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