CHAPTER 39

# DEVELOPMENT OF PHRASE-LEVEL PROSODY FROM INFANCY TO LATE CHILDHOOD

AOJU CHEN, NÚRIA ESTEVE-GIBERT, PILAR PRIETO, AND MELISSA A. REDFORD

## 39.1 Introduction

It was long thought that children master phrase-level prosody (hereafter prosody)—that is, intonational contours, phrasing, and rhythm—before other aspects of language. This belief stemmed from the observation that infants systematically vary pitch, duration, and intensity in their vocalizations according to the interactional context, thus appropriately responding to prosodic variation in adults' speech. Yet, recent theoretical advances in the field have encouraged more detailed study of children's prosodic abilities, leading researchers to question the view of early mastery. Researchers now recognize that mastery of a native prosodic system requires the ability to produce formal properties of a prosodic system, vary these appropriately in response to communicative context, perceive their meanings, and process prosodic information in language comprehension (Kehoe 2013; Prieto and Esteve-Gilbert 2018). Research over the past two decades has shown that prosodic acquisition starts with perceptual attunement to native tonal patterns and with cues to prosodic boundaries in infant vocalizations, but formal properties are not acquired until age 2 years and phonetic implementation of some prosodic features is not mature until middle childhood. The development of the full communicative functions of prosody takes even longer. The goal of this chapter is to present the state-of-the-art research on phrase-level prosodic development from infancy to late childhood (§39.2, §39.3, and §39.4). To this end, we describe developmental trajectories for the formal and functional properties of prosody, identify the factors that may explain why prosodic development is a gradual process across languages, and examine why cross-linguistic differences nevertheless arise early. We also suggest key issues for future research (§39.5).

## 39.2 Prosody in infancy

Infants are sensitive to prosodic variation within and across languages from birth. This sensitivity undergirds early perceptual attunement to ambient language patterns. Infant vocalizations incorporate prosody from a very early age. Whereas the earliest prosodic adaptations to social context may reflect links between acoustic vocal properties and emotional states, it is clear that infants also begin to incorporate language-specific patterns into their vocal repertoires before they begin to speak.

### 39.2.1 Infants' perception of tonal patterns and prosodic phrasing

Infants are acutely sensitive to prosody from birth (see chapter 40). They can distinguish between rhythmically different languages shortly after birth and prefer to listen to a language that is rhythmically similar to their own at 3.5 to 5 months of age (Nazzi and Ramus 2003; Molnar et al. 2013). French newborns detect changes in pitch movement (rise vs. fall) in Japanese words (Nazzi et al. 1998b); English-listening 1- to 4-month-old infants can discriminate a pair of English words whose final vowels differ in pitch, duration, or intensity (Bull et al. 1984, 1985; Eilers et al. 1984). This sensitivity no doubt provides the foundation for the perception of tonal patterns (i.e. phonologically contrastive pitch patterns in a language).

Research on early perception of tonal patterns has focused on infants' perception of non-native tonal patterns from a tone language (Tsao et al. 2004; Mattock and Burnham 2006; Yeung et al. 2013; Liu and Kager 2014; Singh and Fu 2016; see also chapter 38). The main finding is that infants' tonal perception narrows in the first year of life, a phenomenon known as 'phonological attunement' or 'perceptual reorganization'. Whereas all infants can discriminate most native and non-native patterns from birth to between 4 and 6 months of age, sensitivity to native tonal patterns is retained or enhanced after 6 months of age but sensitivity to non-native tonal patterns declines, albeit only in infants whose ambient language does not have lexical tone. Thus, perceptual attunement to tonal patterns is similar to that reported for vowels and consonants, though it occurs at a slightly earlier age (see Kuhl 2004; Saffran et al. 2006; Yeung et al. 2013).

Limited research on infants' perception of native tonal categories in intonation languages shows that early tonal perception may be influenced by how consistently a tonal pattern occurs in a certain communication context and how salient acoustic differences between tonal patterns are (Soderstrom et al. 2011; Frota et al. 2014; Butler et al. 2016; see Frota and Butler 2018 for a review). For example, European Portuguese-learning infants can discriminate between nuclear contours with different boundary tones, such as fall and fall-rise, as early as 5 months of age (Frota et al. 2014), but English-learning infants aged 4 to 24 months do not easily discriminate between rising nuclear contours and non-rising nuclear contours (Soderstrom et al. 2011). Though not directly comparable due to methodological differences (e.g. one-word utterances with different nuclear contours presented in different blocks of stimuli in the Portuguese study vs. multiword utterances with different nuclear contours mixed in the same blocks of stimuli in the English study), the different findings provide initial evidence for infants' sensitivity to the coupling between tonal patterns and communication contexts. In particular, falling nuclear contours are frequently used when making statements and fall-rise nuclear contours are frequently used when asking questions in European Portuguese, whereas rising and non-rising patterns are used in tandem in both situations in English (Soderstrom et al. 2011; Frota et al. 2014). More recently, Butler et al. (2016) have shown that European Portuguese-learning infants do not show sensitivity to contrasts in peak alignment (i.e. the difference between a pitch peak at the end of a stressed syllable vs. on the following syllable) until 8 to 12 months of age, even though these alignment contrasts consistently signal different focus types. Together, the studies on European Portuguese-learning infants suggest that infants acquire tonal contrasts with more salient acoustic differences (i.e. direction of final pitch movement) earlier than tonal contrasts with less salient acoustic differences (e.g. peak alignment) if there is little difference in how consistently a tonal pattern occurs in a certain communication context.

Research on early perception of prosodic phrasing suggests perceptual narrowing around boundary cues. Major prosodic boundaries, such as intonational phrase boundaries, are demarcated by several acoustic cues, including pre-boundary lengthening, pausing, and pitch changes (Ferreira 2007; Mo 2010; Wagner and Watson 2010; Holzgrefe-Lang et al. 2016; Zhang 2012). Languages differ in the distribution of these cues and so listeners weight them differently when identifying boundaries (Peters et al. 2005; Holzgrefe-Lang et al. 2016). For example, English- and Mandarin-speaking listeners treat pitch change as a more reliable cue to boundary perception than pre-boundary lengthening and pause, but Dutch-speaking listeners appear to treat pause as a more reliable cue to boundary perception than pitch changes and boundary tones (Sanderman and Collier 1997; Swerts 1997). As with tonal patterns, infants become attuned to language-specific boundary marking over developmental time. Between 4 and 6 months of age, infants exposed to different languages can discriminate well-formed and ill-formed prosodic units (e.g. 'rabbits eat leafy vegetables' spoken as one intonational phrase or as part of two different intonational phrases with a boundary between 'eat' and 'leafy') only if all boundary cues are present (Seidl and Cristià 2008; Wellmann et al. 2012); between 6 and 8 months of age, infants begin to rely only on those cues that adults find most critical for boundary marking (Seidl 2007; Johnson and Seidl 2008; Wellmann et al. 2012). To take another specific example, English-learning 6-month-old infants attend more to pitch changes than to the other cues at intonational phrase boundaries, but Dutch-learning 6-month-old infants attend more to pausing than to other cues at intonational phrase boundaries (Johnson and Seidl 2008).

## 39.2.2 Prosody in pre-lexical vocalizations

Infants modulate prosodic features to express emotional affect well before they produce identifiable consonant-vowel sequences (Oller et al. 2013): as early as 3 months of age, infants' vocal productions differ as a function of context. For example, the vocalizations produced during positive interactions (i.e. playing, reunion, feeding) differ from those produced when the infant is under stress (i.e. in pain, isolation, or hungry) (Lindová et al. 2015).

Infants express negative affect by vocalizing for longer and across a wider pitch span than during positive interactions (Scheiner et al. 2002).

Around 7 to 9 months of age, infants' vocalizations begin to sound like articulated speech. Consonant-like sounds alternate with vowel-like sounds to form long strings of syllable-like utterances. This is the babbling stage of infant vocal development. This stage overlaps with first-word acquisition in that infants continue to produce long strings of babble intermixed with words well into the second year of life (Vihman 2014). Babbling has a clear prosody that is influenced by the ambient language (e.g. Whalen et al. 1991; Engstrand et al. 2003; DePaolis et al. 2008). For example, DePaolis and colleagues (2008) showed that duration, intensity, and pitch patterns differed in babbling produced by 10- to 18-month-old American English-, Finnish-, French-, and Welsh-learning infants in directions that were clearly consistent with the ambient language prosody.

At around 9 months of age, some infants' vocalizations appear to shift from play-like to intentionally communicative (Vihman 2014; Esteve-Gibert and Prieto 2018). At around 11 months of age, infants clearly use duration, pitch range, and the direction of the pitch movement to specify their specific pragmatic intent. For example, Catalan- and English-learning infants produce longer vocalizations with larger pitch span when requesting or expressing discontent than when responding to a caregiver or producing a statement (Papaeliou et al. 2002; Papaeliou and Trevarthen 2006; Esteve-Gibert and Prieto 2013). Italian-learning 12- to 18-month-olds tend to produce falling contours when making declarative pointing gestures and rising contours when making requestive pointing gestures (Aureli et al. 2017). Moreover, cross-linguistic differences have been observed in the combination of pitch and gestures, possibly reflecting the intonation patterns of the ambient language. For example, in contrast to Italian-learning infants, Dutch-learning 14-month-olds most frequently produce level contours accompanying requestive pointing, and rising contours in declarative pointing (Grünloh and Liszkowski 2015).

In sum, prosody in infancy is characterized by infants' sensitivity to variations in prosodic parameters at birth, the development from language-independent perception to language-specific perception, an evolution from affectively linked vocalizations to more clearly pragmatically driven ones between 3 and 12 months of age, and the approximation of language-specific correlates of prosody during babbling.

## 39.3 Prosodic production in childhood

Developmental studies on prosodic production of toddlers and preschoolers (hereafter children, compared to infants in the preceding sections) are still fragmentary. Researchers have typically investigated a limited set of prosodic phenomena in a small number of children in a narrow age range for a handful of languages. Despite these limitations, a review of the literature suggests an interesting conclusion: children acquire language-specific intonation before they acquire language-specific rhythm. In this section, we briefly review the limited research on prosodic production in childhood that supports this conclusion and then consider what it implies for language development.

## 39.3.1 The acquisition of intonational contours

Contour-based approaches to child prosody have assessed the acquisition of global intonation and prosodic patterns found in languages like English, French, and Japanese (e.g. Hallé et al. 1991; Vihman et al. 1998; Snow 2006). These studies have revealed cross-linguistic differences in the global contours that children produce during babbling and the production of early words (Hallé et al. 1991; Vihman et al. 1998). For example, Hallé and colleagues (1991) investigated the fundamental frequency (f0) and vowel duration properties of bisyllabic babbling sequences and single-word productions of four French- and four Japanese-learning 18-month-old toddlers. The results were that global intonation contours and lexical tonal patterns differed as functions of the ambient language. Snow and Balog (2002) cite this study along with several others (e.g. Crystal 1986; Snow 1995) to support their argument that intonational development reflects the acquisition of language-specific knowledge. The alternative view is that intonational development is a by-product of 'natural tendencies' due to physiological factors (e.g. breathing) or universal socio-emotional ones. This view, common in earlier work on children's intonation production (e.g. Lieberman 1966; Dore 1975; D'Odorico 1984), wrongly predicts a universal pattern of intonational development.

More recent studies on children's speech prosody share Snow and Balog's (2002) focus on knowledge acquisition. These studies have investigated early intonation from the perspective of the autosegmental-metrical (AM) theory, which assumes a prosodic grammar (see chapter 6). The AM perspective encourages a detailed description of children's phonological categories, including mastery over language-specific pitch accents and boundary tones. At this point, children's intonational grammars have been described for Spanish and Catalan (Prieto et al. 2012a; Thorson et al. 2014), Portuguese (Frota et al. 2016), Dutch (Chen and Fikkert 2007; Chen 2011b), and English (Astruc et al. 2013). These analyses suggest that children acquire a good portion of the intonational system between 14 and 19 months, coinciding with the presence of a small, 25-word vocabulary. By age 2 years, children use prenuclear accents appropriately and their speech is characterized by adult-like pitch accent distributions. Adult-like pitch alignment and pitch scaling have also been observed for a few languages (e.g. Spanish for alignment, Portuguese for scaling) (see Frota and Butler 2018 for a review).

Young children also produce cues to major prosodic boundaries at about the same time they correctly produce prenuclear pitch accents. Language-specific patterns of final lengthening have been observed in infant babbling (e.g. Hallé et al. 1991), but it is not fully controlled until age 2 years. For example, Snow (1994) investigated both final-lengthening and falling intonation in a longitudinal study of English-learning 16- to 25-month-olds' speech and observed that control over final falls emerges earlier than control over final lengthening, which is only used systematically as boundary markers at the onset of combinatorial speech.

## 39.3.2 The acquisition of speech rhythm

In contrast to the early acquisition of intonational categories and boundary marking, children do not produce fully adult-like rhythm patterns as measured in terms of interval-based metrics until the age of 4 or 5 years for the 'syllable-timed' Romance languages (Bunta and Ingram 2007; Payne et al. 2012), and not until after the age of 5 years for the 'stress-timed'

Germanic languages (Grabe et al. 1999; Sirsa and Redford 2011; Payne et al. 2012; Polyanskaya and Ordin 2015). Part of the apparent delay is due to the fact that language-specific rhythm patterns are very much a running speech phenomenon. Whereas appropriate intonation and boundary marking can be observed at the one- and two-word stages, well-structured rhythm patterns only arise with longer phrases that are chunked into language-appropriate supralexical units. Of course, children typically produce well-formed sentences as single utterances by about age 3 years. So, why is speech rhythm still immature at age 4 in Romance languages and until after age 5 in Germanic languages?

Payne and colleagues (2012) suggest that the delayed acquisition of language-specific rhythm in Germanic languages is due to an interaction between the acquisition of phonological abilities and the acquisition of phonetic abilities; more specifically, it is due to an interaction between the acquisition of complex syllable structure and the development of motor skills necessary for the temporally invariant production of segmental targets. This explanation is plausible in that it aligns well with the view that language-specific rhythm patterns emerge from language-specific syllable structures (e.g. Dauer 1983) and with the observation that speech motor skills develop slowly (e.g. Smith and Zelaznik 2004). But, in so far as motor skill development continues until middle adolescence, the explanation may overstate the influence of phonological acquisition on speech rhythm development. For example, English-speaking 5-year-olds who have mastered English syllable structure do in fact produce adult-like temporal patterns associated with lexical stress even though their overall rhythm patterns remain immature (Sirsa and Redford 2011). This observation suggests that the protracted development of English rhythm may also reflect delays in children's ability to phonetically implement supralexical structures. For example, Redford (2018) found that 5-year-old children produce longer and louder determiner vowels relative to the adjacent nouns than adults, even though measures of anticipatory coarticulation suggest that they chunk the determiners with the following noun, just as adults do. Thus, the difference between English-speaking children's and adults' speech seems to arise from inadequately reduced grammatical words and not from differences in how these are chunked with adjacent content words to create larger rhythmic units. This interpretation of the results conforms to the more general suggestion that by the time children produce multi-word utterances, their speech representations are no longer influenced by motor skill development, even though this development continues into adolescence (Redford and Oh 2017).

To summarize, current work on children's speech supports the hypothesis that the phonological aspects of prosody are acquired early. The protracted acquisition of speech rhythm suggests that immature motor skills may nonetheless impede children's ability to implement in an adult-like manner the representations they have acquired.

## 39.4 Communicative uses of prosody in childhood: production and comprehension

Research on children's communicative uses of prosody has centred on the interface between prosody and information structure and the expression of emotions and epistemic meanings typical in their interactions with adults. A consistent pattern that has emerged from various studies is that the mastery of adult-like competence in this area is gradual, despite the early use of prosody for interactional purposes and the early mastery of the phonological aspects of prosody. In this section, we review the key findings and explanations for why mastery over communicative uses of prosody takes so long.

## 39.4.1 Acquisition of prosody and information structure

In many languages, speakers vary prosody in response to changes in information structure (see chapter 31). Listeners rely on these changes to make sense of incoming information (Cutler et al. 1997; Dahan 2015). The specific form–function relationship between prosody and information-structural categories is often language-specific. Developmental research on the interface between prosody and information structure is mostly concerned with how children encode focus prosodically in production, how they react to the prosody-to-focus mapping, and how they use prosodic information to anticipate upcoming referents in online comprehension. Although children appear to vary prosody to distinguish new and given information to their interlocutor at the two-word stage (Chen 2011a), adult-like production and comprehension of the prosody-to-information structure mapping develops very gradually: it is not until the age of 10 or 11 years that children acquire adult-like competence in this domain (Chen 2011a, 2018; Ito 2018). Why does this mapping take so long to acquire?

Chen (2018) proposes that differences in prosodic systems and in how to encode information structure result in differences in both the rate and the route of acquisition in children acquiring different native languages. Her proposal is based on a review of recent studies on children's prosodic focus marking in typologically different languages, including Mandarin, Korean, Swedish, Finnish, English, German, and Dutch (Hornby and Hass 1970; Wonnacott and Watson 2008; Sauermann et al. 2011; Arnhold et al. 2016; Romøren 2016; Yang 2017; Chen and Höhle 2018). More specifically, children acquire the use of phonetic means (i.e. phonetic implementation of phonological categories such as lexical tones in Mandarin, lexical pitch accents in Swedish, and pitch accents in English) to distinguish narrow focus from non-focus and to differentiate different focus types (i.e. broad focus, narrow focus, narrow contrastive focus) at an earlier age in a language (e.g. Mandarin) that exclusively relies on phonetic means for focus marking than in a language that uses both phonological and phonetic means for focus marking (e.g. English, Dutch). Furthermore, children acquire phonological encoding of narrow focus at an earlier age in languages with a more transparent form–function mapping between the phonological means and focus conditions (e.g. Swedish and Korean vs. Dutch). The effect of transparency is also present in the phonological marking of focus in different sentence positions within the same language. For example, Dutch-speaking children acquire phonological focus marking earlier in sentence-initial and -final positions than in sentence-medial position, where the form–function mapping is blurrier. Moreover, children acquire the use of 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 of prosody and word order for focus marking has an effect on children's use of phonetic means in distinguishing focus types in different word orders. For example, 4- to 5-year-olds acquiring languages that use word order in conjunction with prosody to mark focus use prosody more extensively and are less restricted by the word order of the sentences (e.g. Finnish) than children acquiring languages where prosody plays a primary role in focus marking (e.g. German and Dutch).

With respect to the comprehension of the prosody–information structure interface, a much studied information-structural condition is focus. Adults take the focus-to-prosody mapping into account in online language comprehension such that an appropriate focus-to-prosody mapping leads to faster comprehension than an inappropriate focus-to-prosody mapping (Birch and Clifton 1995; Cutler et al. 1997). It is widely believed that, at the age of 4 or 5, children are not able to interpret or efficiently use the focus-to-prosody mapping in comprehension, but they use prosody to realise focus in production (Cruttenden 1985; Cutler and Swinney 1987; Hendriks 2005; Müller et al. 2006). Chen (2010b) reviewed the comprehension studies cited as evidence for children's failure in comprehension and found that none of these studies has directly examined children's comprehension of the focus-to-prosody mapping. Besides, the test materials used in the earlier comprehension studies were usually syntactically more complex and semantically more demanding than the materials used in related production studies. Controlling for syntactic complexity and task complexity, Chen (2010a) and Szendröi et al. (2018) found that children can process the focus-to-prosody mapping at the age of 4 or 5 years. There are, however, substantial individual differences in the comprehension of focus-to-prosody mapping in children under the age of 11 years (Chen 2014; Chen and van den Bergh, 2020) and they take more time than adults to reach a decision regardless of focus conditions in comprehension (Chen 2010a).

The other frequently studied information-structural condition is contrast. With the increasing accessibility of eye-tracking techniques, researchers have been able to use the visual-world paradigm (Trueswell and Tanenhaus 2005) to study how children aged 4 to 11 years process prosodic manipulations in a short stretch of material (e.g. the adjective of an adjective-and-noun phrase, the first syllable of a word) to predict the upcoming referent in different languages (Arnold 2008; Sekerina and Trueswell 2012; Ito et al. 2012, 2014). For example, Sekerina and Trueswell (2012) reported a facilitative effect of prominence in the adjective on the detection of the correct target referent in colour adjective-and-noun phrases in 6-year-old Russian-speaking children. Ito and colleagues (2012, 2014) found both facilitative and misleading effects of prominence in the adjective in similar tasks in English- and Japanese-speaking 6- to 11-year-olds. However, children take more time to respond to the prosodic information in the stimuli than adults and are not as fast as adults even at the age of 11. They also need more time to recover from misguided interpretations than adults (Ito 2018). According to Ito (2018), the slow acquisition of contrastive prosody comprehension may be related to underdeveloped executive function, such as attention allocation and inhibition. To respond quickly to the prominence in the speech, children need to switch their attention quickly from the previous referent or the referent that they have considered initially to something new. This is hard to achieve when their executive functions are still developing. This proposal highlights the relation between prosodic development and cognitive development.

## 39.4.2 Acquisition of prosody and sociopragmatic meanings

The meaning of an utterance goes beyond its information-structural interpretation. Speakers also convey and infer via prosody and body language sociopragmatic meanings in communication, such as emotions, irony, politeness, and epistemic stances (e.g. uncertainty, disbelief) (Barth-Weingarten et al. 2009; Prieto 2015; Brown and Prieto 2017; and see also chapter 32). In recent years, researchers have begun to study how children express and understand sociopragmatic meanings via speech prosody and body language in a few languages (see Armstrong and Hübscher 2018 for a review). Complex cognitive abilities are required to infer the others' affective and epistemic stances. It is not until the age of 4 to 5 years that children fully develop the so-called theory of mind—that is, the ability to understand others' emotions, beliefs, and desires—which enables them to use prosody as a tool to infer these complex meanings (e.g. Kovács et al. 2010; Apperly 2012; Ruffman 2014). Research on children's use of prosody in expressing and understanding sociopragmatic meanings has focused on 3-year-olds and older children.

Studies on Catalan- and Dutch-speaking children show that children use prosody to encode epistemic stances such as uncertainty between the ages of 3 and 5 years, but their exact use of prosody is not adult-like even at the age of 7 or 8. For example, Catalan-speaking children use prosodic cues to express uncertainty at age 3 (before they actually use lexical cues for that purpose) (Hübscher et al. 2019). Dutch-speaking 7- to 8-year-olds use mostly delays and high pitch, while adults also use filled pauses, eyebrow movements, and funny faces (Krahmer and Swerts 2005).

In speech comprehension, children appear to rely more on prosodic cues than lexical cues to detect others' epistemic stance at the ages of 3 to 5 (Hübscher et al. 2017). For example, 3-year-old English-speaking children can correctly identify which speaker is requesting an action in a polite way on the basis of the intonation contour (i.e. rising contours for polite requests and falling contours for impolite requests) (Hübscher et al. 2016, 2018). Children's interpretation becomes more accurate when they can also access the interlocutor's facial expression (Armstrong et al. 2014; Hübscher et al. 2017).

The earlier reliance on cues in prosody and body language to interpret epistemic stances and politeness has led to the suggestion that prosody, along with gestural and bodily marking, can serve a bootstrapping function, helping children to express complex sociopragmatic meanings at young ages (Hübscher and Prieto 2019). However, other work has found partially conflicting evidence showing that the ability to interpret prosodic cues to other indirect pragmatic inferences emerges at a later age. For example, it is not until the age of 5 years that children explicitly match sad-sounding prosody with unfamiliar broken objects and happy-sounding prosody with unfamiliar nicely decorated objects (Berman et al. 2010, 2013). When prosodic cues conflict with contextual and/or lexical cues, as in irony, children under the age of 10 tend to rely more on contextual and lexical cues than on prosodic cues to reach an ironic interpretation (Gil et al. 2014). However, they give prosodic cues more weight than contextual cues when prosodic cues are combined with a reinforcing facial expression (e.g. González-Fuente 2017).

In summary, the above-reviewed work shows that children appear to learn multiple functions of prosody simultaneously between age 3 and age 11. The mastery of adult-like competence in the use of prosody in communication is a gradual process and it requires a certain level of competence in other cognitive domains; the form–function mapping is not necessarily transparent in everyday speech. Cross-linguistic differences in the rate and route of acquisition can arise due to differences in prosodic systems and specific uses of prosody.

## 39.5 Future research

This brief review of past research on children's use of prosody in production, perception, and comprehension from infancy to late childhood describes fundamental knowledge about prosodic acquisition, which gives rise to interesting theoretical insights. It has become apparent from this review that development of phrase-level prosody takes years despite early sensitivity to acoustic variations in prosodic features and early language-specific patterns in infants' pre-lexical vocalizations.

However, existing work is almost solely concerned with establishing what children can do at which age in specific languages. The question as to which learning mechanisms drive the developmental changes await attention in future research. Relatedly, research on the acquisition of the use of prosody for information-structural purposes and illocutionary force has been separated from research on the acquisition of the use of prosody to express and interpret sociopragmatic meanings, as also noted by Esteve-Gibert and Prieto (2018) and Ito (2018). Consequently, we are still far from having a comprehensive picture of children's use of prosody in communication. A holistic approach whereby different functions of prosody are studied in interaction will probably be an interesting and rewarding challenge for future research. Furthermore, much development appears to have taken place before the age of 4 years in most of the communicative uses of prosody reviewed above. In general, more research is needed on toddlers' prosodic development, for which new suitable research paradigms are called for (Chen 2018).

Another avenue for further research departs from the observation that children appear to master the intonational aspects of speech prosody earlier than the rhythmic aspects. Initial study of the alignment of phrasal accent and lexical stress confirms that the intonational and rhythmic systems are in fact still segregated in English-speaking school-aged children's speech (Shport and Redford 2014). Future work should investigate how the systems come to be integrated over time in production.

Finally, children develop a range of abilities in linguistic and non-linguistic domains in the same period when their prosodic abilities develop. Research is needed to disentangle how prosodic development interacts with development of other aspects of language (e.g. vocabulary, syntax) and in other cognitive domains (e.g. perspective-taking abilities, empathy).