharda, M., Subhadra, T. P., Sahay, S., Nagaraja, C., Singh, L., Mishra, R., Sen, A., Singhal N., Erickson, D., & Singh, N. C. (2010). Sounds of melody: Pitch patterns of speech in autism. *Neuroscience Letters*, 478, 42–45. doi:10.1016/j.neulet.2010.04.066

Shriberg, L. D. (1975). *Preliminaries to a social learning theory view of deviant child phonology*. Paper presented at the annual convention of the American Speech and Hearing Association, Washington, USA.

Shriberg, L. D., Kwiatkowski, J., & Rasmussen, C. (1990). *Prosody-Voice Screening Profile (PVSP)*. Tucson, AZ: Communication Skill Builders.

Shriberg, L. D. (2010). Childhood speech sound disorders: From postbehaviorism to the postgenomic era. In R. Paul, & P. Flipsen (Eds.), *Speech sound disorders in children* (pp. 1–33). San Diego, CA: Plural Publishing.

Shriberg, L. D., Paul, R., Black, L. M., & van Santen, J. P. H. (2011). The hypothesis of apraxia of speech in children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 41, 405–426. doi:10.1007/s10803-010-1117-5

Shriberg, L. D., & Widder C. J. (1990). Speech and prosody characteristics of adults with mental retardation. *Journal of Speech and Hearing Research*, 33(4), 627–653. doi:10.1044/jshr.3304.627

Snow, D. (2001). Imitation of intonation contours by children with normal and disordered language development. *Clinical Linguistic and Phonetics*, 15(7), 567–584. doi:10.1080/02699200110078168
Snow, D. (1998). Prosodic markers of syntactic boundaries in the speech of 4-year-old children with normal and disordered language development. *Journal of Speech, Language and Hearing Research*, 41, 1158–1170. doi:10.1044/jslhr.4105.1158
Sparks, R. V., & Holland, A. L. (1976). Melodic Intonation Therapy. *Journal of Speech and Hearing Disorders*, 41(3), 287–297. doi:10.1044/jshd.4103.287
Stark, R. E., & Tallal, P. (1981). Selection of children with specific language deficits. *Journal of Speech and Hearing Disorders*, 46, 114–122. doi:10.1044/jshd.4602.114
Stathopoulos, E., Duchan, J., Sonnenmeier, R., & Bruce, N. (1986). Intonation and pausing in deaf speech. *Folia Phoniatrica*, 38, 1–12. doi:10.1159/000265814
Stojanovik, V., Setter, J. & van Ewijk, L. (2007). Intonation abilities of children with Williams syndrome: A preliminary investigation. *Journal of Speech Language and Hearing Research*, 50(6), 1606–1617. doi:10.1044/1092-4388(2007/108)
Stojanovik, V. (2010). Prosodic deficits in children with Down syndrome. *Journal of Neurolinguistics*, 24(2), 145–155. doi:10.1016/j.jneuroling.2010.01.004
Strand, E. A., Stoeckel, R., & Baas, B., (2006). Treatment of severe childhood apraxia of speech: A treatment efficacy study. *Journal of Medical Speech-Language Pathology*, 14(4), 297–307.
Tager-Flusberg, H., Paul, R., & Lord, C. (2005). Language and communication in autism. In F. Volkmar, R. Paul, A. Klin, & D. J. Cohen (Eds.), *Handbook of autism and pervasive developmental disorders* (pp. 650–681). Hoboken, NJ: John Wiley & Sons.
Thomas-Stonell, N., McClean, M., & Dolman L. (1991). *IBM SpeechViewer*.
Van der Meulen, S., Janssen, P., & den Os, E. (1997). Prosodic abilities in children with specific language impairment. *Journal of Communication Disorders*, 30(3), 155–170. doi:10.1016/S0021-9924(96)00059-7
Van Santen, J. P. H., Prud'hommeaux, E. T., Black, L. M. (2009). Automated assessment of prosody production. *Speech Communication*, 51(11), 1082–1097. doi:10.1016/j.specom.2009.04.007
Van Santen, J. P. H., Prud'hommeaux, E. T., Black, L. M., & Mitchell, M. (2010). Computational prosodic markers for autism. *Autism*, 14(3), 215–236. doi:10.1177/1362361310363281
Vohr, B. (2003). Overview: infants and children with hearing loss–part I. *Mental Retardation Developmental Disabilities Research Reviews*, 9, 62–64. doi:10.1002/mrdd.10070
Wang, A., Dapretto, M., Hariri, A., Sigman, M., & Brookheimer, S. (2001). Processing affective and linguistic prosody in autism: An fMRI study. *Neuroimage*, 13, S621–S621. doi:10.1016/S1053-8119(01)91964-0
Weber-Fox, C., Hampton Wray, A., & Arnold, H. (2013). Early childhood stuttering and electrophysiological indices of language processing. *Journal of Fluency Disorders*, 38(2), 206–221. doi:10.1016/j.jfludis.2013.01.001
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. doi:10.1044/1092-4388(2003/001
Whitehill, T., Patel, R., & Lai, J. (2008). The use of prosody by children with severe dysarthria: A Cantonese extension study. *Journal of Medical Speech Language Pathology*, 16, 293–301.

Index

Prosodic development is increasingly recognized as a fundamental stepping stone in first language acquisition. Prosodic sensitivity starts developing very early, with newborns becoming attuned to the prosodic properties of the ambient language, and it continues to develop during childhood until early adolescence. In the last decades, a flourishing literature has reported on the varied set of prosodic skills that children acquire and how they interact with other linguistic and cognitive skills. This book compiles a set of seventeen short review chapters from distinguished experts that have contributed significantly to our knowledge about how prosody develops in first language acquisition. The ultimate aim of the book is to offer a complete state of the art on prosodic development that allows the reader to grasp the literature from an interdisciplinary and critical perspective. This volume will be of interest to scholars and students of psychology, linguistics, cognitive science, speech therapy, and education.

ISBN 978 90 272 0059 4

9 789027 200594

JOHN BENJAMINS PUBLISHING COMPANY