Prosodic development during the early school-aged years

Jeffrey E. Kallay\textsuperscript{1}, Laura Dilley\textsuperscript{2}, and Melissa A. Redford\textsuperscript{*\textsuperscript{,1}}

* Corresponding Author

\textsuperscript{1} Department of Linguistics  
University of Oregon  
Eugene, OR 97403-1290  
redford@uoregon.edu

\textsuperscript{2} Department of Communication Sciences & Disorder  
Michigan State University  
East Lansing, MI 48824

There are no conflicts of interest.

This research was supported by the Eunice Kennedy Shriver National Institute of Child Health & Human Development (NICHD) under grants R01HD061458 (PI: Redford) and R01HD087452 (PI: Redford). The content is solely the authors’ responsibility and does not necessarily reflect the views of NICHD.
Abstract

Purpose: The current study used a cross-sequential design to identify developmental changes in narrative speech rhythm and intonation. The aim was to provide a robust, clinically-relevant characterization of normative changes in speech prosody across the early school-aged years.

Methods: Structured spontaneous narratives were elicited annually from 60 children over a three year period. Children were aged 5 to 7 years at study outset, and so were aged 7 to 9 years at study offset. Articulation rate, prominence spacing, and intonational phrase length and duration were calculated for each narrative to index speech rhythm; measures of pitch variability and pitch range indexed intonation. Linear mixed effects (LME) models tested for cohort-based and within-subject longitudinal change on the prosodic measures; linear regression was used to test for the simple effect of age-in-months within year on the measures.

Results: The LME analyses indicated systematic longitudinal changes in speech rhythm across all measures except phrase duration; there were no longitudinal changes in pitch variability or pitch range across the school-aged years. Linear regression results showed an increase in articulation rate with age; there were no systematic differences between age cohorts across years in the study.

Conclusions: The results indicate that speech rhythm continues to develop during the school-age years. The results also underscore the very strong relationship between the rate and rhythm characteristics of speech, and so suggest an important influence of speech motor skills on rhythm production. Finally, the results on pitch variability and pitch range are interpreted to suggest that these are inadequate measures of typical intonation development during the school-aged years.
Prosodic development

Prosody is the integrated rhythm and intonation patterns of spoken language. It is impacted in speech produced by children with a wide range of developmental disorders (Paul et al., 2020)—from dysarthria and childhood apraxia of speech (Patel et al., 2012; Peter & Stoel-Gammon, 2005) to autism spectrum disorder and William’s syndrome (Catterall et al., 2006; Ito & Marten, 2017; Martinez-Castilla et al., 2011; Peppé, 2009, 2018). The breadth of impact reflects the extended domain of the prosodic system. Prosody emerges from stress patterns at the syllable and word level and from lengthening, accenting, and pausing patterns at the phrase and discourse level. The extended domain of prosody makes it one of the most salient features of any individual’s speech and so also a salient marker of difference from one’s peers. We know that atypical rhythm and intonation patterns increase the perception of disorder (Olejarczuk & Redford, 2013; Redford et al., 2018; Shriberg et al., 2001); and that perception of disorder adversely affects a child’s access to positive peer interactions (e.g., McCabe & Meller, 2004; Redford et al., 2018; Rice et al., 1991). Isolation from one’s peers undermines a child’s social-emotional development and cognitive functioning (Rubin et al., 2015). Speech intelligibility also suffers with atypical prosody (Klopfenstein, 2009)—an observation that has been made many times and across many different populations with disordered speech prosody, including in children with Down syndrome (Kent & Vorperian, 2013; Stojanovik, 2011; Wilson et al., 2019), autism spectrum disorder (ASD; Redford et al. 2018), childhood apraxia of speech (Connaghan & Patel, 2012), dysarthria (Patel et al., 2012), and hearing impairment (Chin et al., 2012; Parkhurst & Levitt, 1978). Given the breadth of impact of prosody on communication, there is a recognized need to equip speech language pathologists with tools to assess and treat prosodic disorder in order to improve children’s access to positive social interaction and speech intelligibility (Hawthorne & Fischer, 2020; Kalathottukaren et al., 2015; Peppé, 2009).
Tools and strategies for prosodic assessment must reference normative development to be useful and effective (see Diehl & Paul, 2009). But the empirical description of prosody in typically developing children’s speech is incomplete. Extant research has mostly focused on very early acquisition (e.g., Arcuili & Colombo, 2016; Jusczyk et al., 1993; Kehoe, 2000; Kehoe et al., 1995; Payne et al., 2011; Schwartz et al., 1996; Snow, 1994, 1997; Vihman et al., 2006). There is a relative dearth of information on the later stages of development. The present study seeks to address this deficit. The aim is to provide a clinically-useful description of prosodic development during the school-aged years. To do this, we focus on narrative speech rather than on isolated words and sentences, and on global measures of rhythm and intonation rather than on the linguistic and phonetic detail of these prosodic attributes.

Much of the prior research on prosody in children’s speech addresses its acquisition as part of the linguistic grammar. Our specific interest is in the emergence of speech prosody at the intersection of language and speech motor skills. Both sets of skills show protracted developmental trajectories: higher-level syntax and discourse skills continue to change throughout the school-aged years (Nippold, 2016); speech motor skills show significant change during the early school-aged years and continue to be refined until at least age 14 years (Smith & Zelaznik, 2004). The influence of developing speech motor skills on prosody production is likely to be especially evident in narrative speech where the task demands are high relative to single word or single sentence elicitations. This suggestion is based on the effect of language complexity and task difficulty on speech motor performance in children and adults: articulatory timing becomes more variable with increasing linguistic complexity and speech task difficulty (Maner et al., 2000; Saletta et al., 2018). Under the assumption that prosody emerges at the
intersection of language and speech motor skills, the expectation is that global speech prosody will change in measurable ways across the school-aged years.

Background

Speech Rhythm Development

Most of what we currently know of the typical development of speech rhythm is based on words and phrases elicited from very young children (e.g., Arcuili & Colombo, 2016; Jusczyk et al., 1993; Kehoe, 2000; Kehoe et al., 1995; Schwartz et al., 1996; Snow, 1994, 1997; Vihman et al., 2006). For example, Kehoe (2000) elicited multisyllabic words from 22-34 month old children to investigate the developmental trajectory of weak syllable deletion. Earlier studies had established the development of acoustic correlates of lexical stress in single word productions elicited from children aged 18 to 30 months old (Kehoe et al., 1995; Schwartz et al., 1996). Vihman (1996) used perceptual measures to investigate the development of adult-like rhythm patterns in disyllabic words produced by even younger infants (11-17 months old). Other studies have investigated rhythm development in multiword phrases elicited from older children, but still mostly in the pre-school age range (e.g., Grabe et al., 1999; Lleo et al., 2007; Payne et al., 2011). These studies generally suggest that children acquire a stress-timed language rhythm more slowly than they acquire a syllable-timed language rhythm. English is considered a stress-timed language.

The focus of prior studies on very young children’s speech might be taken to suggest that rhythm acquisition is complete before the school-age years. But a handful of studies with older children strongly suggest that this is not the case for children who are acquiring English. More specifically, study findings indicate the protracted development of phonetic detail associated with lexical stress production (Arciuli & Ballard, 2017; Ballard et al., 2012), function word
Prosodic development

reduction (Redford, 2018), and rhythm-related temporal patterning (Polyanskaya & Ordin, 2015; Sirsa & Redford, 2011). Like most studies with young children, the studies with older children are based on highly-controlled speech materials and use cross-sectional study designs to infer patterns of development. For example, Arciuli & Ballard (2017) used a picture naming task to elicit 4 multisyllabic words with contrastive stress patterns from 77 Australian English-speaking children who were between 8 and 11 years old at the time of study. They found that, as a group, older children did not modulate amplitude in weak-strong sequences in the same way as adults. Redford (2018) elicited simple subject-verb-object sentences from 24 school-aged children and 12 adults to study function word reduction. Using measures of relative amplitude and duration, she found that even 8-year-old children did not reduce unstressed function words to the same extent as adults.

The above-mentioned differences between school-aged children’s and adults’ stressed syllable production contribute to measured differences in the temporal patterning of child and adult speech (e.g., Sirsa & Redford, 2011). Setting aside questions about the link between temporal patterning and speech rhythm perception (Arvaniti, 2012), interval-based rhythm metrics clearly indicate change across the school-aged years. For example, Polyanskaya and Ordin (2015) elicited a set of sentences from 10 British English-speaking adults and 42 British English-speaking children, who ranged in age from 4 to 11 years old. The sentences were segmented into consonantal and vocalic intervals; mean segment duration, mean syllable duration, and rhythm metrics were calculated based on these intervals (e.g., VarcoC, %V; see White & Mattys, 2007). Metrics that captured interval variability (e.g., VarcoV and VarcoC) varied systematically with age and distinguished younger children’s speech from older children’s
speech; mean segment and syllable durations distinguished between every age group, including between older children’s speech and adults’ speech.

Polyanskaya and Ordin’s (2015) findings suggest a relationship between articulation rate and rhythm. The same relationship has been identified in adult language. For example, Arvaniti (2012) found significant and consistent differences within a language when interval-based rhythm metrics were calculated for spontaneous speech versus connected read speech versus sentence reading, presumably because of differences in articulation rate. More relevant to the present study, Dellwo (2010) showed that articulation rate is as effective as different interval-based rhythm metrics for distinguishing between language rhythm classes, such as stress-timed versus syllable-timed languages (German versus French). Moreover, he found that listeners rely on rate more than on variability in segment durations to discriminate timing patterns. We take this latter finding to suggest that articulation rate can be used to index the temporal patterning in speech associated with language rhythm.

The finding that articulation rate differentiates rhythm classes is consistent with the hypothesis that prosody emerges at the intersection of language and speech motor skills. After all, increasing articulation rate indexes better motor skills in children’s speech (Lee et al., 1999; Mahr et al., 2021; Redford, 2014), and articulation rate is positively correlated with more mature rhythm-related temporal patterning in children’s speech (Polyanskaya & Ordin, 2015). The hypothesis also helps to make sense of why rhythm (and rate) are impacted in developmental disorders that are classified as neuromotor speech disorders rather than language disorders (e.g., childhood apraxia of speech and dysarthria). The hypothesis also predicts that other global
Prosodic development

rhythm measures besides articulation rate will change across the school-aged years with the
development of language and speech motor skills.

Development of Intonation

Similar to the literature on speech rhythm development, the literature on intonational
development is mainly focused on words and phrases elicited from very young children’s speech
(for relevant reviews, see: Chen et al., 2020; Frota & Butler, 2018). But, unlike the
developmental literature on speech rhythm, there is less evidence for the protracted development
of intonation patterns. In fact, Frota and Butler argue that intonation develops “independent(ly)
from the onset of combinatorial speech (p. 156).” More specifically, they argue that the
acquisition of language-specific accentual patterns and boundary tones is largely complete by
age 2 years. And yet, other work suggests that speech motor control over fundamental frequency
(F0), the acoustic correlate of intonation, has a protracted developmental trajectory.

Lee and colleagues (1999) demonstrated that pitch variability, measured as the standard
deviation in F0, declined from age 5 years to around age 14 years before leveling out; the decline
between age 5 and 8 years was found to be especially dramatic. The results were based on single
words elicited from a very large sample of typically developing children. Consistent with a motor
interpretation of the results, Lee and colleagues reported similar age-related declines in the
temporal and spectral variability with which the same children produced consonants and vowels in
the elicited words. The motor interpretation of decreasing variability in measures of pitch also
suggests a particular interpretation of the finding that pitch variability and pitch range are greater in
speech produced by children with ASD compared to speech produced by their typically-developing
peers (Bone et al., 2014; Bonneh et al., 2011; Fosnot & Jun, 1999). In particular, it suggests a
pattern of prosodic delay rather than disorder in children with ASD (Diehl et al., 2009). But the
Prosodic development

main point is that, even in typical development, the intonational aspect of speech prosody may continue to change into the school-aged years due to the protracted development of F0 control.

Another reason to suspect that intonation may continue to develop into the school-aged years is because some of its linguistic aspects only develop with the acquisition of higher language abilities (Wells et al., 2004; Chen et al., 2020). For example, Wells and colleagues (Wells et al., 2004) used a battery of tasks designed to assess intonation comprehension and production in 120 British English-speaking children between the ages of 5 and 13 years of age. They found that children were at ceiling on most production tasks (e.g., contrastive focus marking); but younger school-aged children had trouble marking familiarity versus surprise when repeating back a word that was either familiar (e.g., carrot) or unfamiliar (e.g., pargle); that is, the children had trouble with a specific semantic-pragmatic function of intonation. Relatedly, Chen and colleagues have shown the delayed acquisition of certain semantic-pragmatic functions of intonation (e.g., Chen, 2009; Chen, 2011; Romøren & Chen, 2014). For example, Chen (2011) found that, whereas the overall intonation pattern of different types of phrases had been acquired by age 5 years, Dutch-speaking children did not systematically mark topic versus focus correctly until age 8 years. The protracted development of this aspect of intonation likely depends on cognitive development and the acquisition of higher-level language skills (Chen et al., 2020). Given that Dutch is like English in that it is an intonational language, these findings reinforce the expectation that the intonational aspects of speech prosody will continue to change during the school-aged years.

Present Study

The typical development of prosody during the school-aged years has not been studied systematically. There is a particular gap in the description of prosody in children’s spontaneous speech. Gaps in the description of typical prosodic development have negative consequences for
Prosodic development

the assessment and treatment of prosodic disorder in children (Diehl & Paul, 2009; Kalathottukaren et al., 2015). The present study seeks to address these gaps. To do this, we calculated global measures of rhythm and intonation from narrative speech samples that were collected once a year for 3 years from English-speaking school-aged children who were between 5 and 7 years old at the start of the 3-year period. The description of typical prosody in school-aged children’s narrative speech is likely to be especially useful for clinical purposes. After all, this type of speech is frequently elicited during assessment to characterize the global patterns of a child’s spontaneous speech (Heilmann et al., 2010). Also, prosody is no different from other aspects of language: longer speech samples provide a better foundation for its description than shorter speech samples. Importantly, some aspects of prosody – such as rhythmic phrasing – only emerge in longer speech samples. Narrative speech also facilitates the study of prosody production at the intersection of developing speech motor and language skills: narratives are associated with more complex language than conversational speech (Nippold et al., 2014); a narrative task is also more challenging than other commonly used speech elicitation tasks and so is more likely than these other tasks to tax the speech motor system (see Saletta et al., 2018), rendering speech motor effects on prosody production more evident than they might otherwise be.

Study predictions

The global rhythm measures calculated from the children’s narrative speech samples were articulation rate and perceptually-based measures of stress-timing and rhythmic phrasing. The global intonation measures calculated were pitch variability and pitch range. The predictions were that the global measures of rhythm would increase over developmental time with age-related development in motor speech and language skills, while the global measures of intonation would
Prosodic development

decrease with the development of these skills. In what follows, we give the motivation for each measure and for its predicted increase or decrease.

Articulation rate. As already noted, articulation rate is at least as good of a measure of rhythm-related temporal patterning as the more difficult-to-extract interval-based rhythm metrics (Dellwo, 2010; Polyanskaya & Ordin, 2015). The specific prediction was for an increase in articulation rate over developmental time. This prediction follows directly from previous cross-sectional and longitudinal study findings on articulation rate (Amir & Grinfeld, 2011; Mahr et al., 2021; Polyanskaya & Ordin, 2015; Scholderle et al., 2021; Walker & Archibald, 2006) and from the assumption that rhythm production is influenced by the development of speech motor skills.

Since language rhythm cannot be reduced to temporal patterning, it cannot be reduced to articulation rate. In linguistics, rhythm is understood to emerge from the spacing of prominences and from the timing of prosodic boundaries (Arvaniti, 2012; Ladd, 2008). Prominence is a perceptual construct (Ladd, 2008). The perception of prominence is cued by a combination of syllable lengthening (= syllable-to-syllable differences in acoustic duration), syllable strengthening (= syllable-to-syllable differences in acoustic intensity), and accenting (= phrase-internal F0 minima or maxima). Prosodic boundaries are also multiply cued, including by syllable lengthening, boundary tones (e.g., a final fall or rise in F0), and pausing. Accordingly, in the present study, prominence spacing and the timing of prosodic boundaries was calculated based on perceptual judgments.

Prominence spacing. This measure captures the alternation in prominences that is typical of a stress-timed language like English. It has long been reported that younger English-speaking children produce speech rhythms that are more equally stressed (or, syllable-timed) than the speech rhythms produced by older children and adults (Allen & Hawkins, 1980; Grabe et al., 1999). This
Prosodic development

perception is likely tied to the slow development of function word reduction (Allen & Hawkins, 1980; Sirsa & Redford, 2011), which itself is likely tied to speech motor skill development (Redford, 2018). The prediction was therefore that prominence spacing in narrative speech would increase with children’s age from a situation where nearly every syllable is perceived as prominent to one where prominent syllables occur (at most) every other syllable.

Phrasing. The specific measures of phrasing in the current study were phrase length, phrase duration, and variability in phrase length. Like the other global measures of rhythm, the prediction was that these would increase with developmental time. But, unlike with the other measures, this prediction was based on the assumption that developing cognitive and language skills are the most important factor driving changes in the rhythmic phrasing of children’s speech. Prosodic phrase boundaries are, after all, usually aligned with strong syntactic boundaries (Turk & Shattuck-Hufnagel, 1996) and so their timing will depend on the length of a clause and/or communication unit. Along these lines, we have previously found a positive relationship between syntactic complexity and the mean length of pause delimited utterances in children’s narrative speech (see, e.g., Kallay & Redford, 2021). This relationship predicts a developmental increase in phrase length and phrase duration with developing language skills. Also, as syntactic complexity increases, more diverse expressions are possible. A greater diversity of expression should result in a greater range of phrase lengths. The prediction is therefore that phrase lengths will also become more variable with development because language complexity increases with age.

Pitch variability and pitch range. The global intonation measures of pitch variability and pitch range have been used extensively in studies designed to characterize speech differences between children with typical and atypical prosodic development (e.g., Bone et al., 2014; Bonneh et al., 2011; Diehl et al., 2009; Fosnot & Jun, 1999). Testing for developmental changes in these
measures is thus particularly well-aligned with the goal of providing a clinically-useful description of typical prosodic development across the early school years. Pitch variability and pitch range were predicted to decline with development, under the assumption that these measures are tied to speech motor skills in extended speech samples. If the measures are instead more closely tied to the semantic-pragmatic function of intonation, they will likely increase with development time as children become increasingly sophisticated in their marking of semantics and discourse structure in their narratives.

The core study predictions of a developmental increase in global rhythm measures and a developmental decrease in global intonation measures were tested in narratives elicited from a cross-sectional sample of children across a 3-year period. In other words, the study used a cross-sequential design. This design provides the independent advantages of cross-sectional and longitudinal designs (see Galbraith et al., 2017), while minimizing the drawbacks of both. Stritch (2017:222) notes that “researchers cannot use static data to directly test dynamic theories” because cohort effects may mask developmental ones or because they may be inappropriately interpreted as developmental effects. A longitudinal design is the most appropriate design for describing developmental trajectories. But longitudinal studies are time-consuming. The cross-sequential design allows for the assessment of development across a larger range of ages in a shorter amount of time, with the advantages of increased sample size and a reduction in the dropout rate typically associated with more traditional longitudinal designs (see also Verbeke & Lesaffre, 1999). Moreover, the results are comparable to those obtained from a longitudinal study design (see, e.g., Duncan et al., 1996). Here, the cross-sequential design allows us to describe the development of prosody across most of the primary school years; specifically, we
provide data that captures prosody production in narrative speech from age 5 years to age 11 years.

Methods

Speech Material

The speech samples analyzed in the present study were drawn from the Eugene Children’s Story Corpus (ECSC; Kallay & Redford, 2021), which is publicly available though the CHILDES database (MacWhinney, 2000). The corpus includes 367 audio recordings and transcriptions of structured spontaneous narratives elicited from a total of 188 typically-developing, English-speaking school-aged children and 26 adults (i.e., accompanying caregivers). Narratives were elicited using the well-known frog story picture books (Mayer, 1967, 1969, 1973; Mayer & Mayer, 1975). The narrative storytelling sessions began with the experimenter presenting the 4 books to the child and a caregiver. Each then selected a book to narrate. A small number of the children (N = 4) chose the same book across all 3 study years. A far larger number (N = 20) chose a different book to narrate in each of the study years. The majority (N = 36) happened to choose the same book in 2 of the 3 years. The experimenter assisted the child in conceptualizing a coherent narrative around the pictures in their chosen book by asking questions about predetermined events on the pages. Each participant then told their story twice to each other in the presence of the experimenter, alternating between child and caregiver as storyteller or vice versa. The tellings themselves were free from prompts. The second telling was different than the first, but the telling itself benefited from practice (e.g., pause durations were shorter; Redford, 2013). It is the 2nd telling of each story that has been made publicly available, and so it is also the 2nd telling upon which the current description of prosody is based.
Participants

The ECSC speakers were all recruited from the Eugene-Springfield area in Oregon, USA via word of mouth. A subset of 71 children (32 male) were followed longitudinally over a 3-year period. Those children returned to the lab at 12-month intervals within 2 weeks of the date of their previous visit. The study was approved by the institutional review board at the University of Oregon and informed consent was obtained from all speakers prior to participation. The current study is based on the 60 youngest children (25 male, 35 female) from the subset of children in the longitudinal study who were in kindergarten through 2nd grade in the first study year. The sample of 60 children allowed for an even split of the data into three non-overlapping starting age cohorts: youngest (Y), in-between (B) and oldest (O). The children in the Y cohort (12 male, 8 female) ranged in age from 5;2 to 6;3 in the first study year ($M = 5;8$, $SD = 4$ months); those in the B cohort (6 male, 14 female) ranged in age from 6;4 to 7;3 in the first study year ($M = 6;9$, $SD = 3$ months); and those in the O cohort (7 male, 13 female) ranged in age from 7;3 to 8;1 in the first study year ($M = 7;7$, $SD = 3$ months). There were two children in the sample who were aged 7;3 in the first study year; one was randomly assigned to the B cohort and one to the O cohort for the analyses.

The majority of the 60 children were identified by caregivers as White only ($N = 45$). The remaining children were identified as Asian ($N = 2$), Pacific Islander ($N = 2$), Hispanic ($N = 2$), or mixed ($N = 9$). Typical language development was reported by all caregivers and was confirmed using the Peabody Picture Vocabulary Test (PPVT-4; Dunn & Dunn, 2007) and two subtests from the Clinical Evaluation of Language Fundamentals (CELF-4; Semel et al., 2003): Recalling Sentences (language production) and Sentence Structure (language comprehension). All of the children’s scores on these tests were no more than 1 standard deviation from the mean.
standard score (PPVT: $M = 118.83$, $SD = 11.45$; RS: $M = 11.83$, $SD = 2.72$; SS: $M = 11.91$, $SD = 2.19$).

**Measurement**

**Rhythm measures**

The 180 audio recordings from the 60 children in the longitudinal study were suitable for calculating articulation rate and for prosodic transcription. The stories were first segmented into pause-delimited utterances following the criteria laid out in Redford (2013) and orthographically transcribed using Praat software (Boersma & Weenink, 2018). Articulation rate was defined as the number of syllables per second in speech without pauses. This measure was extracted with a Praat script for each narrative based on the pause-delimited transcriptions. Specifically, it was calculated as the total number of syllables produced divided by the total speaking time in seconds, minus pauses.

The other rhythm measures were based on perceptual judgments of prominences and boundaries. These judgments were made in the context of prosodic coding, which was based on 30 second samples from each narrative. The samples were taken from roughly the middle of each narrative specifically to exclude the stereotyped prosody associated with story beginning and endings. The samples were also chosen so that they would be representative of the child’s fluent speech and free from non-speech interruptions (e.g., coughs) or from interruptions by an interlocuter (i.e., parent or experimenter). The onsets of the samples were chosen to correspond with a strong phrase boundary and the offsets were selected to align with the nearest sentence boundary to the 30 second point. The choice to use shorter samples of the full narrative recordings was motivated by the time-consuming nature of hand-coding prosodic features.
A team of trained analysts marked off all syllables in the samples and then used the Rhythm and Pitch (RaP) system (Dilley & Brown, 2005) to label the locations and types of perceptual prominences and prosodic phrase boundaries in the speech samples. The team of analysts was comprised of undergraduate and graduate students from the department of Communicative Sciences and Disorders at Michigan State University (MSU), who received in-lab training in RaP after participating in a seminar on prosody and prosodic transcription. Prominences were marked as strong or weak or absent for each syllable. Phrase boundaries were also marked as strong or weak or absent. In linguistic theory, strong and weak prominences and strong and weak phrase boundaries are often thought to reflect different types of prosodic units (see, e.g., Ladd, 2008). For example, a strong prominence marks cumulative prominence due to phrase- and lexical-level stresses; a strong boundary marks an intonational phrase (IP) and a weak boundary marks an intermediate phrase (ip). In the present analyses, any syllable marked as prominent was treated as a strong syllable; similarly, we made no distinction between weak and strong boundary judgments. In the case of prominences, collapsing judgments into a single “strong” category aligned with our interest in speech rhythm, which is assumed to arise from the alternation of stressed and unstressed syllables at the level of the metrical foot—not at the level of the intonational phrase. In the case of boundaries, the strong and weak categories were collapsed because there is little agreement on what type of unit the intermediate phrase might be or on how to define it other than with reference to boundary strength. Additionally, some have suggested that the distinction between strong and weak phrase boundaries is not relevant from a psycholinguistic perspective (Choe & Redford, 2012; Watson & Gibson, 2004). The kappa (\(\kappa\)) metric was used to assess reliability across all prosodic transcriptions. This metric controls for chance agreement between analysts (McHugh, 2012). There was moderate to
substantial inter-rater agreement across the different prominence types ($K = .60$) and boundary types ($\kappa = .62$). As an additional check, acoustic measures of duration were taken on a subsample of the data. These measures confirmed a clear acoustic contrast between syllables that were judged prominent versus those that were not prominent [$F(1,625) = 529.5, p < .001; M = 345$ msec. v. $M = 199$ msec.] – a contrast that is consistent with the reported correlation between duration and perceived prominence in stress-timed languages like English (Kochanski et al., 2005; Sluijter et al., 1997). Third-party inspection of the acoustics associated with labeled boundary data were also conducted. This inspection indicated that perceived strong boundaries (nearly) always preceded a labeled silent pause. Strong boundaries represented 80% of the total boundaries coded. The additional acoustically-based checks on coding were undertaken at the University of Oregon.

Once the samples had been prosodically labelled, the dependent measures of prominence spacing, mean phrase length, mean phrase duration, and phrase length variability were calculated as follows: prominence spacing was the total number of syllables across the entire speech sample divided by the total number of perceived prominences; mean phrase length was average number of syllables between intonational phrase boundaries; mean phrase duration was the average duration in seconds between boundaries, exclusive of pauses; variability in phrase length was the standard deviation of the phrase lengths in syllables within the sample.

**Intonation measures**

Due to background noise in some recordings or tangential discussions between experimenter and child or between caregiver and child, a reduced sample of materials was used for calculating mean F0 variation and range from across entire narratives (see below). In order to maintain the benefits of a cross-sequential design, we included only those children with recordings that were
of consistently high quality across all three study years. This strict inclusion criteria resulted in a sample of 108 narratives from 36 children (16 male, 20 female) who ranged in age from 5;2 to 7;9 in the first study year. As in the larger sample, the majority of these children were also identified by caregivers as being White only ($N = 25$). The children were evenly split into cohorts of 12, with children in the Y cohort (6 male, 6 female) ranging in age from 5;2 to 6;2 in the first study year ($M = 5;8$, $SD = 4$ months), children in the B cohort (4 male, 8 female) ranging in age from 6;2 to 6;11 in the first study year ($M = 6;7$, $SD = 3$ months), and children in the O cohort (6 male, 6 female) ranging in age from 7;1 to 7;9 in the first study year ($M = 7;5$, $SD = 2$ months). Two of the children were aged 6;2 in the first study year; one was randomly assigned to the Y cohort and the other to the B cohort.

F0 values were extracted automatically over 10 millisecond intervals across the sonorant-only portions of the utterances in each narrative to avoid spurious F0 values associated with changes in voicing and with frication. Sonorant intervals were identified within the utterances using the vuv (voiced/unvoiced) textgrid function in Praat (Boersma and Weenink, 2018). To further avoid spurious values in the automatic extraction of F0, outliers were removed according to the 1.5 IQR rule prior to calculating each of the intonation measures. Once this was done, intonational variability was calculated in two ways: as the standard deviation of F0 within narratives and as the coefficient of variation of F0 within narratives. The overall standard deviation of F0 has been widely used in previous studies to investigate differences in intonation variability between typical and atypical populations (e.g., Diehl & Paul, 2012; Diehl et al., 2009). The coefficient of variability (= standard deviation of F0 divided by mean F0) was calculated to control for a developmental- or cohort-based effect of mean F0 on the standard deviation in F0.
Prosodic development was also calculated for each narrative. This measure was simply the minimum F0 value in the narrative subtracted from the maximum F0 value in the same narrative.

**Statistical Analyses**

**Cross-sequential analyses**

Linear mixed-effects models were constructed for each measure of rhythm and intonation discussed above to test for cross-sectional and longitudinal effects that would indicate developmental change. The cross-sectional effect was Age Cohort, a between-subjects factor with 3 levels (Y, B, O). The longitudinal effect was Study Year, a within-subjects factor with 3 levels (YR1, YR2, YR3). Models were built using the lme4 package in R software (Bates et al., 2015; R Core Team, 2021). Full models included both Age Cohort and Study Year as fixed main effects and Speaker as a random effect. For each fixed effect, null models were constructed by removing the effect of interest from the full model. The full model was then compared to the null model with ANOVA to assess significance of the removed fixed effects (see Brown, 2021; Winter, 2013). The $\chi^2$-statistic and p-value are reported for all significant differences between the models. Pairwise comparisons were also performed within Age Cohort and Study Year using the same linear mixed effects models with Bonferroni corrections.

**Additional analyses**

In addition to the cross-sectional analyses, simple linear regression models were built to assess the continuous effect of Age-in-Months on the rhythm and intonation measures. Separate models were constructed for each measure within each Study Year; Age-in-Months was the lone predictor variable. The $t$-statistic and p-value are reported for all significant effects. Linear mixed-effects models were not used because of the strong relationship between Age-in-Months and speaker in
Prosodic development

the present study \([H(59) = 179, p < .001]\); that is, the fixed effect of Age-in-Months and the random effect of speaker were virtually indistinguishable.

Results

The mean and standard deviation values for each prosodic measure are presented by Age Cohort and Study Year in Table 1. The patterns evident in the summary data are consistent with the prediction that global rhythm measures would increase with development. The patterns do not clearly support the predicted developmental decrease in global intonation measures. The detailed cross-sequential analyses confirm this general conclusion: articulation rate, prominence spacing, mean phrase length, and variability in phrase length all increase with age; pitch variability and pitch range do not. The detailed results are presented next.
Prosodic development

1. **Table 1.** Mean (SD) values for each of the global rhythm and intonation measures by Age Cohort (the cross-sectional factor) and Study Year (the longitudinal factor). Measurement units indicated in parentheses (syll = syllables, sec = seconds, SD = standard deviation, CV = coefficient of variation, Hz = Hertz, F0 = fundamental frequency).

<table>
<thead>
<tr>
<th>Rhythm</th>
<th>Y</th>
<th>B</th>
<th>O</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Year 1</td>
<td>Year 2</td>
<td>Year 3</td>
</tr>
<tr>
<td>Articulation Rate (syll/sec)</td>
<td>3.04</td>
<td>3.37</td>
<td>3.41</td>
</tr>
<tr>
<td></td>
<td>(0.45)</td>
<td>(0.43)</td>
<td>(0.41)</td>
</tr>
<tr>
<td>Prominence Spacing (syll)</td>
<td>1.72</td>
<td>1.83</td>
<td>1.83</td>
</tr>
<tr>
<td></td>
<td>(0.23)</td>
<td>(0.19)</td>
<td>(0.24)</td>
</tr>
<tr>
<td>Mean Phrase Length (syll)</td>
<td>3.29</td>
<td>3.79</td>
<td>3.66</td>
</tr>
<tr>
<td></td>
<td>(0.92)</td>
<td>(1.08)</td>
<td>(0.89)</td>
</tr>
<tr>
<td>Mean Phrase Duration (sec)</td>
<td>1.08</td>
<td>1.13</td>
<td>1.07</td>
</tr>
<tr>
<td></td>
<td>(0.23)</td>
<td>(0.32)</td>
<td>(0.24)</td>
</tr>
<tr>
<td>Phrase Length Variability (syll)</td>
<td>2.23</td>
<td>2.30</td>
<td>2.29</td>
</tr>
<tr>
<td></td>
<td>(0.90)</td>
<td>(0.77)</td>
<td>(0.65)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Intonation</th>
<th>Y</th>
<th>B</th>
<th>O</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Year 1</td>
<td>Year 2</td>
<td>Year 3</td>
</tr>
<tr>
<td>SD of F0 (Hz)</td>
<td>53.03</td>
<td>58.22</td>
<td>52.13</td>
</tr>
<tr>
<td>CV of F0</td>
<td>0.18</td>
<td>0.20</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td>(0.08)</td>
<td>(0.07)</td>
</tr>
<tr>
<td>Range of F0 (Hz)</td>
<td>500</td>
<td>571.26</td>
<td>538.36</td>
</tr>
<tr>
<td></td>
<td>(88.33)</td>
<td>(21.10)</td>
<td>(79.97)</td>
</tr>
</tbody>
</table>
Prosodic development

Rhythm

Articulation Rate

Results of the full model indicated significant main effects of Age Cohort \( \chi^2 = 14.51, p < .001 \) and Study Year \( \chi^2 = 45.32, p < .001 \) on mean articulation rate. When the data were split by Study Year, pairwise comparisons indicated significant differences between each year across age-based cohorts of children: Year 1 v. Year 2 \( \chi^2 = 15.71, p < .001 \); Year 1 v. Year 3 \( \chi^2 = 41.41, p < .001 \); Year 2 v. Year 3 \( \chi^2 = 9.18, p < .01 \), reflecting systematic increases in articulation rates across developmental time. Pairwise comparisons indicated that the longitudinal effect was significant within each of the three age cohorts [Y cohort: \( \chi^2 = 9.78, p < .01 \); B cohort: \( \chi^2 = 21.36, p < .001 \); O cohort: \( \chi^2 = 15.33, p < .001 \)]. Figure 1 shows these results.

Figure 1. Longitudinal effect of Study Year on articulation rate is shown by the cross-sectional effect of Age Cohort (Y = youngest, B = in-between, O = oldest).

Cohort effects indicate that the different starting ages of children had an effect on articulation rate—an effect that was independent of (though presumably compounded by) longitudinal development. Recall that children in the Y cohort were 5 years old in the first year of study, those in the B cohort were 6 years old, and those in the O cohort were 7 years old. The effect of Age Cohort indicates that this cross-sectional age group difference mattered. Over the 3 study
Prosodic development

years, the Y cohort produced narratives with a slower articulation rate than the B cohort [$\chi^2 = 14.74, p < .001$]; their articulation rate in narrative speech was also slowered overall all than that produced by the O cohort [$\chi^2 = 7.15, p < .01$]. Related to this cross-sectional effect on articulation rate, there was a significant linear increase in articulation rate by children’s age within a year. The continuous between-subjects effect of Age-in-Months was significant within Year 1 [$t(58) = 2.40, p = .02$] and Year 3 [$t(58) = 3.51, p < .001$], but not within Year 2 of the study. These results are shown in Figure 2, where the most noticeable increase in articulation rate occurred in children who were between the ages of 86 and 121 months in Year 3 of the study.

Figure 2. The continuous between-subjects effect of Age-in-Months on articulation rate is shown by Study Year.

Prominence Spacing

Results of the mixed effects model indicated a significant main effect of Study Year on prominence spacing [$\chi^2 = 13.81, p < .01$], but there was no overall effect of Age Cohort. The longitudinal effect was due to a significant difference between Year 1 and Year 2 values [$\chi^2 = 7.89, p = .01$] and between Year 1 and Year 3 values [$\chi^2 = 11.59, p < .001$]. The difference between Year 2 and Year 3 values was not significant. When the data were split by Age Cohort, analyses indicated that the effect of Study Year only reached significance in the B cohort [$\chi^2 = 18$]
Prosodic development

16.00, $p < .001$. Still, the same longitudinal trend is evident in the narrative produced by the Y cohort. Figure 3 shows the longitudinal increase in prominence spacing by the cross-sectional age groups.

Figure 3. The longitudinal effect of Study Year on prominence spacing is shown by the cross-sectional effect of Age Cohort (Y = youngest, B = in-between, O = oldest).

Consistent with the absence of a cross-sectional effect of age group on the development of prominence spacing, the continuous between-subjects effect of Age-in-Months was not significant either. Figure 4 nonetheless shows the hint of a cross-sectional pattern in Year 1 when children ranged in age from 5 years old (62 months) to 8 years old (97 months).

Figure 4. The continuous between-subjects effect of Age-in-Months on prominence spacing is shown by Study Year.
Prosodic development

**Phrasing**

Results of the mixed effects model indicated a significant overall main effect of Study Year on phrase length [$\chi^2 = 13.10, p < .001$]. Again, there was no effect of Age Cohort. When the analyses were split by the cross-sectional factor, the longitudinal effect only reached significance for B cohort [$\chi^2 = 12.58, p < .001$]. Though not significant, mean phrase length also increased by Study Year within the Y and O cohorts (see Figure 5). As with prominence spacing, individual variability was such that there were no significant linear effects of Age-in-Months on mean phrase length within Study Year (Figure 6).

![Figure 5](image.png)

**Figure 5.** The longitudinal effect of Study Year on phrase length is shown by the cross-sectional effect of Age Cohort (Y = youngest, B = in-between, O = oldest).
Prosodic development

Figure 6. The continuous between-subjects effect of Age-in-Months on mean phrase length is shown by Study Year.

Unlike for phrase length, there were no main effects of children’s age on phrase duration. Phrases were roughly one second in duration ($M = 1.1$ seconds; $SD = 290$ msec) across longitudinal time and cross-sectional age groups. The minimum average phrase duration was 617 milliseconds; the maximum duration was 2.69 seconds.

In contrast to the results on phrase length and phrase duration, the mixed effects model on variability in phrase length indicated a significant main effect of Age Cohort [$\chi^2 = 8.15, p = .02$], but no overall effect of Study Year. Pairwise comparisons between the different cross-sectional age groups indicated that phrase length variability was higher in the B cohort than in the Y cohort [$F(1,116) = 5.02, p = .03$]; it was also higher in the O cohort than in the Y cohort [$F(1,116) = 9.39, p < .01$]. The difference between the B and O cohorts were not statistically significant. The results are shown in Figure 7.
Prosodic development

Figure 7. The longitudinal effect of Study Year on variability in phrase length is shown by Age Cohort (Y = youngest, B = in-between, O = oldest).

The cross-sectional effect of age was confirmed in the analysis by Age-in-Months. There was a significant linear increases in variability with age in Year 1 \( t(58) = 2.11, p = .039 \) and Year 3 \( t(58) = 2.02, p = .048 \). The effect did not reach significance in Year 2. Figure 8 shows phrase length variability by the continuous between-subjects age variable for each Study Year.

Figure 8. The continuous between-subjects effect of Age-in-Months on variability in phrase length is shown by Study Year.
Prosodic development

1 **Intonation**

2 **Pitch Variability**

3 Results of the mixed effects model indicated no main effects of Age Cohort or Study Year on pitch variability, as measured by the standard deviation of F0 values for each narrative. Figures 9 and 10 show the distribution of the data by the fixed factors. Figure 10 shows a large amount of individual variability in pitch variability across the 60 narratives within each study year. As is evident from that figure, there was no significant effect of Age-in-Months on pitch variability either.

![Diagram](image)

9 **Figure 9.** The longitudinal effect of Study Year on pitch variability, measured as the standard deviation in F0 values, is shown by the cross-section effect of Age Cohort (Y = youngest, B = in-between, O = oldest).
Figure 10. The continuous between-subjects effect of Age-in-Months on pitch variability, measured as the standard deviation of F0, is shown by Study Year.

Similar results were obtained when the measure of pitch variability, controlled for mean F0 (= CV of F0), despite a significant cross-sectional effect of Age Cohort on this measure [$\chi^2 = 6.37, p = .04$]; the effect of Study Year was not significant. The significant cross-sectional effect was due to higher pitch variability in narrative elicited from children in the B cohort compared to those elicited from children in either the Y or O cohorts. Given that children in the B cohort were older than those in the Y cohort and younger than those in the O cohort, the difference is not what one would expect if the results were age related. The analysis by Age-in-Months confirms the absence of a developmental effect. The distribution of the CV of F0 measure is shown by the fixed factors in Figure 11 and by Age-in-Months in Figure 12.
Prosodic development

Figure 1. The longitudinal effect of Study Year on pitch variability, measures as the coefficient of variation in F0 values, is shown by the cross-sectional effect of Age Cohort (Y = youngest, B = in-between, O = oldest).

Figure 2. The continuous between-subjects effect of Age-in-Months on pitch variability, measured as the coefficient of variation of F0, is shown by Study Year.

Pitch Range

There were also no significant main effects of Age Cohort or Study Year on pitch range. The (null) results are shown in Figure 13.
Prosodic development

Figure 13. The longitudinal effect of Study Year on pitch range, measured as the average difference between max and min F0, is shown by the cross-sectional effect of Age Cohort (Y = youngest, B = in-between, O = oldest).

As expected from the cross-sequential analyses, the effect of Age-in-Months on pitch range was not significant. But, unlike with the measures of pitch variability, the null effect seems to stem from a general lack of individual variability rather than to excessive variability (see Figure 14).

Figure 14. The continuous between-subject effect of Age-in-Months on pitch range, measured as the difference between max and min F0, is shown by Study Year.
Prosodic development

Discussion

The aim of the current study was to provide a clinically-useful characterization of the typical age-related changes in speech prosody that occur across the early school years. Our approach was to calculate global measures of rhythm and intonation based on narrative speech samples elicited from a cross-sectional sample of children, who were followed longitudinally for 3 years. The focus on narrative speech allowed us to capture aspects of prosody that have been previously studied, such as the development of temporal patterning, as well as those aspects that are poorly studied, such as the development of rhythmic phrasing. In addition, the narrative speech task was expected to elicit language that was sufficiently complex to make manifest developmental patterns that emerge with the maturation of speech motor skills and the development of higher-level language skills. The study predictions were that (1) measures of temporal patterning and prominence spacing (i.e., “stress-timing”) would increase with age due to the development of speech motor skills; (2) measures of pitch variability and pitch range (i.e., global intonation) would decrease with age for the same reason; and that (3) measures of rhythmic phrasing would increase with age due to the development of higher-level language skills. The cross-sequential analyses confirmed the predictions that temporal patterning and prominence spacing increase with age and that prosodic phrases become longer and more variable. In contrast to the developmental changes in global rhythm, pitch variability and pitch range showed no systematic effects of development. Below, we discuss these results in the context of developing speech motor and language skills – beginning with the absence of developmental effects on the global measures of intonation.
The meaning of pitch variability and pitch range

The global measures of intonation used in the present study were pitch variability and pitch range. These measures were chosen for their clinical value: they are easy to calculate – especially when compared to a phonological analysis of intonation; and, they have been used extensively in previous research on prosody production in clinical populations – especially in studies of children with ASD (e.g., Bone et al., 2014; Bonneh et al., 2011; Fosnot & Jun, 1999), but also in studies of children with stuttering (Fosnot & Jun 1999), in children who have received cochlear implants (Nakata et al., 2012), and in adults with schizotypal personality disorder (Dickey et al., 2012). The prediction was that pitch variability and pitch range would decrease with developmental time. This prediction was based on a finding from a large-scale acoustic study on typical development of articulatory timing skills, which showed that pitch variability decreases dramatically from age 5 years to 8 years of age (Lee et al., 1999). It was also based on a particular interpretation of the well-established finding that higher-than-average pitch variability and range indexes atypical prosody in children with ASD; namely, that this finding is due to prosodic delay rather than disorder (see e.g., Diehl et al., 2009:398).

Contrary to the prediction, we found no systematic changes in pitch variability or pitch range across study year, age cohort, or by age-in-months within study year. Instead, we found extensive individual differences in pitch variability (see e.g., Figure 10), with measures ranging from a low of 11 Hz differences across the sonorant portions of the narrative to a high of 66 Hz. We interpret these finding to indicate that the F0-based measures of pitch variability and pitch range are insufficiently sensitive to age-related changes in intonation during the school years. This interpretation is motivated by our firm expectation that all aspects of speech prosody continue to develop through the school-aged years with the maturation of speech motor skills and
the development of language. This expectation is motivated by the Lee et al. (1999) findings that indicate the slow maturation of motor skills relevant for F0 manipulation and by other studies that document immature intonational patterns for some aspects of discourse prosody in young school-aged children’s speech (e.g., Chen, 2009; Chen, 2011; Romøren & Chen, 2014). The latter study findings are also consistent with the observation that discourse skills themselves continue to develop into adolescence (Nippold, 2016). Relatedly, Zampini and colleagues (2016) found that while children with Down syndrome produce largely typical phrase-level intonation patterns, they struggle specifically with production of interrogative contours relative to their typically developing peers. Children with Down syndrome also produce shorter and less syntactically complex phrases than their typically developing peers (ibid), which again suggests that influence of language development on intonation production.

Since the specific data presented herein are meant to provide developmental norms against which other data can be compared, it is worth pointing out that pitch variability values in the present study (i.e., SD of F0) were substantially lower on average than pitch variability values reported in Diehl et al.’s (2009) study on narrative speech, and substantially higher on average than the values reported for typically developing 5-year-olds in the Lee et al. (1999) study. Our mean SD F0 value was 37 Hz; Diehl et al.’s was approximately 48 Hz in narratives elicited from children aged 6 to 14 years (42 Hz for TD males and 53 Hz for TD females; Tables 2-3, p. 392-93); Lee et al. report a mean of 23 Hz for 5-year-old speech (Figure 3, p.1460). Although it is possible that Diehl et al.’s participants were more animated in their storytelling than our participants, we suspect that the difference in values reflects a difference in the way in which the F0 measurements were made. Diehl et al. extracted average F0 across 250 millisecond intervals of their narratives, yielding 4 measures per second. Pitch variability was the standard deviation
Prosodic development

of these measures. Since the methods say nothing about pitch tracking correction or a focus on
sonorant-only intervals, it is possible that their larger values reflect the inclusion of spurious F0
values due to voicing changes at obstruent-sonorant edges or simple mistracking. As spurious
values are likely to be randomly distributed across groups, their inclusion does not impact their
study results. Still, we are suggesting that the values we report here may more accurately reflect
the perceptual experience of pitch variability. Either way, the present study results can be used as
yet another baseline against which to compare the higher pitch variability and pitch range values
associated with ASD and prosodic disorder in the literature.

Turning now to the difference between the values reported here and those reported in Lee et
al. (1999), we expect that our larger SD F0 values relative to that study are meaningfully related
to task differences. Lee et al. extracted their measurements from words uttered in isolation; we
extracted our measurements from across an entire narrative. It is intuitively plausible that pitch
variability will be higher in narratives compared to isolated words because speakers deploy the
full prosodic system during narrative language. By contrast, isolated words that are monosyllabic
only provide information on the extent of F0 change during the pitch accenting that might
accompany an in-focus production of said monosyllabic word. It is also likely that task
differences and sample sizes account for why we find no developmental change in pitch
variability while Lee et al. report a significant change. Note that the change from 5-year-old SD
F0 values to 8-year-old SD F0 values in the Lee et al. study is of about 8 Hz. By contrast, our
range in SD F0 values across ages was on the order of 55 Hz. It is easy to see how this level of
individual difference in pitch variability associated with the full prosodic spectrum of storytelling
might swamp any evidence of a developmental decrease in pitch variability due to the
development of motor control over pitch and age-related changes in vocal fold physiology.
Prosodic development

1. Future normative work on pitch variability in the context of language production (as opposed to single word production) will need to be based on much larger sample sizes if the goal is to detect differences due to changes in motor control and vocal tract physiology.

2. **Age-related change in global rhythm**

   Turning now to the results on rhythm, our findings indicate gradual age-related changes in articulation rate, prominence spacing, and phrase length across study years as well as a gradual age-related change in phrase length variability. These results stand in contrast to the more dramatic nonlinear changes that are frequently reported in the developmental speech-language literature (e.g., articulatory timing: Lee et al., 1999; phonological development: Burchinal & Appelbaum, 1991; vocabulary growth: Ganger & Brent, 2004). We suspect that the more gradual changes in speech rhythm across the early school-aged years compared to other types of speech-language change reflects the many different influences on speech rhythm, under the assumption that these influences hold to different developmental schedules that vary across children.

   Rhythm emerges at the intersection of speech motor and linguistic systems, which is to say that the effects of both are evident in its structure (see “Background” to the present study).

   Despite the known interaction between the speech motor and linguistic systems, it is reasonable to treat them as quasi-independent systems, with the former intimately tied to perceptual-motor processes and the latter to domain-specific and domain-general cognitive processes. Similarly, subsystems within the linguistic system may also be quasi-independent: some are well-described as domain-specific; others clearly depend on domain-general processes. Consider the case of syntax versus pragmatics: both are bound to semantics, but syntax is much more domain-specific than pragmatics. In particular, syntax is the learned, language-specific sequencing of meaningful elements – the sequencing patterns may have their origins in lexical semantics, but once fixed by
Prosodic development

historical forces the particular language-specific sequencing patterns are arbitrary from a
synchronic perspective. In contrast, despite language-specific (i.e., formal) elements to
pragmatics, it is clear from both typical and atypical development that pragmatic learning relies
very heavily on the ability to track conceptual relations at a high level and on social cognition.
Traces of all of these quasi-independent systems are seen in the production of speech rhythm.
We speculate that differences in the rate of development across these systems accounts for the
extensive individual differences, gradual developmental change, and conflicting results obtained
in the analyses of global rhythm measures.

Consider, for example, the results on phrasing. At first glance, the systematic longitudinal
increases that were observed for phrase length (measured in syllables) would seem to be at odds
with the finding that phrase duration (measured in milliseconds) remains stable across
developmental time, especially when one considers that phrase length and phrase duration were
strongly correlated in the present data \[r(179) = .853, p < .001\]. But, in fact, the conflicting
results are consistent with the combined influence of language and speech factors on the
emergence of rhythm.

In particular, developmental changes in prosodic phrase length must be at least partially due
to the development of more complex language. For instance, a richer vocabulary and the ability
to deploy this vocabulary to render more complex thoughts will likely result in longer syntactic
phrases on average (e.g., the more frequent use of adjectives and adverbs to modify a noun or
verb). Given the relationship between strong prosodic boundaries and strong syntactic
boundaries (Turk & Shattuck-Hufnagel, 1996), an increase on clause-internal syntactic phrase
lengths will result in longer prosodic phrases. Support for this hypothesis comes from the
Prosodic development

correlation between mean clause length and pause-delimited utterance length in narrative speech (e.g., Kallay & Redford, 2021).

Of course, the length of a prosodic phrase in syllables should correlate perfectly with its length in duration. But this is only true if articulation rate is held constant. If instead articulation rate increases with age, then the relationship between phrase length and phrase duration will be weaker. And, in fact, it is an increase in articulation rate that accounts for the divergent results on phrase length and phrase duration in the present study: phrase length and articulation rate explain 96% of the variance in phrase duration \( [F(2,177) = 2321; p < .001; R^2 = .963] \). The combined effect of phrase length and articulation rate on phrase duration is also consistent with the positive relationship between length and rate that has been reported elsewhere (Mahr et al., 2021). In so far as phrase length indexes language development and mean articulation rate indexes speech motor skill in children’s speech, the absence of an effect of development on phrase duration despite the presence of an effect of development on phrase length exemplifies the combined influence of motor and language factors on the emergence of speech rhythm. Again, we also assume that these factors may develop at different rates and that there is extensive individual difference in the relative rate of development of these factors.

The interaction between language and speech motor factors on rhythm is noted elsewhere in the literature with reference to the acquisition of different languages. For example, Payne et al. (2011) argued that the delayed acquisition of a stress-timed language, like English, relative to syllable-timed languages, like Spanish and Catalan, has to do with the different phonotactics of the languages. Following Dauer (1983), they adopted the view that a stress-timed rhythm emerges in languages that allow both complex syllable structures and unstressed syllable reduction and that a syllable-timed rhythm emerges in languages that have simpler syllable
Prosodic development

structures and full vowels. In so far as children with immature speech motor skills are more likely to accurately produce simple syllables than complex syllables, they are also more likely to produce adult-like rhythm earlier in syllable-timed languages, like Spanish, than in stress-timed languages, like English. Relatedly, typically developing children learning Italian, a syllable-timed language, are mostly adult-like in their productions of contrastive stress as early as 3 years of age (Arciuli & Colombo, 2016); similar measures applied to children learning English, suggest a much more protracted developmental course (Arciuli & Ballard, 2017). The point is that rhythm emerges in interaction with the phonological patterning of a language. This observation suggests that yet another factor to consider when evaluating the longitudinal rhythm structure a child’s language, namely, the interaction with phonological development.

**Longitudinal versus cross-sectional age effects**

Although longitudinal studies have a privileged status in the developmental literature, it is not entirely clear that all longitudinal data should be treated as developmental data. Children may simply get better at certain assessment or experimental tasks with repeated exposure to the task itself. This type of learning is interesting and important, but it does not necessarily provide the developmental insights we seek when studying age-related speech and language changes. Relatedly, it is common knowledge that cross-sectional effects can emerge from uncontrolled (i.e., non-developmental) differences in population samples. These differences are known as cohort effects. Of course, the knowledge that cohort effects are always possible in a cross-sectional study rarely impacts the interpretation of cross-sectional age effects. Instead, cross-sectional effects are nearly always interpreted as developmental effects when an interest in development motivates the study. The validity of a developmental interpretation of cross-sectional age effects on speech-language variables is strengthened with replication and with
Prosodic development

longitudinal data. An important benefit of the cross-sequential design used in the present study is that it allows for both of these. Age cohort effects were subject to quasi-replication by examining the within year effect of Age-in-Months on the variables of interest. A developmental interpretation of one or both of these cross-sectional age effects was further strengthened by the longitudinal aspect of the study. The strongest evidence for development is provided when the cross-sectional (between-subject) effects and the longitudinal (within-subject) effects align within the data set. With this in mind, we make some additional observations about the present study findings.

The present study results provide very strong evidence for developmental changes in the articulation rate of narrative speech across the early school-aged years, with the largest increases occurring between the ages of 7 and 9 years old (cf. Mahr et al. (2021) who report faster rates of change between the ages of 2 and 5 years and more gradual change thereafter until age 9 years). The results provide weaker evidence for significant developmental change in prominence spacing. In particular, the longitudinal effect was significant overall and within the in-between age cohort (children who were 6 years old on average at the start of the study), but the between-subject effects were not significant despite some evidence of age-related differences in the Year 1 data (i.e., between 62 and 97 months in age). A similar pattern of results was observed for phrase length. The longitudinal effect was significant, but the cross-sectional age effect was not and individual variability within each study year was substantial enough to obscure an effect of age-in-months in the Year 1 study data, despite some suggestion in the data of an age-related increase in phrase length. The combined results were stronger for the measure of variability in phrase lengths. Although the overall analyses indicated only an effect of age cohort, the within age cohort analyses indicated longitudinal increases in variability in the older cohorts. The effect
Prosodic development

was strongest in children who were 7 years old on average at the start of the 3-year study. Also, there were linear increases in the variability of phrase lengths within a narrative across Age-in-Months within the Year 1 and Year 3 data; that is, variability in the timing of prosodic boundaries increased between the ages of 5;2 and 8;1 (years;months) and then again between the ages of 7;2 and 10;1.

In the context of the preceding discussion regarding the multifactorial nature of speech rhythm, we take the stronger developmental effects on articulation rate and the internally replicated effect of age-related change on variability in phrase length to suggest that these measures may index a narrower range of specific influences on rhythm production than the other measures. For example, the across-the-board effects on articulation rate are consistent with previous findings, many of which argue that this measure provides an index of speech motor skill (Lee et al., 1999; Redford, 2014). Accordingly, we interpret this finding to indicate that articulation rate represents a reasonably pure index of the influence of speech motor development on rhythm production; which is to say that it does not index the influence of language development on rhythm production to the same extent. By contrast, we interpret the significant, but weaker evidence for developmental change in prominence spacing to suggest that this measure indexes both speech motor and language influences on rhythm production in a more equal fashion: a child’s ability to reduce weak syllables to the same extent as the adult relies on the development of speech motor skills (Redford, 2014; 2018); but the number of weak syllables that are reduced in a sample also depends on language complexity. In particular, the production of more disyllabic words and more prepositional phrases in a sample will result in more opportunities for weak syllable reduction in that sample. A greater number of weak syllables in
Prosodic development

1. the sample will result in a developmental increase in prominence spacing, assuming that the child is able to appropriately reduce said syllables.

2. By a similar logic, the stronger evidence for a developmental change in phrase length variability compared to phrase length itself suggests that variability in phrase length may be the purer index of language development – again, under the assumption that strong prosodic boundaries are usually aligned with strong syntactic boundaries (i.e., clause boundaries) in fluent spontaneous speech. It was suggested that the variable timing of phrase boundaries will increase with the greater diversity of expression that follows from the development of language skills. This suggestion is really a hypothesis. The assumed relationship between phrasing and diversity of expression is understudied. For that matter, given the clinical value of global prosodic measures, the underlying reasons for developmental changes in prosodically-related variability of any kind merits further research.

3. **Future research**

4. Regarding the need to better understand the underlying factors that contribute to developmental changes in measures of prosodic variability, we make the more general observation that global measures of speech prosody provide incomplete information about prosodic development. This is because the measures separate rhythm and intonation from the linguistic and semantic-pragmatic functions they serve. We interpreted the null results on pitch variability and pitch range as an instance of this limitation, but it is also relevant to the discussion of rhythm. For example, we implied that prominence spacing may be more dense in younger children’s speech than in older children’s speech not simply because younger children are unable to reduce weak syllables to the same extent as older children, but also because younger children favor monosyllabic words and simpler syntactic constructions, including – for example – fewer
Prosodic development

prepositional phrases. This possibility points to the importance of future research that seeks to better understand exactly what global measures of prosody are actually measuring. The clinical implication is also clear: a thorough prosodic assessment must also include an analysis of the language that is being produced. Absent further research into the underlying reasons for developmental increases or decrease to global rhythm and intonation measures, we will lack insight into the origin of deviant rhythm or intonation in development. Relatedly, prosodic profiles are most accurately drawn on the basis of extended speech samples. Although single word elicitations can provide insight into a child’s access to language-specific metrical structures (i.e., the prosodic grammar), spoken language rhythms are influenced by a variety of factors that combine to give rise to the perception of typical or disordered prosody. Extended speech samples are needed to detect the combined influence of these factors on speech. Thus, future research should focus on describing the linguistic and communicative context of global rate and pitch changes in extended speech samples as well as on the distribution of prosodic boundaries in these types of samples.

Acknowledgments

We are grateful to the families that participated in the longitudinal study that underpins the present study and to a number of research assistants (RAs) who passed through the Speech & Language Lab at the University of Oregon (PI: Redford) and the Speech Perception & Production Lab at Michigan State University (PI: Dilley). Those RAs who contributed the most to the data collection and transcriptions on which the present study was based were: Aubrianne Carson, Rachel Ciccerella, Faire Holliday, and Jayme Monroe in the Speech & Language Lab; and Jessica Gamache in the Speech Perception & Production Lab. Jessica Fanning, Ph.D., CCC-
Prosodic development

SLP helped to design the original storytelling task and oversaw the training and administration of speech-language assessments in the Speech & Language Lab.

Data Availability Statement

The speech samples analyzed in the present study were drawn from the Eugene Children’s Story Corpus (ECSC; Kallay & Redford, 2021), which is publicly available through the CHILDES database (MacWhinney, 2000). The ECSC includes both the audio files and pause-delimited orthographic transcriptions of the stories. The Rhythm and Pitch (RaP) transcriptions of the stories and raw measurement data extracted from the audio files and from the transcriptions are available upon request made to the corresponding author.

References


Prosodic development


Prosodic development


Prosodic development


Prosodic development


Prosodic development


Prosodic development


Prosodic development


Prosodic development


Prosodic development


Prosodic development


Prosodic development


Prosodic development
