CHAPTER 13

Gradual development of focus prosody *and* affect prosody comprehension: a proposal for a holistic approach

Running title: Gradual development of prosody comprehension

Kiwako Ito

Ohio State University

Excellence in communication skills requires an ability to appropriately represent the discourse structure including focus, as well as good comprehension of speaker affect. Both focus and affect are communicated in large part through prosody, so comprehension and production of the accompanying prosody is essential. However, past studies on focus prosody have been both theoretically and methodologically separated from the research on affect prosody. (In this chapter, I use the term 'focus prosody' to refer to prosodic phenomena that are either produced or perceived as the cue to a specific part of speech that conveys the focal content of a message. This includes 'narrow focus', which is defined in terms of the informational scope (e.g., answers to Wh-questions), and 'contrastive focus', which is a subtype of narrow focus that evokes interpretational alternatives.) This chapter argues that the suggested difference in the developmental trajectory (i.e., focus prosody develops slower as compared to affect prosody) may be an artifact of the perspective divergence, and points out that the mastery of prosodic skills in both these domains must be necessarily *gradual* – though they may not develop hand-in-hand. A holistic approach that considers the interaction between affect prosody and focus prosody is proposed as a future direction of the research on prosodic development within and across individuals.

Introduction

Prosodic prominence provides dynamicity in speech acts at multiple levels of communication: it may accompany facial expressions and gestures for displaying affect during face-to-face interaction, convey the speaker's mood, level of excitement or formality in the presence or absence of visual cues, reflect the speaker's attitude toward the topic of a conversation or the listener, and signal the relative importance of particular parts of speech, indicating to the listener what is worth paying attention to and what is a simple reminder or a confirmation. Not surprisingly, understanding prosodic development in children has been a challenge for linguists, psychologists, speech therapists, neuroscientists and psychiatrists due to its intertwined relationship to individuals' social cognition and pragmatic knowledge. Experts in various scientific disciplines have studied prosody to describe its linguistic structure (Beckman, 1996; Beckman & Pierrehumbert, 1986; Jun, 2005; Ladd, 2008), to identify neuro-anatomical specificity of brain function (Belyk & Brown, 2013; Grossmann, Striano, & Friederici, 2005; Ross, 1981; Wildgruber, Ackermann, Kreifelts, & Ethofer, 2006), or to achieve better machine recognition of speaker's emotions (Fernandez & Picard, 2011; Oudeyer, 2003; Schuller et al., 2010; Schuller, Steidl, & Batliner, 2009). Researchers have also tried to assess the prosodic skills in individuals with communication disorders and identify the similarities and differences in their processing mechanisms as compared to those in their typically-developing peers (Edwards, Pattison, Jackson, & Wales, 2001; Hargrove, 2013; Ito & Martens, 2016; Peppé, 2009, this volume; Peppé & McCann, 2003; Wells, Peppé, & Goulandris, 2004). While efforts in each of these fields have certainly advanced our understanding of prosody in the last few decades, the extreme theoretical and methodological diversity in their approaches may have prevented the construction of a model of prosodic

development that incorporates empirical evidence gathered across this wide range of research fields.

In the past, I have argued that the acquisition of contrast-marking prosody takes time, and have proposed some accounts as to why (Ito, 2014; Ito, Bibyk, Wagner, & Speer, 2014; Ito et al., 2012). This chapter reviews a wider range of related research to advance this view and adds a claim that while comprehension of focus prosody develops gradually through childhood (and possibly through adulthood), so do other abilities that involve perception and processing of prosodic prominence – such as affect recognition. I argue that accurate mapping of prosodic prominence onto message structure takes many years to develop, whether it is for computing the discourse structure (e.g., focus) or for identifying the intention or emotion of the speaker (e.g., affect). This is because human oral communication gains complexity and sophistication through social experience, yet its elements are susceptible to cognitive constraints and informational screening. The intentions and nuances that can be expressed via prosody may receive more or less weight according to the reliability of the cues produced by the speaker (Kurumada, 2013) and the listeners' cognitive capacities such as attention or memory (Fraundorf, Watson, & Benjamin, 2010, 2012). In addition, listeners may tune differently to particular aspects of prosodic cues according to their percept of the discourse context and communicative demands (Kurumada, et al., 2014; for speaker's prosodic adjustment according to informational predictability, see Turnbull, 2016). While processing of prosodic prominence in any discourse context involves both the computation of informational weight and the computation of speaker's emotional status, there has been a clear division between these subfields of prosody research: when listeners' comprehension of focus prosody is studied, their understanding of speaker affect is rarely discussed, whereas when listeners' emotion recognition is investigated, how they represent the informational structure tends to be

neglected. To date, research on the interaction of so-called 'linguistic' prosody and affect prosody is extremely sparse (Pihan, Tabert, Assuras, & Borod, 2008). However, simultaneous consideration of those two aspects is critical for modeling prosodic development, as correct identification of speaker's affect may facilitate the comprehension of informational foci and vice versa.

While the ability to perceive affect in speech seems to start developing from a preverbal stage (Flom & Bahrick, 2007; Grossmann, Striano, & Friederici, 2005; Walker-Andrews & Grolnick, 1983; Walker-Andrews & Lennon, 1991; see Esteve-Gibert & Prieto, this volume, for an updated summary), the ability to use prosodic cues to comprehend the informational structure seems to take a much slower, gradual developmental trajectory (Cruttenden, 1985; Cutler & Swinney, 1987; Grassmann & Tomasello, 2010; Ito, 2014; Ito, Bibyk, Wagner, & Speer, 2014; Ito et al., 2012; Sekerina & Trueswell, 2011; Wells, Peppé, & Goulandris, 2004). This apparent contrast in the developmental time course, however, may have resulted from differences in the theoretical assumptions and research directions across these 'sub-fields'. The first section of the present chapter reviews how scholars have framed their research questions and described their study implications to elicit two distinct traditions: while scientists who study affect prosody focus on *how early children start exhibiting their sensitivity to vocal cues to speaker emotion*, those who study focus prosody attempt to demonstrate *when children start representing the referential statuses of linguistic entities (e.g., morphemes, words and phrases) in an adult-like manner*. There, I argue that a particular assumption about cognitive development (i.e., that understanding affect prosody must precede comprehending prosodic cues to informational structure) may have driven research on affect prosody to focus on infancy and studies on focus prosody to explore later developmental stages.

Following a summary of the theoretical trends, the second section of the chapter discusses the methodological constraints that may have led to the contrast in the research outcome - early affect prosody and late focus prosody. Here, I emphasize that we must carefully evaluate the age-appropriateness of tasks and the meta-linguistic skills required by the tasks before making conclusions about the onset and completion of prosodic development. The third section of the chapter will discuss the cognitive factors that underlie the development of prosodic semantics and processing efficiency. It will also explore the danger in assuming prosodic categories that are composed of fixed sets of acoustic cues that directly map onto semantic categories (whether affect or informational status). In the final section, I will propose a holistic approach that investigates how children's perception of speaker affect influences their discourse structure representation. In this approach, individual differences in the executive function as well as social cognition must be considered together. The chapter will conclude with a set of questions and research topics for future studies.

Theoretical division: fundamental affect function vs. complex information-structuring function of prosody?

Studies on affect prosody development often emphasize how essential the processing of affect is for the general development of social cognition and communication skills. An example comes from the first line of the article by Vaillant-Molina, Bahrick, and Flom (2013): "Perception of emotional expressions is fundamental to social development." The view that the sensitivity to emotional expressions and its early comprehension are universal seems to have collected support from cross-linguistic studies on infant-directed speech (Fernald, 1985,

1993, 2004; Fernald et al., 1989; Fernald & Mazzie, 1991; Katz, Cohn, & Moore, 1996), as well as from modern brain science which provides evidence for spontaneous neuronal responses to affective prosody in both adults and children (Ethofer et al., 2012; Ethofer, Kreifelts, & Wiethoff, 2009; Grandjean et al., 2005; Grossmann, Striano, & Friederici 2005). Research on primates' distinctive responses to affective vocalization (e.g., Owren & Rendall, 1997; Scherer & Kappas, 1988) has also supported the view that the ability to learn to recognize emotion in voice may have an evolutionarily traceable genetic basis. In a review of studies on affect prosody, Grossmann, Striano, and Friederici (2005: 1828) state, "it has been proposed that enhanced sensory responses to emotional facial and vocal stimuli might be a fundamental neural mechanism". This echoes the earlier claim by Walker-Andrews (1997: 437): "although processing only rudimentary capacities to detect, discriminate, and recognize others' emotional expressions, the human infant is born well prepared to rapidly develop these competencies during the first year." Developmental language researchers are generally driven by the question of 'how early children start responding to linguistic stimuli' and are expected to advance methodologies for detecting the behavioral traits earlier than previously reported. It is therefore not surprising that many studies of affective vocal perception test infants younger than 5 months to examine whether they can respond to changes in affect prosody (Flom & Bahrick, 2007; Villant-Molina, Bahrick, & Flom, 2013; Walker, 1982; Walker-Andrews, 1997).

In comparison, the field of research on focus prosody seems to be less bound to this 'how early' question. This may be due to the notion that the understanding of prosody that signals the information structure of words (such as novelty or givenness, narrow focus and contrast) must await the development of abilities to segment words, allocate attention to relevant referents, and represent referential status of linguistic entities. In other words, we

tend to think that understanding which words bear relative importance in a given discourse demands a wider range of cognitive prerequisites than understanding affect. Numerous studies have shown that the use of language-specific rhythmic structure for word segmentation develops in the latter half of the first year (Bull, Eilers, & Oller, 1984, 1985; Eilers, Bull, Oller, & Lewis, 1984; Höhle et al., 2009; Jusczyk, 1997; Jusczyk, Cutler, & Redanz, 1993; Nazzi, Jusczyk, & Johnson, 2000; see also Frota & Butler, this volume, for a more detailed review). This coincides with the phase where infants exhibit the ability to extract transitional probabilities (Marcus, Vijayan, Bandi Rao, & Vishton, 1999; Saffran, Aslin, & Newport, 1996) and gain further sensitivity to their native phonemic categories while losing sensitivity to foreign categories (Kuhl et al., 2006; Werker & Tees, 2002). Interestingly, a shift in the gaze more to the mouth than to the eyes takes place in this narrow time window as well (Tomalski, 2015). Thus, infants who will start producing one-word utterances in a few months are busy forming their native phonemic categories, learning the language-specific rhythmic patterns and phonotactics, and mapping the sound strings to visually identifiable oral gestures. While this seems a lot to deal with for their very young brains, they may be able to handle more. In theory, infants who are learning to detect word boundaries in running speech may be simultaneously developing the ability to compute gross informational weight for parts of speech that bear prosodic prominence. While young infants become capable of tuning to durational, pitch, and spectral cues for segmenting words, they may be able to benefit from exaggerated, emphatic cues to more clearly represent the referential relationships.

Fernald and Mazzie (1991) once hypothesized that focus prosody, which tends to mark word newness, must bear a critical role in lexical learning in infants who cannot rely on a rich semantic network or contextual knowledge. Fernald's research on infant-directed speech (IDS) emphasizes the effect of IDS on word learning via infants' preferences for

motherese (Cooper & Aslin, 1990; Fernald & Kuhl, 1987). Later studies provided direct and indirect evidence for this view. For example, Thiessen, Hill and Saffran (2005) confirmed that statistical learning of word-forming syllables is facilitated by IDS. Ma, Golinkoff, Houston, and Hirsh-Pasek (2011) showed that 21-month-old children learned novel words better with IDS than with adult-directed speech (ADS), a finding which was replicated in 17-month-olds by Graf Estes and Hurley (2013). Importantly, Graf Estes and Hurley (2013) demonstrated that IDS was effective only when target words exhibited prosodic variability across tokens: mere repetitions of identical IDS tokens did not lead to word learning, and performance was no better than for children tested with ADS. Shukla, White, and Aslin (2011) showed that 6-month-old infants gazed longer at an object that was labeled with a disyllabic string (e.g., mu:ra:) when the string was grouped into the same intonational phrase (IP) than when it straddled an IP boundary. In addition, Sakkalou and Gattis (2012) revealed that 18-month-old English-learning toddlers better followed the actions labeled by foreign words (Greek "*Ochi*" and "*Nato*") when they were presented with prosody that expressed a speaker's intention to guide them (i.e., a tune that is similar to what is produced with the English directive utterance "There.") than when they were accompanied by prosody that expressed an accidental event (the tune that accompanies "Whoops"). Thus, studies suggest that children younger than two years are not only sensitive to various prosodic cues (on this topic see Esteve-Gibert & Prieto, this volume; Frota & Butler, this volume), but they are also able to *use* prosody to allocate their attention to novel events and learn object names. What remains unclear is whether they can use prosodic prominence to *represent discourse structure* even at a very rudimentary level (e.g., whether some entities or events are more worth paying attention to than others for the current communicative purpose).

As for the development of the ability to map prosodic prominence to discourse structure, studies show mixed results regarding when it emerges and how it is mastered. The age of participants also has varied widely, and the conclusion as to whether children can or cannot comprehend focus prosody seems largely dependent on the experimental tasks and measures. In one of the early production studies, Hornby and Hass (1970) tested 4-year-old children, who had not been explicitly taught the use of contrastive prosody in school. Because their participants used contrastive stress more often for the contrastive element of scenes they were asked to describe than for other non-contrastive elements, Hornby and Hass concluded that production of contrastive stress was mastered by 4 years of age. Hornby (1971) put 1st, 3rd , and 5h graders (approximately 6, 8, and 10 years old) through a task where participants must provide an explicit correction of the experimenter's scene narration, and found that the use of contrastive stress reduced as children gained a wider range of syntactic means to express focus (such as passive and cleft structures). Later, a study by MacWhinney and Bates (1978) reported that discourse newness, which was elicited by a switched element across otherwise identical scenes (e.g., the recipient in a scene where a person gives something to another person), increased the use of sentential stress in children between ages 3 and 5 years. Their results also showed that English-speaking children used stress a lot more often than Italian and Hungarian-speaking children, who preferred to use other means such as word order to express newness. The steady increase of stress with age was found only in English-speaking children, indicating that the frequency of focus prosody in child language may depend on the language-specific repertoire of focus expressions children develop with age. Recent production studies summarized by Frota and Butler (this volume) and by Chen (this volume) suggest that the developmental trajectory of prosodic production is even more complex than what these early studies depicted decades ago. These studies show that both language-specific

constraints and individual variability impact the use of particular phonetic cues, while leaving little room to dispute the early onset of functional prosodic production in child speech. The detection of early onset, however, should not be interpreted as the evidence of immediate mastery.

As for the comprehension of focus prosody, the findings from Solan (1980), Cutler and Swinney (1987), Cruttenden (1974, 1985), and Wells, Peppé, and Goulandris (2004) all point to difficulties for preschoolers and school-age children in understanding the meaning of sentential stress. These findings therefore provide counter-evidence to the general observation of language development that comprehension precedes production. In fact, this paradox of focus prosody acquisition was explicitly discussed by Cruttenden (1985) and Cutler and Swinney (1987). Cruttenden (1985) acknowledges the studies by Hornby and Hass (1970) and Crystal (1979) that argue for an early mastery of production of focus prosody, yet claims that his experimental results, in which 10-year-old children underperformed adults in comprehension of prosody, "dispel the myth that children master the adult intonation system very early in their linguistic life [which Cruttenden had already claimed in 1974] p.657)". Cruttenden defies Bolinger's (1978) view of 'intonation as an innate gesture-like reflex' as rather too strong and argues that certain intonation meanings that require mature grammatical and contextual knowledge (e.g., 'indignation' meaning for a fall-rise with a word 'might') should appear later in production. Cruttenden also points out that children's underperformance in some comprehension tasks (e.g., distinction between *hotdog* and *hot dog*) does not necessarily reflect their lack of knowledge (of stress assignment), but may instead show their uncertainty in how to weigh various cues available at the moment. Cutler and Swinney (1987) also argue that children younger than age 5 to 6 have yet to develop the skill to map discourse structure onto linguistic structure, and that truly intentional use of focus

prosody must await the development of semantic and pragmatic knowledge: “Only once this [referring to the development of discourse representation] has occurred can the prosodic production system approximate the adult system, in which the underlying physiological basis has become ‘socialized’ (p.163).” Considering Bolinger’s view mentioned above, Cutler and Swinney suggest that the early production of sentential stress or focus prosody is ‘qualitatively different’ from later productions that are based on the computation of discourse-level factors.

In sum, earlier studies that discussed the discrepancy between children’s production and comprehension of focus prosody emphasized that the integration of semantic and pragmatic knowledge with the prosodic structure may not be acquired early, because sufficient development of those components is necessary before they can be linked effectively. Since we cannot address when children acquire adult-like skills to express and interpret prosodic meaning without a fair grasp of adults’ prosody and its function, researchers’ attention may have been drawn more toward a finer-grained description of adult prosody and simultaneously drifted away from the development of focus prosody during early childhood. In contrast, researchers of affect prosody tend to assume that perception of prosodic cues to emotion must develop (and be mastered) early, and thus they have not explored how children gain prosodic means to express and understand emotion in later stages of development. However, both tracing the symptomatic behavior of information structuring in early developmental stages and investigating the processing of affect in later developmental stages are essential for achieving a more accurate description of prosodic development within and across individuals. The methodological challenges are, of course, non-trivial.

Methodological division: passive tasks to test affect detection vs. interactive tasks to test focus comprehension

Due to the focus on early infancy, the majority of traditional investigations of affect prosody in infancy has used preferential looking paradigms (Flom & Bahrick, 2007; Vaillant-Molina, Bahrick, & Flom, 2013; Walker-Andrews, 1998; Walker-Andrews & Grolnick, 1983; Walker-Andrews & Lennon, 1991). There, the experimental effects are inferred from the gaze duration and its proportion, which is assumed to reflect the degree of infants' interest in the stimuli. Since the values of these dependent measures themselves do not reveal the processing mechanisms that lead to the differences across conditions, statistical inference must be necessarily indirect: for example, if tested infants spent 54% of stimuli presentation time attending to the target in condition A and 46% of time in condition B and their difference is statistically significant, researchers may conclude, despite the overall quasi-chance-level performances, that infants preferred condition A to condition B or that they at least discriminated the tested affect categories. Designs of infant studies are typically constrained by the small number of trials, and the interpretation of small effect sizes requires caution due to close-to-chance-level baselines. Research with preschoolers that investigates the mapping between prosodic tunes and affect types also tends to be constrained by a small number of stimuli (see Armstrong & Hübscher, this volume, for a review of studies testing children aged 3 years and older). In many studies, the experimental task often forces one-to-one mapping between prosodic tunes and labels or facial expressions (e.g., Berman, Chambers, & Graham, 2010, 2016). While this seems a feasible strategy, forced-choice tasks entail a problem of methodological adequacy because the prosody-to-meaning mapping is 'context-dependent and defeasable' in nature (Hirschberg, 2002), and the performance may rely on the

participants' understanding of the experimenter's intention (see below for a discussion on a similar problem with focus prosody research). If the task measures the skill for building mutually exclusive links between stimulus sets and response options within an experiment, the findings may or may not directly reflect children's spontaneous interpretations of the prosodic cues in question.

The indirectness of inference also applies to a neurolinguistic approach. For example, a study with the event-related brain potential (ERP) technique by Grossmann, Striano, and Friederici (2005) reports larger positive wave shifts for the temporal electrodes for happy and angry prosody as compared to neutral prosody, and negative wave shifts for angry, as compared to happy and neutral prosody for the frontal-central sites in 7-month-old infants. These differences in the size of brainwave components show that the infants responded to the prosodic changes and indicate that they may attend to a specific category of valence differently. However, the differences in the size or direction of brainwaves do not show whether the infants recognized the emotion, i.e., whether the happy and angry prosody were *interpreted* by the infants as the expression of happiness and anger. Thus, implications of the studies with preverbal infants are often limited to the indication of the sensitivity to prosodic manipulation.

In contrast, methods for testing focus prosody research in older children may be easier to evaluate against study hypotheses, because unlike deducing affect recognition from indirect physiological measures such as gaze duration and ERPs, the semantics of prosody can be more directly observed with interactive tasks. Solan (1980), Cruttenden (1985), Cutler and Swinney (1987), and Wells, Peppé, and Goulandris (2004) all employed tasks that required behavioral responses from children that can be coded as either accurate or inaccurate. Except for Cutler and Swinney who used a word-monitoring task, these studies used visual stimuli or

real world objects that had to be selected or evaluated according to speech input. Solan (1980) used toy animals in a game where the experimenter and a child participant took turns in acting out the pre-recorded narratives (e.g., *The camel hit the lion, and then HE hit the elephant*). While 5-, 6-, and 7-year-old children all performed correctly when they heard the stress on the pronoun (e.g., by grabbing the lion), 5-year-olds performed incorrectly when the unstressed pronoun was expected to lead to a parallel interpretation (while 6- and 7-year-olds grabbed the previous agent camel upon hearing *he*, 5-year-olds grabbed the lion). Since the contrastive stress did not change the behavior in 5-year-olds, Solan (1980: 696) speculates that they are in a phase in which they "assume that reciprocity governs events in the world". While this proposal remains to be empirically attested, I suspect that this outcome may also have to do with a difference in task comprehension between the youngest and older children. It is possible that the youngest children simply thought that the animals were supposed to take turns in this particular game, and thus focused on the action to be repeated and paid little attention to prosody. Once they believed that they had learned the rule of the game, the perseverance tendency in younger children (Trueswell, Sekerina, Hill, & Logrip, 1999) may have made them respond in a consistent manner for the remaining trials. The older children, in contrast, may have better guessed what the experimenter expected them to do in the task: differentiate actions according to different speech input.

While measures of interactive tasks may more directly speak to the research questions than indirect physiological measures, earlier studies on focus prosody (e.g., Cruttenden, 1985; Wells, Peppé, & Goulandris, 2004) were not free of methodological problems. Since I have discussed them elsewhere (Ito, 2014; Ito, Bibyk, Wanger, & Speer, 2014; Ito et al., 2012), I refrain from repeating the details here. In a nutshell, children have shown difficulty selecting the picture that matches the intended interpretation of the spoken stimuli, yet their poor

performance may have resulted from (1) a failure to comprehend out-of-the-blue prosodic prominence that was presented in isolated sentences, and (2) a problem linking the speaker's intention to one of the multiple contrastive relations among the visual prompts, which were susceptible to ranking according to visual salience (for example, a boy holding four oranges makes a more salient contrast with a girl holding four bananas than with a girl holding two oranges, thus the former picture set is more likely to lure children's attention when they hear *John's got FOUR oranges*). Importantly, although successful performances in these interactive tasks may suggest children's correct contrastive interpretation of the prominence, their incorrect responses do not necessarily indicate a lack of contrastive interpretation. It is possible that children noticed the emphasis in the narrative, but decided to weigh visual salience more. Ironically, these interactive tasks for gauging the semantics of prosody in older children are not adequate for detecting their sensitivity to prosody, which may be identified quickly by passive preferential looking paradigms.

While performance in offline interactive tasks often depends on child participants' interpretation of the task, using eye-tracking techniques such as the visual-world paradigm can overcome the problem of individual differences in meta-linguistic skills and improve methodological adequacy for investigating children's responses to the presence or absence of prosodic prominence (Arnold, 2008; Ito, Bibyk, Wagner, & Speer, 2014; Ito, et al., 2012; Sekerina & Trueswell, 2011). The primary advantage of the visual-world paradigm is its capacity to trace participants' spontaneous reactions to speech input before they follow the commands or make selections for a visual search (Trueswell & Tanenhaus, 2005, for a summary). Visual-world studies commonly report that child participants (aged between 4 and 11 years) respond to prosodic manipulations within a short period of time after the critical speech input. For example, Arnold (2008) detected fixations to the anaphoric (or already

mentioned) visual target within 1000 ms from the onset of the unaccented target word in 4- and 5-year-old children. Sekerina and Trueswell (2011) found facilitation of target detection for nouns that immediately followed accented color adjectives in 6-year-old Russian-speaking children. Ito and colleagues found both facilitative and misleading effects of prominence on pre-nominal color adjectives within 400-600 ms after the offset of the adjectives, in Japanese-speaking (Ito et al., 2012) as well as English-speaking (Ito, Bibyk, Wagnerm, & Speer, 2014) 6-year-old children: a prominent adjective in a sequence that repeated the noun (pink cat → GREEN cat) led to faster detection of the target animal set (cat), whereas the prominence in the sequence that switched both the adjective and the noun (purple lion → ORANGE monkey) led to initial fixations on the previously mentioned animal set (lion), which resulted in a slower detection of the target set (monkey). In eye-tracking data, the timing of fixations can reveal participants' sensitivity to prosodic (as opposed to segmental) cues, while the direction of fixations can determine whether prosodic prominence is appropriately interpreted as the cue to contrast. In Ito, Bibyk, Wager, and Speer (2014), the robust immediate effect of prominent accent was confirmed in all age groups (6- and 7-year-olds, 8- and 9-year-olds, 10- and 11-year olds, and adults), but importantly, the fixation timings of child participants approached those of adults *gradually* with age, and even the oldest child group (10- and 11-year-olds) were not as swift as adults. In addition, the oldest group's recoveries from the misguided fixations were clearly delayed as compared to those of adults. For example, while adults' initial incorrect fixations to the lion cell in (purple lion → ORANGE monkey) peaked at the midpoint of the noun 'monkey' and decreased from there on, the incorrect fixations kept increasing throughout the noun in 10- and 11-year-olds.

In sum, eye-tracking techniques with better temporal resolution have advanced our understanding of prosodic processing in children: the data demonstrate immediate responses

to prosodic cues, spontaneous detection of particular visual targets that reflects the interpretation of the cues, and a clear effect of age on processing efficacy. The developmental trajectory that these findings depict, however, is not very different from the claims of three decades ago: comprehension of focus prosody may emerge early, but takes time to develop (Cruttenden, 1985; Cutler & Swinney, 1987). To date, use of interactive eye-tracking paradigms such as Arnold (2008), Sekerina and Trueswell (2012), Ito et al. (2012), and Ito, Bibyk, Wagner, and Speer (2014) to investigate younger children (e.g., 2- and 3- year olds) is rather sparse, because experimental materials are limited by such factors as the child participants' vocabulary size, ability to comprehend instructions, and attention span. Future studies must explore whether using a smaller number of within-subject conditions with between-subject designs can overcome the problems of data loss and low statistical power that typically challenge the studies with toddlers.

Even though the use of online methods has enabled fine-grained examination of spontaneous responses to prosodic cues in children, such studies have yet to overcome a methodological pitfall that is common across experimental paradigms: the use of acted or read speech. In most of the aforementioned studies on affect prosody and focus prosody, participants were presented with a set of carefully handpicked speech stimuli produced by an actor or trained phonetician. This has been the methodological norm, in order to overcome the problem of using live face-to-face interactions that do not control for prosodic consistency within and across experimental sessions (e.g., Cruttenden, 1985). Using pre-recorded speech stimuli facilitates objective assessment of prosodic skills and reduces inter-experimenter and environmental variability (e.g., computerized version of Profiling Elements of Prosodic Systems-Child (PEPS-C), originally developed by Peppé and McCann (2003) and advanced by Peppé and colleagues, see http://www.peps-c.com/). However, as long as we use non-

spontaneous stimuli, we cannot escape from the question of whether the experimental findings are generalizable to the daily oral communication that children experience. One strategy we may adopt is the use of measurable materials from spontaneous speech corpora. Some studies have used speech from pre-existing corpora and tested the prominence ratings of naïve adult listeners (Cole, Mo, & Hasegawa-Johnson, 2010; Turnbull, Royer, Ito, & Speer, 2017), as well as their eye-movement responses to tree-decoration instructions (Ito, Turnbull, & Speer, 2017), in order to confirm listeners' sensitivities to natural speech. With the growing number of open resources such as CHILDES (http://childes.psy.cmu.edu/), and advanced recording devices such as the LENA system (https://www.lena.org/), experimentation with more naturalistic (yet controlled) stimuli is certainly becoming feasible.

In sum, affect prosody and focus prosody have been studied with different experimental paradigms, largely due to differences in the age of targeted populations. Measures of responses to affect prosody in pre-verbal children are necessarily indirect (e.g., gazes and brainwaves), whereas responses to focus prosody can be more directly observed with interactive tasks with older children. While applications of a visual-world paradigm have demonstrated children's immediate interpretation of contrastive focus prosody and its gradual development, experimental paradigms can be further advanced with more naturalistic discourse speech materials.

Slow development of prosodic skills and slow development of developmental theory: why does it take so long?

To explain the factors underlying the gradual developmental trajectory of contrastive prosody comprehension, Ito et al. (2012) and Ito, Bibyk, Wagner, and Speer(2014) have proposed

data-driven accounts that extend the views of Cruttenden (1985) and of Cutler and Swinney (1987). First, these studies found that 6-year-olds, both Japanese- and English-speaking, have a perseveration tendency that makes them take time to shift attention from the previous referential set to the new referential set. Thus, the efficacy of prosodic processing seems tightly related to the development of attention allocation that controls the speed of discourse representation (e.g., new/less accessible vs. old/accessible), which in turn affects the effectiveness of prosody-to-discourse mapping. Second, the fact that children take longer to recover from prosodic garden-pathing indicates that they are not as efficient as adults in detecting conflicts in the signal, revising prosody-based expectations, and letting segmental information guide the referential resolution. This requires inhibition of salient percepts of contrastive prosody and a switch of attention to segmental cues, which demands general cognitive flexibility. Executive function that includes attention allocation, inhibition, and cognitive flexibility is known to develop slowly throughout childhood and adolescence (Prencipe et al., 2011; Zelazo & Müller, 2002). Thus, it is reasonable to assume that it takes years to achieve the level of ability that mature processors have to integrate prosodic cues with referential information and to revise analyses when necessary.

Assuming that executive function has an intertwined relationship with various aspects of oral communication skills, it does not make sense to believe that the ability to process affect prosody is mastered in early childhood. As summarized in Armstrong and Hübscher (this volume), studies show that young children rely more on salient lexical information, facial expressions and gestural cues than on prosodic cues for interpreting affect, and even older children often exhibit difficulty understanding a speaker's emotion expressed by prosody. These findings may reflect an overall shortage of cognitive resources (such as memory and attention span) for multi-dimensional information processing in children, and

their tendency to ignore less direct and more variable cues. A cognitive capacity constraint on audio-visual processing is also suggested by unique eye-movement patterns in infants with Autism Spectrum Disorder, who look less at the face when speech signals accompany the visual stimuli than when they are presented silently (Shic, Macari, & Chawarska, 2014). As for the link between memory function and general language skills, a study by Vulchanova, Foyn, Nilsen, and Sigmundsson (2014) reports strong correlations between verbal working memory capacity (measured with a forward digit recall task) and grammar, vocabulary, and L2 spoken sentence comprehension measures in 10-year-old Norwegian children. Individual differences in verbal working memory are also known to affect adult sentence processing of complex syntactic structures such as object-extracted relative clauses (Fedorenko, Gibson, & Rohde, 2006; King & Just, 1991). While prosody conveys information about lexical identity, phrase structure, discourse context and social dynamics, it is likely that processing of signals related to information weight or not-so-evident expressions of affect is compromised in processors with limited capacity. Children must learn to attend to referential contexts and social factors (such as who is talking to whom) while holding on to the linguistic contents of utterances during conversation. The ability to quickly map prosodic cues to context-dependent informational structures and complex affect statuses must develop, to some degree, with the growth of cognitive resources.

Another fundamental factor that tends to be oversimplified in research on affect prosody is the granularity of affect categories. While many studies of infant affect detection test the distinction between canonical categories such as happy/positive vs. sad/negative (Grossmann, Striano, & Friederici, 2005; Villant-Molina, Bahrick, & Flom, 2013; Walker-Andrews & Grolnick, 1983), real-world communication often requires much finer distinctions of emotions along the valence and arousal dimensions (e.g., tired vs. bored, miserable vs.

depressed, delighted vs. pleased, content vs. calm, etc.: Russell, 1980). Research has shown that 5-year-olds can reliably recognize happy and sad faces yet have difficulty identifying expressions for fear, disgust and anger (Durand et al., 2007). Complex affect categories beyond these five basic emotions are therefore predicted to develop even later.

As for the auditory processing of affect, Flom and Bahrick (2007) have shown that 5- and 7-month-old infants discriminate happy, sad, and angry prosodies, and Villant-Molina, Bahrick, and Flom (2013) found that 5-month-olds can correctly map prosody to positive and negative facial expressions. It is yet to be discovered, however, when children acquire other affect categories such as fear and disgust and recognize them with particular prosodic cues. A study by Demenescu, Kato, and Mathiak (2015) showed that adults recognize happiness and anger better than sadness, disgust and fear in vocal expressions, and negative emotional recognition generally deteriorates with aging. Their fMRI data suggest that the reduction of sensory function (in the superior temporal gyrus) may underlie this decline. While the development of sensory function is considered as the primary neurological component of automatic affect detection, another recent study by Voyer, Thibodeau, and Delong (2014) showed that the interpretation of sarcasm in adults heavily relies on context as well as prosody. Thus, processing of affect prosody requires both quasi-automatic perception of acoustic cues (such as energy, voice quality and tune) and a fair grasp of discourse context, which may also rely on executive function abilities such as working memory and attention.

Finally, for the construction of a more empirically adequate theory of prosodic development, it is important to remember that different affect categories are expressed by different acoustic cues, which are not uniformly salient across different types and levels of emotions (Banse & Scherer, 1996; Juslin & Laukka, 2001; Laukka, Juslin, & Brestin, 2005; Sauter, Eisner, Calder, & Scott, 2010). For example, while Banse and Schere (1996) show

that panic, hot anger, happiness, sadness are respectively expressed with increased mean F0, mean energy, low frequency energy, and duration, Juslin and Laukka (2001) reveal that emotional intensity affects the use of the same acoustic cues differently for different emotion types: strong emotional intensity boosts the F0 floor for both fear and happiness, but it boosts the F0 expansion and F0 maxima only for fear but not for happiness, which already reaches the speaker's ceiling F0 level with weak emotional intensity. Sauter, Eisner, Calder, and Scott (2010) provide a diagram of which acoustic cues are mainly used for which types of emotion (e.g., pitch for surprise, anger, achievement, and relief but not for amusement, sadness, contentment, pleasure and disgust, which are expressed by spectrum and amplitude instead. Fear is expressed solely by spectrum according to this diagram). However, these cues do not always discern between different affect types: Sauter, Eisner, Calder, and Scott's (2010) experimental data demonstrate that both human listeners and discriminant analysis of acoustic measures tend to confuse anger with disgust, and contentment with pleasure or relief. Furthermore, while humans may misidentify fear as amusement (13%), discriminant analysis of acoustics mistakes fear for achievement (31%) and anger (25%).

These studies point to the fact that prosodic characterization of affect types is difficult, and acoustic cues can be unreliable and misleading at times. Based on the currently available empirical data, we may not want to assume any fixed bundle of acoustic features for any given affect category, as an expression of a particular emotional status can be achieved via various combinations of multi-modal cues. We may in fact achieve a better scientific description of human cognition by identifying the conditions under which particular prosodic features are processed as dominant cues to the observed emotional status rather than by seeking evidence for prosodic categories that directly map onto affect categories. Affect is a primary source for changes in speech rate, overall intensity, voice quality, and other

articulartory gestures as well as body language and facial expressions of the speaker: these behavioral signals in turn lead to the percept of a specific emotional status in the listener, who may not necessarily share the ranges of valence and arousal with the speaker. Thus, affect, either as the speaker's emotional status or the listener's percept, can never be defined by a finite set of absolute prosodic cues. This basic observation is often overlooked in the studies of affect prosody acquisition. To advance the theory of prosodic development, we must walk away from assumptions about direct mappings between affect categories and prosodic categories (for the acoustic difference between lab speech and spontaneous speech that lead to similar contrastive interpretation, see Ito, Turnbull, and Speer (2017).

In this section, I reviewed studies that suggest affect recognition is not easy once it requires more than a simple discrimination between happiness and sadness. I argue that maturity of executive functions underlies the slow development of both focus prosody and affect prosody, which are orchestrated flexibly with cues expressed through other modalities according to each communicative purpose and context. Importantly, we should bear in mind that seeking for a set of acoustic cues that invariably labels a type of speech act – whether it is for marking information structure or recognizing affect – is not a very rewarding approach, as compared to efforts to identify the mechanism of cue tuning or balancing.

Interaction and integration of affect prosody and focus prosody

At the beginning of this chapter, I suggested a direction of future studies that considers developmental trajectories of affect prosody and focus prosody simultaneously. This was in fact inspired by a comment from an individual with Williams syndrome (WS), who

participated in our eye-tracking study (Ito, Martens, & McKenna, 2014, March): “This person sounds very loud. I felt she was scolding me”. The stimuli were identical to those of Ito, Bibyk, Wagner, and Speer (2014), for which a young female phonetician produced questions such as “*Where is the pink monkey? Now, where is the GREEN monkey?*” imagining talking to a young child. To everybody up to that point, she sounded like a happy preschool teacher rather than a grumpy lady, and none of the typically developing child participants had expressed discomfort with her voice. While the above comment from an adult participant with WS was thus unexpected, it certainly provided food for thought. Individuals with WS are known to be hypersensitive to social cues (Dykens, 2003), and thus this participant may have paid particular attention to cues to speaker affect. However, this may also happen in everyday conversations among typically developing individuals. If rudimentary sensitivity to affect prosody develops during infancy, young children may well be capable of automatically detecting basic affect status (e.g., positive vs. negative) while attending to prosodic signals related to discourse structure.

To date, studies that simultaneously examine more than one function of prosody have been sparse. Pihan, Tabert, Assuras, and Borod (2008) asked participants to listen to a pair of sentences and indicate which one sounds more like a question. Three statements were produced by two speakers (one male, one female) in happy, neutral and fearful prosody, and their pitch contours were artificially modified such that each sentence ends with a rising, falling or level tone. Pihan et al. (2008) found that the rising tones generally increased the ‘question’ responses. However, the happy prosody that was characterized by the largest F0 changes throughout the utterances interfered with the speaker intention judgments, resulting in the lowest ‘question’ responses within the rising tone set. (The predicted right-lateralization for happy and fear prosody was not detected in the EEG signals.) The findings of Pihan et al.

(2008) suggest that unconsciously detected speaker emotion can affect the perception of illocutionary force. Thus, while laboratory experimenters typically make participants attend to particular dimensions of speech signals, listeners may assess the emotional state of a speaker automatically and make responses accordingly. This implication needs to be considered carefully for the study of focus prosody in toddlers and infants. On one hand, the function of accentuation in highlighting specific discourse entities may be blurred by happy or overly excited prosody with exaggerated pitch excursions. On the other hand, accents embedded in a cold angry tone with a compressed pitch range may be more efficiently processed if they stand out acoustically, although some listeners may interpret the emphasis as part of the expression of anger.

The interaction between affect and focus prosody may also modulate the effect of memory encoding in children. Fraundorf, Watson, and Benjamin (2010, 2012) have shown that contrastive accent leads to better recall of narratives in both younger and aging adults. Lee and Snedeker (2016) have partially replicated this effect of contrastive accent in 5-year-old children. Another recent study by Lee and Fraundorf (2016) confirmed some sensitivity to contrastive accent only in high-proficiency, but not in low-proficiency, L1-Korean learners of English. This finding adds to a general observation of difficulty in using prosodic cues for interpreting the speaker's intention in language learners (see Armstrong & Hübscher, this volume, for a summary of research on children's comprehension of prosody for belief state). A recent study by Igualada, Esteve-Gibert, and Prieto (in press) reports the effect of 'beat gestures' (gestures that highlight a part of speech) on the recall of verbal information in 3- to 5-year-old children, suggesting that informational retrieval can be facilitated by associated visual cues. Since this finding demonstrates preschoolers' ability to bind cues for retrieving spoken information, we may hypothesize that prosodic prominence can also function as a

retrieval cue in young children. However, the critical question is whether prosodic cues alone can flag the discourse status of a referential expression and facilitate the retrieval of it. Based on the results of Ito, Bibyk, Wagner, and Speer (2014), I doubt that prosodic prominence is as effective as visually-associated cues for memory retrieval in young children, because segmental and prosodic cues must compete for the limited cognitive resources available for auditory processing. In addition, if the listener happens to attend to the emotional state of the speaker while processing the segmental information for syntactic and lexical semantic structures of the spoken message, the resources for auditory input (or the 'phonological loop' component in Baddeley and Hitch's (1974) model of working memory) may be overloaded with an underdeveloped processor. This 'limited resource hypothesis' needs to be tested with careful experimental designs, because particular emotional percepts may encourage or discourage memory encoding as well as informational representation in listeners. Caution is required when preparing experimental stimuli, because accents may not stand out in an overly expanded pitch range that may express a high level of arousal (as in IDS), while a specific combination of acoustic cues for expressing a particular degree of valence may sharpen or dilute the effect of prosodic emphasis.

While these hypotheses about the interaction of affect prosody and focus prosody with limited cognitive resources remain to be explored, the research outcome for such questions would be very beneficial not only to the field of developmental intonational phonology, but also to the research fields of educational psychology, language pedagogy and communication disorders. One common goal across these applied research fields is to find a way to improve individuals' communication skills. There is little room to debate whether excellence in communication skills comprises of good understanding of speaker emotion and a fair comprehension of discourse context and structure. Since prosody provides robust cues to both

these components, developmental research in this area should take a more holistic approach in which the bi-directional interaction between affect prosody and focus prosody is examined across multiple developmental stages. This approach should investigate whether and how children's percept of speaker affect impacts their understanding of message structure, and how this interaction changes as they acquire fine-grained affect categories and discourse representations. It is also important to examine how efficacy of sensory processing, sensitivity to social cues, cognitive flexibility, and memory impact the way emotion recognition influences message comprehension. Although it is impossible to assess all these inter-related abilities of an individual in one study, strategic experimental designs and inter-disciplinary collaborations that draw on multiple expertises have strong potential to overcome methodological challenges and advance our understanding of communication development.

Acknowledgments

I thank the two anonymous reviewers who gave me very encouraging and constructive comments.

References

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

Baddeley, A. D., & Hitch, G. (1974). Working memory. *The psychology of learning and motivation: Advances in research and theory,* 8, 47–89.

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, 70, 614–636.

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

Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English, *Phonology Yearbook*, 3, 255–309.

Belyk, M., & Brown, S. (2013). Perception of affective and linguistic prosody: An ALE meta-analysis of neuroimaging studies. *Social Cognitive and Affective Neuroscience*, 9, 1395–1403.

Berman, J. M., Chambers, C. G., & Graham, S. A. (2010). Preschoolers' appreciation of speaker vocal affect as a cue to referential intent. *Journal of Experimental Child Psychology*, 107(2), 87–99.

Berman, J. M., Chambers, C. G., & Graham, S. A. (2016). Preschoolers' real-time coordination of vocal and facial emotional information. *Journal of Experimental Child Psychology*, 142, 391–399.

Bolinger, D. (1978). Intonation across languages. In , J. H. Greenberg, C. A. Ferguson & E. Moravcsik (Eds.), *Universals of human language: Phonology* (Vol. 2 pp. 471–524). Stanford: Stanford University Press.

Bull, D. H., Eilers, R. E., & Oller, D. K. (1984). Infants' discrimination of intensity variation in multisyllabic stimuli. *Journal of the Acoustical Society of America*, 76, 13–17.

Chen, A. (2014). Production-comprehension Asymmetry: Individual differences in the acquisition of prosody focus-marking. *Proceedings of the 7th International Conference on Speech Prosody – Dublin*, 423–427.

Cole, J., Mo, Y., & Hasegawa-Johnson, M. (2010). Signal-based and expectation-based factors in the percetion of prosodic prominence. *Laboratory Phonology*, 1, 425–452.

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development*, 61, 1584–1595.

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

Cruttenden, A. (1974). An experiment involving comprehension of intonation in children from 7 to 10. *Journal of Child Language*, 1, 221–232.

Cruttenden, A. (1985). Intonation comprehension in ten-yea-olds. *Journal of Child Language* 12, 643–661.

Crystal, D. (1979). Prosodic development. In P. Fletcher, & M. Garman (Eds.), *Language acquisition: Studies in first language development* (pp. 33–48). Cambridge: Cambridge University Press.

Edwards, J., Pattison, P. E., Jackson, H. J., & Wales, R. J. (2001). Facial affect and affective prosody recognition in first-episode Schizophrenia. *Schizophrenia Research*, 48, 235–253.

Demenescu, L. R., Kato, Y., & Mathiak, K. (2015). Neural processing of emotional prosody across the adult lifespan. *BioMed Research International*. Article ID 590216.

Durand, K., Gallay, M., Seigneuric, A., Robichon, F., & Baudouin, J. Y. (2007). The development of facial emotion recognition: The role of configural information. *Journal of Experimental Child Psychology*, 97, 14–27.

Dykens, E. M. (2003). Anxiety, fears, and phobias in persons with Williams syndrome. *Developmental Neuropsychology*, 23, 291–316.

Eilers, R. E., Bull, D. H., Oller, D. K., & Lewis, D. C. (1984). The discrimination of vowel duration by infants. *Journal of the Acoustical Society of America*, 75, 1213–1218.

Ethofer, T., Bretscher, J., Gschwind, M., Kreifelts, B., Wildgruber, D., & Vuilleumier, P. (2012). Emotional voice areas: anatomic location, functional properties, and structural connections revealed by combined fMRI/DTI. *Cerebral Cortex*, 221, 191–200.

Ethofer, T., Kreifelts, B., Wiethoff, S., Wolf, J., Grodd, W., Vuilleumier, P., & Wildgruber, D. (2009). Differential influences of emotion, task, and novelty on brain regions underlying the processing of speech melody. *Journal of Cognitive Neuroscience*, 217, 1255–68.

Fedorenko, E., Gibson, E., & Rhode, D. (2006). The nature of working memory in linguistic, arithmetic, and spatial integration processes. *Journal of Memory and Language*, 56, 246–269.

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

Fernald, A. (1993). Approval and disapproval: Infant responsiveness to vocal affect in familiar and unfamiliar languages. *Child Development*, 64, 657–674.

Fernald, A. (2004). Hearing, listening, and understanding: Auditory development in infancy. In G. Bremner, & A. Fogel (Eds.), *Blackwell handbook of infant development* (pp. 35–70). London: Blackwell Publishing.

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

Fernald, A., & Mazzie, C. (1991. Prosody and focus in speech to infants and adults.

Developmental Psychology, 27 2, 209–221.

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*, 16(3), 477–501.

Fernandez, R., & Picard, R. (2011). Recognizing affect from speech prosody using hierarchical graphical models. *Speech Communication,* 53, 1088–1103.

Flom, R., & Bahrick, L. E. (2007). The development of infant discrimination of affect in multimodal and unimodal stimulation: The role of intersensory redundancy. *Developmental Psychology,* 431, 238–252.

Fraundorf, S. H., Watson, D. G., & Benjamin, A. S. (2010). Recognition memory reveals just how CONTRASTIVE contrastive accenting really is. *Journal of Memory and Language,* 63, 367–386.

Fraundorf, S. H., Watson, D. G., & Benjamin, A. S. (2012). The effects of age on the strategic use of pitch accents in memory for discourse: A processing-resource account. *Psychology and Aging*, 27(1), 88–98.

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

Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M., Scherer, K. R., & Vuilleumier, P. (2005). The voices of wrath: Brain responses to angry prosody in meaningless speech. *Nature Neuroscience,* 8*2,* 145–146.

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

Grossmann, T., Striano, T., & Friederici, A. D. (2005). Infants' electric brain responses to emotional prosody. *Neuroreport,* 16, 1825–1828.

Hargrove, P. M. (2013). Pursuing prosody interventions. *Clinical Linguistics & Phonetics*, 27, 647–660.

Hirschberg, J. (2002). The pragmatics of intonational meaning. *Proceedings of Speech Prosody 2002 – Aix-en-Provence*, 65–68.

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

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

Hornby, P., & Hass, W. (1970). Use of contrastive stress by preschool children. *Journal of Speech and Hearing Research*, 13, 395–399.

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

Ito, K., & Martens, M. (2017). Contrast-marking prosodic emphasis in Williams syndrome: results of detailed phonetic analysis. *International Journal of Language and Communication Disorders*, 52, 46–58.

Ito, K. (2014). Children's pragmatic use of prosodic prominence. In D. Matthews (Ed.), *Pragmatic development in first language acquisition* (pp. 199–218). Amsterdam: John Benjamins.

Ito, K., Bibyk, S., Wagner, L., & Speer, S. R. (2014). Interpretation of contrastive pitch accent in 6- to 11-year-old English speaking children and adults. *Journal of Child Language*, 41(1), 84–110.

Ito, K., Jincho, N., Minai, U., Yamane, N., & Mazuka, R. (2012). Intonation facilitates contrast resolution: Evidence from Japanese adults & 6-year olds. *Journal of Memory*

and Language, 66(1), 265–284.

Ito, K., Martens, M., & McKenna, E. (2014, March). *Processing of pitch prominence in Williams syndrome.* Talk presented at 27th Annual CUNY Conference on Human Sentence Processing, Columbus, USA.

Ito, K., Turnbull, R., & Speer, S. R. (2017). Allophonic tunes of contrast: Lab and spontaneous speech lead to equivalent fixation responses in museum visitors. *Laboratory Phonology*, 8(1): 6, 1–29 .

Jun, S-A. (Ed.) (2005). *Prosodic typology: The phonology of intonation and phrasing.* Oxford: Oxford University Press.

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

Jusczyk, P. W., Cutler, A., & Redanz, N. (1993). Preference for the predominant stress patterns of English words. *Child Development*, 64, 675–687.

Juslin, P. N., & Laukka, P. (2001). Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion. *Emotion*, 1, 381–412.

Katz, G., Cohn, J., & Moore, C. (1996). A combination of vocal F0 dynamic and summary features discriminates between three pragmatic categories of infant-directed speech. *Child Development*, 67, 205–217.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, 30, 580–602.

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, 9, F13–F21.

Kurumada, C. (2013). *Navigating variability in the linguistic signal: Learning to interpret contrastive prosody*. PhD dissertation. Stanford University.

Kurumada, C., Brown, M., Bibyk, S., Pontillo, D., & Tanenhaus, M. K. (2014). Is it or isn't it: Listeners make rapid use of prosody to infer speaker meanings. *Cognition*, 133, 335–342.

Laukka, P., Juslin, P., & Bresin, R. (2005). A dimensional approach to vocal expression of emotion. *Cognition & Emotion*, 19, 633–653.

Lee, E-K., & Fraundorf, S. (in press). Effects of contrastive accents in memory for L2 discourse. *Bilingualism: Language and Cognition.* http://dx.doi.org/10.1017/S1366728916000638

Lee, E-K., & Snedeker, J. (2016). Effects of contrastive accents on children's discourse comprehension. *Psychonomic Bulletin & Review*, 23, 1589–1595.

Ladd, R. D. (2008). *Intonational phonology*. 2nd edition. Cambridge: Cambridge University Press.

Ma, W., Golinkoff, R. M., Houston, D., & Hirsh-Pasek, K. (2011). Word Learning in Infant- and Adult-Directed Speech. *Language Learning and Development*, 7, 209–225.

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

Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton P. M. (1999). Rule learning by 7-month-old infants. *Science*, 283, 77–80.

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

Oudeyer, P-Y. (2003). The production and recognition of emotion in speech: Features and algorithms. *International Journal of Human-Computer Studies*, 59, 157–183.

Owren, M. J., & Rendall, D. (1997). An affect-conditioning model of nonhuman primate vocalizations. In D. W. Owings, M. D. Beecher, & N. S. Thompson (Eds.), *Perspectives in ethology* (Vol. 12, pp. 299–346). New York: Plenum Press.

Peppé, S. J. E. (2009). Aspects of identifying prosodic impairment. *International Journal of Speech-Language Pathology*, 11(4), 332–338.

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

Pihan, H., Tabert, M., Assuras, S., & Borod, J. (2008). Unattended emotional intonations modulate linguistic prosody processing. *Brain and Language*, 105, 141–147.

Prencipe, A., Kesek, A., Cohen, J., Lamm, C., Lewis, M.D., & Zelazo, P. D. (2011). Development of hot and cool executive function during the transition to adolescence. *Journal of Experimental Child Psychology,* 108, 621–637.

Ross, E. D. (1981). The aprosodias: functional-anatomic organization of the affective components of language in the right hemisphere. *Archives of Neurology*, 38, 561–9.

Russell J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39, 1161–1178.

Sakkalou, E., & Gattis, M. (2012). Infants infer intentions from prosody. *Cognitive Development,* 27, 1–16.

Saffran, J.R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month-Old Infants. *Science*, 274, 1926–1928.

Sauter, D. A., Eisner, F., Calder, A. J., & Scott, S. K. (2010). Perceptual cues in nonverbal vocal expressions of emotion. *The Quarterly Journal of Experimental Psychology*, 63(11), 2251–2272

Scherer, K. R., & Kappas, A. (1988). Primate vocal expression of affective state. In D. Todt, P. Goedeking, & D. Symmes (Eds.), *Primate vocal communication* (pp. 171–194). Berlin: Springer.

Schuller, B., Steidl, S., & Batliner, A. (2009). The INTERSPEECH 2009 Emotion Challenge. *Proceedings of the 10th International Speech Communication Association Annual Conference - Brighton*, 312–315.

Schuller, B., Steidl, S., Batliner, A., Burkhardt, F., Devillers, L., Müller, C., & Narayanan, S., (2010). THE INTERSPEECH 2010 Paralinguistic Challenge. *Proceedings of the 11th International Speech Communication Association Annual Conference - Chiba*, 2794–2797.

Sekerina, I. E., & Trueswell, J. C. (2012). Interactive processing of contrastive expressions by Russian children. *First Language,* 32(1-2), 63–87.

Shic, F., Macari, S., & Chawarska, K. (2014). Speech disturbs face scanning in 6-month-old infants who develop Autism Spectrum Disorder. *Biological Psychiatry,* 75, 231–237.

Shukla, M., White, K. S., & Aslin, R. N. (2011). Prosody guids the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. *Proceedings of the National Academy of Sciences of the United States of America,* 108, 6038–6043.

Solan, L. (1980). Contrastive stress and children's interpretation of pronouns. *Journal of Speech and Hearing Research,* 23, 688–698.

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

Tomalski, P. (2015). Developmental trajectory of audiovisual speech integration in early infancy. A review of studies using the McGurk paradigm. *Psychology of Language and Communication*, 19, 77–100.

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition,* 73, 89–134.

Trueswell, J. C., & Tanenhaus, M. K. (2005). *Approaches to studying world-situated language use*. Cambridge, MA: MIT Press.

Turnbull, R. (2016). The role of predictability in intonational variability. *Language and Speech,* 60, 123-153.

Turnbull, R., Royer, A., Ito, K., & Speer, S. R. (2017). Prominence perception is dependent on phonology, semantics, and awareness of discourse. *Language, Cognition and Neuroscience.* DOI: 10.1080/23273798.2017.1279341

Vaillant-Molina, M., Bahrick, L. E., & Flom, R. (2013). Young Infants Match Facial and Vocal Emotional Expressions of Other Infants. *Infancy*, 18(suppl. 1), E97–E111.

Voyer, D., Thibodeau, S-H., & Delong, B. J. (2014). Context, contrast, and tone of voice in auditory sarcasm perception. *Journal of Psycholinguistic Research,* 45, 29–53.

Vulchanova, M., Foyn, C. H., Nilsen, R. A., & Sigmundsson, H. (2014). Links between phonological memory, first language competence and second language competence in 10-year-old children. *Learning and Individual Differences,* 35, 87–95.

Walker A. S. (1982). Intermodal perception of expressive behaviors by human infants. *Journal of Experimental Child Psychology,* 33, 514–535.

Walker-Andrews, A. S. (1997). Infants' perception of expressive behaviors: Differentiation of multimodal information. *Psychological Bulletin,* 121, 437–456.

Walker-Andrews, A. S., & Grolnick W. (1983). Discrimination of vocal expression by young infants. *Infant Behavior & Development*, 6, 491–498.

Walker-Andrews A. S., & Lennon, E. (1991). Infants' discrimination of vocal expressions: Contributions of auditory and visual information. *Infant Behavior & Development,* 14, 131–142.

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

Werker, J. F., & Tees, R. (2002). Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development,* 25, 121–133.

Wildgruber, D., Ackermann, H., Kreifelts, B., & Ethofer, T. (2006). Cerebral processing of linguistic and emotional prosody: fMRI studies. *Progress in Brain Research,* 156, 249–268.

Zelazo, P. D., & Müller, U. (2002). Executive function in typical and atypical development. In U. Goswami (Ed.), *Handbook of childhood cognitive development* (pp. 445–469). Oxford, UK: Blackwell.