Developmental change in the temporal patterning of vowels in school-age children’s speech

Melissa A. Redford & Hema Sirsa
University of Oregon

Running title:
Temporal patterning of vowels in children’s speech

Please address correspondence to:
Melissa Redford
Linguistics Department
University of Oregon
Eugene, OR 97403-1290
redford@uoregon.edu
Abstract
The motivating hypothesis for the present study is that the acquisition of articulatory timing interacts with the acquisition of language rhythm. Thirty school-aged children and 10 adults completed a delayed sentence repetition task. Twenty of the children completed the same task one year later. Fluent sentences from the cross-sectional and longitudinal speech samples were segmented to provide measures of segmental duration, speech rate, and overall rhythm. Several words and phrase were identified and segmented to provide temporal measures of lexical stress, phrase-level rhythmic grouping, and phrase-final lengthening. Results confirmed previous findings, namely, that children speak more slowly and produce longer and more variable vowel durations than adults. Results also indicated age-related differences in overall rhythm patterns that were best explained by phrase-level rhythmic grouping. Finally, the results indicated the continued evolution of lexical stress and final lengthening into middle childhood. The new findings on temporal patterns in children’s speech suggest the prolonged acquisition vowel reduction. Also suggested is an early strategy of vowel lengthening to attain adult-like temporal patterns. Middle childhood may represent an intermediate stage in development wherein temporal contrasts are exaggerated because children begin to appropriately reduce unstressed vowels and continue to lengthen stressed and phrase-final vowels.
1. Introduction

Classic studies on segmental durations and temporal variability in child speech indicate that adult-like articulatory timing is not attained until sometime after middle childhood (Smith, 1978; Kent & Forner, 1980; Lee, Potamianos, & Narayanan, 1999). For example, results from Lee and colleagues’ large-scale study on English-speaking children’s speech indicate that average vowel durations are greater in younger children than in older children. The results also indicate that vowel durations decrease with age until 12 years, and continue to decrease after this, past adult norms, before increasing again after age 15. The hypothesis that motivates the current study is that the very protracted acquisition of articulatory timing, exemplified by the slow acquisition of adult-like vowel durations, interacts with the acquisition of temporal patterns that are correlated with perceptions of language rhythm. Specifically, we expect school-age children’s rhythm patterns to deviate from adult rhythm patterns not because they have failed to acquire language-specific phonological patterns, but because they are continuing to refine control over articulatory timing. Rhythm here refers to the temporal spacing of prominences in a language. And, at least in English, perceptions of prominence are highly correlated with vowel duration (Cooper, Eady, Mueller, 1985; Turk & Sawuch, 1996; Kochanski, Grabe, Coleman, & Rosner, 2005).

Ramus, Nespor, & Mehler (1999) and Low, Grabe, & Nolan (2000) stimulated renewed interest in the so-called rhythm class hypothesis when they presented acoustic measures that successfully distinguished between languages that had been perceptually categorized by linguists as either stress-timed (e.g., British English and Russian), syllable-timed (e.g., Singapore English and Spanish), or mora-timed (e.g., Japanese and Tamil). The measures were strictly temporal in nature, and grounded in the assumption that the critical distinction between stress-timed languages on the one hand and syllable- and mora-timed languages on the other is in the presence versus absence of vowel reduction and its covariate, syllable structure complexity: canonical stress-timed languages have vowel reduction in non-prominent syllables and more complex syllable structures; canonical syllable- and mora-timed languages have only full vowels and simpler syllable structures (Dauer, 1983).

The temporal measures that Ramus and colleagues (1999) and Low and colleagues (2000) proposed are obtained by summarizing patterns of vocalic and consonantal interval duration across a sentence. Given vowel reduction and more complex syllable structure in stress-timed languages compared to syllable- and mora-timed languages, vowel durations should be more variable on average and
intervocalic consonantal durations longer on average in a stress-timed language compared to a syllable- or mora-timed language. Indeed, Ramus and colleagues found that distinct language rhythms were well distinguished by the mean proportion of vocalic duration in a sentence (%V—lower in stress-timed languages). Low, Grabe, and Nolan argued that language rhythm is best captured by successive vowel duration patterns, and proposed a measure called the Pairwise Variability Index (PVI). They further noted that measures like %V are likely highly influenced by speaking rate, which disproportionately affects vowel durations (Gay, 1981), and so normalized the PVI for speaking rate. The rate-normalized PVI is calculated by taking the difference between successive vowel durations, dividing this difference by the mean duration of the two vowels, summing these values across a sentence, and multiplying the resulting value by 100. Since the same operation has been applied to consonantal durations (Grabe & Low, 2002), it is useful to distinguish between the rate-normalized index for vowels (nPVI-V) and the rate-normalized index for consonants (nPVI-C). White and Mattys (2007), borrowing in part from Dellwo (2006), proposed two additional rate-normalized measures. These are merely the standard deviation of interval durations in a sentence divided by the mean interval duration for the sentence, multiplied by 100. The normalized measures are referred to as Varco-V and Varco-C, depending on whether they are based on vocalic or inter-vocalic consonantal intervals. Again, measures of vowel duration variability (e.g., nPVI-V and Varco-V) are expected to be higher in stress-timed languages compared to syllable- and mora-timed languages.

Interval-based rhythm measures have been applied to child speech in a few studies (Grabe, Watson, & Post, 1999; Bunta & Ingram, 2007; Payne, Post, Astruc, Prieto, & del Mar Vanrell, 2011). A major goal of these studies has been to understand the acquisition of language-specific patterns in younger children. For example, Payne and colleagues investigated cross-linguistics differences in the rhythm patterns produced by small groups of 2-, 4-, and 6-year-old children who were acquiring English, Spanish, or Catalan. Although unexpected, Payne and colleagues found an effect of language on several interval-based rhythm measures (e.g., nPVI-V and Varco-V) even in the youngest children’s speech. They also found an effect of age on several of the measures, and some significant differences in the variability of consonantal duration between 6-year-old and adult patterns across the language groups. An examination of the mean measurement values by language and age group given in the Appendix material suggests that nPVI-V and Varco-V values were also lower in English-speaking 6-year-old speech than in
English-speaking adult speech. These differences parallel those reported by Grabe and colleagues in their comparisons of 5-year-old and adult English. Payne and colleagues did not report age by language group comparisons for their measures, perhaps because their study included a limited number of subjects per age group per language (N = 3). Moreover, they concluded that children master prosodically-related vowel duration patterns by age 6, and discussed instead the motor factors that may contribute to developmentally-related differences in consonant production.

Payne and colleagues (2011) cited Allen and Hawkins (1980) in support of their conclusion that vocalic timing is mastered before consonantal timing. But the focus in that chapter, and in other work by Allen and Hawkins (e.g., Allen & Hawkins, 1978), is on English-speaking children’s slow acquisition of vowel reduction in weak syllables. For example, Allen and Hawkins (1978) conducted a transcription-based study of weak syllable reduction in the spontaneous speech of 5 children who ranged in age from 2;2 to 3;9 years. Weak syllable reduction was defined in perceptual terms as a centralized vowel, especially schwa, or as syllable deletion. Allen and Hawkins found that the children did not appropriately reduce English weak syllables. They also found that the children were equally poor at reducing weak syllables in function words and in polysyllabic content words, and best at realizing and reducing weak syllables after a strong syllable. These results led to the hypothesis of a trochaic bias in English-speaking children’s language, which many others have expanded upon since (e.g., Echols & Newport, 1992; Gerken, 1994; Demuth, 1996). This work is not reviewed here because its focus has been on syllable deletion rather than on syllable (vowel) reduction in very young children’s speech. Of interest, though, to the current hypothesis regarding the slow acquisition of rhythm are the preliminary results presented in Allen and Hawkins (1980:248), which show a production bias in the direction of the trochaic hypothesis in 5 and 6 year old children. Although suggestive, data from only 2 such children are discussed and the results are based on transcriptions rather than on acoustic measurement.

Most of acoustic phonetic work on children’s productions of unstressed or weak syllable vowel reduction has focused on preschool-aged children’s speech, and on their production of lexical stress in multi-syllabic words (Pollock, Brammer & Hageman, 1993; Kehoe, Stoe-Gamon, & Buder, 1995; Schwartz, Petinou, Goffman, Lazowski, & Cartusciello, 1996). Mainly, this work suggests that children acquire the contrast between lexically stressed and unstressed syllables very early. Young children do not, however, realize the contrast in exactly the same way as adults. For example, Kehoe and colleagues found
that the relative duration of stressed-to-unstressed vowels was the same in 2-year-old and adult speech, but that unstressed vowels in 2-year-old speech were substantially longer than those in adult speech. Similarly, Pollock and colleagues found that unstressed vowel durations decreased with developmental time. Finally, kinematic data indicate that even older children (4+ years of age) are more variable in their production of unstressed syllables compared to stressed syllables (Goffman, 1999). In all, the literature suggests that the ability to realize reduced vowels in lexically unstressed syllables may require more advanced motor skills than the ability to realize full vowels in lexically stressed syllables. If school-aged children have not mastered vowel reduction in lexically unstressed syllables, then they are unlikely to produce adult-like rhythm patterns.

Unstressed vowel reduction is also a feature of phrasal stress patterns in English. In particular, vowels in function words like “a” and “the” are typically reduced in a phrasal context. The perceptual result is the rhythmic grouping of the determiner and noun, called a clitic group, which is treated as a distinct category in some theories of prosodic phonology (e.g., Nespor & Vogel, 2007). The reader may recall that Allen and Hawkins (1978) found no evidence that children reduce vowels in function words differently from vowels in content words. But this conclusion was based on the percept of a centralized (schwa-like) vowel, which could be why the result has been contradicted by subsequent production work. For example, Goffman (2004) directly investigated whether or not school-aged children (4- to 7-year-olds) and adults realize unstressed syllables differently depending on the morphosyntactic status of the unstressed syllable. She compared the production of function word syllables in determiner noun phrases (e.g., “a BAB”) to lexically unstressed syllables in disyllabic nouns (e.g., “aBAB”) in child and adult speech. Acoustic and kinematic measures indicated that children did not differentiate between unstressed syllables in the two contexts even though adults reduced unstressed syllables in determiner noun phrases more than in disyllabic words. If Goffman’s results apply to more naturalistic speech, then we might expect that children’s failure to rhythmically group function and content words into clitic groups would affect the overall rhythm pattern of their speech.

Phrase-final lengthening is another source of temporal variability that affects rhythm (White & Mattys, 2007). Final lengthening is associated with prosodic boundaries, and is especially notable in English at the intonational phrase boundary. Whereas final lengthening is evident even in pre-speech vocalizations (Robb & Saxman, 1990), it is not clear when it becomes fully adult-like in child language.
Snow (1994) has shown, for example, that final lengthening continues to evolve with the acquisition of other linguistic skills, albeit in very young children. We are unaware of any studies that compare final lengthening in younger and older children’s speech or of work comparing final lengthening in children and adults’ speech, and so cannot state with certainty whether school-aged children show the same degree of final lengthening as adults. However, if final lengthening continues to evolve into middle childhood, then it could contribute to differences in child and adult rhythm.

In sum, school-aged children have yet to master control over articulatory timing and so produce longer and more variable vowel durations than adults (Smith, 1978; Kent & Forner, 1980; Lee, Potamianos, & Narayanan, 1999). Insofar as language rhythm can be adequately characterized using measures based on vowel durations, it is likely that the protracted acquisition of articulatory timing also affects the acquisition of language rhythm. The current cross-sectional and longitudinal study tested this hypothesis. Segmental durations, speech rate, and overall rhythm measures were taken on speech produced by school-aged children and adults. The cross-sectional study tested whether these measures differed in speech produced by younger and older school-aged children and adults. The younger children were 5 years old on average, and the older children were 8 years old on average. The longitudinal study tested for developmental change in a subset of the younger and older children who returned to the lab one year after their first visit.

Since one of the most significant features of English rhythm is the alternation of reduced and full vowels, we might expect that child and adult productions of lexical or phrasal stress could account for any individual or age-related differences in language rhythm. Another major linguistic source of vowel duration variability is phrase-final lengthening, which may or may not differ in child and adult language. Accordingly, temporal measures of lexical stress, phrasal stress, and final lengthening were taken and used to predict variability in overall rhythm patterns. Additional analyses investigated the effect of age on these measures to provide additional information on the acquisition of specific prosodic temporal patterns.
2. Methods

2.1 Participants

The cross-sectional study included 15 typically developing 5-year-old children (M = 5;7, SD = 3.6 months) and 15 typically developing 8-year-old children (M = 8;5, SD = 3.9 months). All children were native American-English speakers. The children had no speech/language or hearing problems as determined by parental report and a pure-tone hearing screening. In addition to the children, 10 undergraduate students from the University of Oregon participated in the study. These students were native-American English speakers and self-reported normal hearing. A subset of ten 5-year-old and ten 8-year-old children returned to the lab a year later. The 5-year-olds had turned 6 (M = 6;4, SD = 3.2 months), and the 8-year-olds had turned 9 (M = 9;3, SD = 4.8 months).

2.2 Materials

Speakers were audio recorded while completing the Recalling Sentences subtest from the Clinical Evaluation of Language Fundamentals (CELF-4; Semel, Wing, & Secord, 2003). The subtest uses delayed imitation of sentences to evaluate a child’s linguistic development. The initial sentences in the test are shorter and syntactically simpler than later sentences, and the test always proceeds from simpler sentences to more complex ones. We administered the test in the normal way, except that we always began with the first test sentence regardless of a child’s age and we used prerecorded stimuli to control for sentence prosody and other phonological and phonetic characteristics of the model sentences.

Overall rhythm measures were based on the first 6 test sentences from the Recalling Sentences subtest. Children and adults almost always produced these first sentences readily and fluently, and without error. The sentences were as follows:

1. The tractor was followed by the bus.
2. Did the girl catch the baseball?
3. Did you remember to bring your lunch?
4. The boy fell and hurt himself.
5. Was the van followed by the ambulance?
6. The rabbit was not put in the cage by the girl.

Out of the 240 sentences collected for the cross-sectional study, 9 of the children’s sentences were excluded due to mismatches with the target or disfluencies. Disfluencies were defined by within
sentence hesitations (prosodically incorrect prolongations), the presence of filled pauses, restarts, or word truncations. Out of the additional 120 sentences collected for the longitudinal study, 8 were excluded due to mismatches with the target or disfluencies, and 3 from an older child were mistakenly never collected (i.e., the tester skipped the first 3 test sentences). Of the remaining 340 sentences, 37 included short silent pauses that accompanied other cues associated with prosodic focus. Robust focus marking was characteristic of the enthusiastic style with which 16 children and 1 adult completed the task. 10 of the children were from the younger group and 6 were from the older group. Six of the 16 children produced at least one sentence with an emphatic pause on both visits.

Measures of specific prosodic temporal patterns were taken by extracting words and phrases from the six sentences listed above and from the other sentences that the children and adults produced. A temporal measure of lexical stress was based on 4 trochaically-stressed disyllabic words (castle, mother, rabbit, tractor). A temporal measure of rhythmic group was based on 4 determiner noun phrases (the boy, the girl, the van, the cage). And, a temporal measure of phrase-final lengthening was based on 4 two-word phrases in final position (the bus, the girl, cat’s food, for class). The disyllabic words and determiner noun phrases were extracted from initial or medial phrase position. If a word was produced incorrectly or in a disfluent context, it was excluded from measurement. Measurements on the younger children’s speech were based on 86 disyllabic words, 94 determiner noun phrases, and 90 two-word phrases in final position. Measurements on the older children’s speech were based on 96 disyllabic words, 98 determiner noun phrases, and 94 two-word phrases in final position. All of these numbers were out of a possible 100 items (15 children × 4 items = 60 items at visit 1 + 10 children × 4 items at visit 2).

Measurements on adult speech were based on the total possible number of items (i.e., there were no exclusions).

2.3. Segmentation

Speakers’ productions were hand segmented in Praat (Boersma & Weenink, 2006). Each sentence was displayed as an oscillogram and spectrogram. Vocalic and consonantal intervals were identified using both visual and auditory cues. The identification of intervals was based primarily on the presence/absence of higher formant structure. Vowels were defined as intervals with substantial energy in F2 and F3. Thus, like Ramus et al. (1999), glides were frequently included as part of the vocalic interval. A portion of liquid articulation was also usually included as part of the vocalic interval. Intervocalic intervals were
segmented into component consonantal intervals based on phoneme specific criteria. Once segmented, interval durations were extracted automatically for each sentence, and the following values calculated: mean vowel and consonant durations, standard deviation of vowel and consonant durations, speech rate (= number of vowel intervals per second excluding silent pauses), %V, raw PVI-V, rate normalized PVI-V, and Varco-V. The values were then averaged across sentences within a speaker for the analyses.

The segmentation procedure was assessed for internal consistency. The first author selected a set of 30 sentences for re-segmentation and stripped them of all subject and age identifying codes. The quasi-random selection was meant to ensure a representative sample of sentences from different speakers of different ages. The second author then blindly re-segmented the sentences and generated the same descriptive statistics and rhythm measures that served as dependent variables in the analyses. Correlations of the measures derived from the new and original sentence segmentations indicated high intra-rater consistency. The Pearson correlation coefficients ranged from a low of 0.965 for the standard deviation of consonantal durations to a high of 0.984 for the standard deviation of vocalic durations. An example of the intra-rater reliability statistics that fell between theses low and high statistics is the correlation coefficients for the rate-normalized measure of sequential vocalic interval variability (nPVI-V), which was 0.977.

When the disyllabic words, determiner noun phrases, and two-word phrases in final position were extracted from test sentences that had not been segmented, they were segmented in the sentences in which they occurred according to the same segmentation criteria. Again, interval durations were extracted automatically and vowel duration ratios were derived to capture patterns associated with lexical stress, phrase-level rhythmic grouping, and phrase-final lengthening. For lexical stress, the ratio was calculated by dividing the duration of the first vowel by the second (V1:V2). For phrase-level rhythmic grouping, the ratio was calculated by dividing the vowel duration of the noun by the vowel duration of the determiner (N:Det). For phrase-final lengthening, the ratio was calculated by dividing the vowel duration in the final word by the vowel duration in the non-final word (Ult:Pen). The ratio values were then averaged across the 4 items within a speaker for the analyses.
3. Results and Discussion

The analyses investigated the effect of age on overall measures of temporal variability in vowel and consonant production as well as in speech rate. The significant effect of age on these measures led us to investigate the effect of age only on the rate-normalized measures of rhythm. The effect of age was found to be significant for nPVI-V and Varco-V. A significant portion of the variance in nPVI-V and Varco-V values was accounted for by a single prosodic temporal pattern, namely, the noun-to-determiner vowel duration ratio. The specific temporal patterns associated with lexical stress and final lengthening, though not correlated with nPVI-V or Varco-V, varied with age. The results from analyses on the cross-sectional data are presented first, followed by results from analyses on the longitudinal data.

3.1. Cross-sectional data

The first set of analyses investigated the fixed effect of age group on vowel and consonantal durations and on speech rate. The effect of age group was significant for mean vowel durations, $F(2,37) = 14.69; p < 0.001$, the standard deviation of vowel durations, $F(2,37) = 4.89; p = 0.013$, mean consonantal durations, $F(2,37) = 5.63; p = 0.007$, and speech rate, $F(2,37) = 16.38, p < 0.001$, but not for the standard deviation of consonantal durations. Figure 1 plots the results for speech rate.

![Figure 1](image.png)

Figure 1. Speech rate is shown as a function of age group. Younger children were 5 years old, older children were 8 years old, and adults were college age. Error bars show the 95% confidence interval.
Note that the effect of age on speech rate shown in Figure 1 was similar to the effect of age on mean vowel and consonantal durations and on the standard deviation of vowel durations, probably because these measures all negatively correlated with speech rate (r’s ranged from –0.708 for the standard deviation of vowel duration to –0.893 for mean vowel duration). Post hoc mean comparisons indicated that speech rate was indeed slower in younger children’s speech than in older children or adults’ speech, but speech rate in older children did not differ from speech rate in adults. Similarly, vowel and consonantal durations were longer in younger children’ speech than in older children or adults’ speech, but mean durations did not differ in the speech of older children and adults. Younger children also produced more variable vowel durations than adults, but variability in consonantal durations was not different across the age groups. The results from the post hoc tests are presented in Table 1 for all dependent measures. Note that the Bonferroni method was used to control for multiple comparisons.

Table 1. Results of post hoc tests are shown for the different comparisons and measures: mean vowel duration (mean V dur), the standard deviation of vowel duration (s.d. V dur), mean consonantal duration (mean C dur), the standard deviation of consonantal duration (s.d. C dur), and speech rate. Interval measures were in milliseconds, speech rate in syllables per second. Younger children were 5 years old, older children were 8 years old, and adults were college age.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Measure</th>
<th>Mean Difference</th>
<th>Std. Error</th>
<th>Adjusted p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Younger vs. Older Children</td>
<td>mean V dur</td>
<td>21.54</td>
<td>4.87</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>s.d. V dur</td>
<td>5.06</td>
<td>2.39</td>
<td>0.123</td>
</tr>
<tr>
<td></td>
<td>mean C dur</td>
<td>9.22</td>
<td>3.32</td>
<td>0.026</td>
</tr>
<tr>
<td></td>
<td>s.d. C dur</td>
<td>5.62</td>
<td>3.65</td>
<td>0.398</td>
</tr>
<tr>
<td></td>
<td>speech rate</td>
<td>–0.49</td>
<td>0.12</td>
<td>0.001</td>
</tr>
<tr>
<td>Younger Children vs. Adults</td>
<td>mean V dur</td>
<td>26.01</td>
<td>5.44</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>s.d. V dur</td>
<td>8.03</td>
<td>2.67</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td>mean C dur</td>
<td>10.85</td>
<td>3.71</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td>s.d. C dur</td>
<td>10.02</td>
<td>4.08</td>
<td>0.057</td>
</tr>
<tr>
<td></td>
<td>speech rate</td>
<td>–0.71</td>
<td>0.13</td>
<td>0.000</td>
</tr>
<tr>
<td>Older Children vs. Adults</td>
<td>mean V dur</td>
<td>4.47</td>
<td>5.44</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>s.d. V dur</td>
<td>2.97</td>
<td>2.67</td>
<td>0.818</td>
</tr>
</tbody>
</table>
Due to significant differences between the age groups on measures of speech rate and the speech rate correlated duration measures, analyses on overall rhythm focused only on the rate-normalized measures, namely, nPVI-V and Varco-V. Results indicated a significant effect of age group on both nPVI, $F(2,37) = 4.98; p = 0.012$, and Varco-V, $F(2,37) = 3.30; p = 0.048$. The nPVI results are shown in Figure 2.

![Figure 2](image.png)

Figure 2. The group mean values for one of the rate normalized rhythm measures, nPVI-V, is shown as a function of age group. Younger children were 5 years old, older children were 8 years old, and adults were college age. Error bars show the 95% confidence interval.

Post hoc comparisons indicated a significant mean difference in nPVI-V values in younger and older children’s speech (Younger – Older = −6.33, $SE = 2.02$, $p = 0.010$), but no significant differences in values associated with younger children and adults’ speech or older children and adults’ speech. The comparisons also indicated that none of the pairwise differences in Varco-V values were significant,
though the difference between younger and older children’s values neared significance (Younger – Older = –3.55, SE = 1.43, p = 0.053).

Results of the mean comparisons indicated that there was substantial overlap between the groups in spite of the significant effect of age on nPVI-V and Varco-V. Figure 3 shows the extent to which the groups overlapped, as well as the degree to which the nPVI-V and Varco-V were correlated (r = 0.72).

Figure 3. The relationship between the two rate normalized rhythm measures, Varco-V and nPVI-V, is shown as a function of age group. Younger children were 5 years old, older children were 8 years old, and adults were college age.

The data in Figure 3 suggest that the effect of age group was driven by a number of younger children who produced speech with low nPVI-V and Varco-V values and a few older children who produced speech with high nPVI-V and Varco-V values. These individual differences were not accounted for by the participants’ age. That is, regression analyses, conducted within age group, indicated that age in months did not account for a significant portion of the variance in nPVI-V and Varco-V.

The next set of analyses on the cross-sectional data investigated whether individual differences in rhythm production, within and between age groups, could be accounted for in terms of differences in the
production of specific prosodic temporal patterns; namely, in terms of individual differences in the production of lexical stress (V1:V2), phrase-level rhythmic grouping (N:Det), and phrase-final lengthening (Ult:Pen).

The 3 measures of specific temporal patterns were used to predict nPVI-V and Varco-V values in regression models that controlled for age group. The full models accounted for 39% of the variance in nPVI-V values (Adjusted $R^2 = 0.32$) and the same amount of variance in Varco-V values (Adjusted $R^2 = 0.32$). The only significant predictor in either model was N:Det. Alone, this measure accounted for 34% of the variance in nPVI-V values and 33% of the variance in Varco-V values. The relationship between the overall rhythm measure, nPVI-V, and the specific temporal pattern, N:Det, is shown in Figure 4.

![Figure 4](image)

**Figure 4.** The relationship between the noun-to-determiner vowel duration (N:Det) and a rate normalized rhythm measure, nPVI-V, is shown as a function of age group. Younger children were 5 years old, older children were 8 years old, and adults were college age.

Figure 4 shows the positive relationship between N:Det and nPVI-V. Speech with higher, more adult-like nPVI-V values was produced by children who were also more likely to produce monosyllabic nouns with especially long vowels and determiners with especially short vowels in the non-final
determiner noun phrases. A long-short, noun-to-determiner vowel ratio is consistent with the production of a determiner noun phrase as a clitic group.

Although N:Det was the only significant predictor of the overall rhythm measures, analyses on N:Det indicated that it was not significantly different across the different age groups. However, the effect of group was significant for the temporal patterns associated with lexical stress (V1:V2), $F(1,37) = 5.00$, $p = 0.012$, and final lengthening (Ult:Pen), $F(1,37) = 3.89$, $p = 0.029$. The group mean, standard error of the mean, and group standard deviation for these measures are presented in Table 2 along with mean comparisons between the groups and the Bonferroni adjust $p$-values for those comparisons. Recall that the measures were averaged across the items within a speaker so that $N$ equals 15 measures for younger children, 15 for older children, and 10 for the adults.

Table 2. Descriptive statistics and the results of post hoc mean comparisons are shown for the different age groups and the specific measures of lexical stress (V1:V2), phrase-level rhythmic grouping (N:Det), and phrase-final lengthening (Ult:Pen). Younger children were 5 years old, older children were 8 years old, and adults were college age.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Statistic</th>
<th>Younger Children</th>
<th>Older Children</th>
<th>Adults</th>
<th>Mean Comparisons</th>
<th>Adjusted $p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>V1:V2</td>
<td>Mean</td>
<td>1.43</td>
<td>1.41</td>
<td>1.01</td>
<td>Younger vs. Older</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>S.E.</td>
<td>0.10</td>
<td>0.10</td>
<td>0.04</td>
<td>Younger vs. Adult</td>
<td>0.020</td>
</tr>
<tr>
<td></td>
<td>s.d.</td>
<td>0.40</td>
<td>0.40</td>
<td>0.14</td>
<td>Older vs. Adult</td>
<td>0.026</td>
</tr>
<tr>
<td>N:Det</td>
<td>Mean</td>
<td>3.46</td>
<td>3.67</td>
<td>3.90</td>
<td>Younger vs. Older</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>S.E.</td>
<td>0.26</td>
<td>0.21</td>
<td>0.20</td>
<td>Younger vs. Adult</td>
<td>0.633</td>
</tr>
<tr>
<td></td>
<td>s.d.</td>
<td>1.02</td>
<td>0.81</td>
<td>0.64</td>
<td>Older vs. Adult</td>
<td>1.000</td>
</tr>
<tr>
<td>Ult:Pen</td>
<td>Mean</td>
<td>2.75</td>
<td>3.52</td>
<td>2.79</td>
<td>Younger vs. Older</td>
<td>0.045</td>
</tr>
<tr>
<td></td>
<td>S.E.</td>
<td>0.20</td>
<td>0.26</td>
<td>0.14</td>
<td>Younger vs. Adult</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>s.d.</td>
<td>0.78</td>
<td>1.04</td>
<td>0.43</td>
<td>Older vs. Adult</td>
<td>0.113</td>
</tr>
</tbody>
</table>

Table 2 indicates that the group differences in V1:V2 were due to higher ratios in children’s speech than in adults’ speech. Higher ratios are consistent with either lengthening of the stressed syllable or vowel reduction in the unstressed syllable or both. Analyses of the effect of age group on unstressed and stressed vowel durations indicated that unstressed vowel durations were the same in children and
adults’ speech, but that stressed vowel durations were different, $F(2,37) = 6.95; p = 0.003$. This result accounts for why $V1:V2$ values were higher in children than in adults’ speech.

The results shown in Table 2 suggest that the effect of age on final lengthening was due to greater lengthening in older children’s speech compared to younger children or adults’ speech. Analyses of penultimate and ultimate vowel durations suggest that the difference between younger and older children’s speech was due to a difference in the duration of penultimate vowel durations (Younger – Older = 21.44; Adjusted $p = 0.004$) rather than to a difference in final vowel durations (Younger – Older = 13.08; Adjusted $p = 1.000$). In particular, older children’s penultimate vowel durations were shorter and no different from adult vowel durations, but their final vowels were as long as younger children’s vowels.

Also evident in Table 2, is that vowel duration ratios were much higher when the stressed and unstressed vowels were compared in determiner noun phrases than when they were compared in disyllabic nouns. The ratios associated with final lengthening were lower on average than those associated with the rhythmic grouping of the determiner and noun. The measure of final lengthening compared vowels in phrase-final, monosyllabic nouns to those in penultimate, monosyllabic content and function words. The mix of penultimate content and function words likely led to longer vowel durations in the divisor for the measure of final lengthening compared to the measure of rhythmic grouping for determiner noun phrases, which included only function word vowel durations in the divisor.

### 3.1. Longitudinal data

We repeated the same analyses on the longitudinal data. The first set of analyses once again investigated the effect of age group on interval durations and speech rate. The repeated measures analyses also included the within-age-group factor, visit (1st or 2nd).

As in the cross-sectional data, the effect of age group was significant for mean vowel durations, $F(1,18) = 23.06; p < .001$, the standard deviation of vowel durations, $F(1,18) = 6.82; p = .018$, mean consonantal durations, $F(1,18) = 9.44; p = .007$, and for speech rate, $F(1,18) = 22.41, p < .001$. As before, the effect of age was not significant for the standard deviation of consonantal durations. There was no effect of visit or of the interaction between visit and age group on any of the measures. Figure 5 plots the representative results of speech rate.
Figure 5. Speech rate is shown as a function of visit and age group. Younger children were 5 years old at the first visit and 6 years old at the second, older children were 8 years old at the first visit and 9 years old at the second. Error bars show the 95% confidence interval.

The analyses on rate normalized rhythm measures also indicated a significant effect of age group on nPVI-V, $F(1,18) = 9.70; p = .006$, and on Varco-V, $F(1,18) = 8.76; p = .008$. The effect of visit on nPVI-V approached significance, $F(1,18) = 4.34; p = .052$, but there was no trend towards a significant interaction between visit and age group. By contrast, there was no trend towards a significant effect of visit on Varco-V, but the interaction between visit and age group trended towards significance, $F(1,18) = 4.35; p = .052$. The results for nPVI-V and Varco-V are shown in Figure 6.
Figure 6. The group mean values for the rate normalized rhythm measures, nPVI-V and Varco-V, are shown as a function of visit and age group. Younger children were 5 years old at the first visit, and 6 years old at the second. Older children were 8 years old at the first visit and 9 years old at the second. Error bars show the 95% confidence interval.

Next, we used the specific temporal patterns to predict individual variability in nPVI-V and Varco-V values in regression models that controlled for visit and age group. The best model of nPVI-V values, which was not significantly different from the full model, accounted for 51% of the variance (Adjusted $R^2 = .47$). The significant predictors were visit, $t = 2.15; p = .038$; df = 39, age group, $t = 3.32; p = .002$; df = 39, and N:Det, $t = 3.97; p < .001$; df = 39. The best model of Varco-V values, which was not significantly different from the full model, accounted for 46% of the variance (Adjusted $R^2 = .41$). The significant predictors were visit, $t = 2.32; p = .026$; df = 39, age group, $t = 2.29; p = .028$; df = 39, and N:Det, $t = 3.98; p < .001$; df = 39. Alone, N:Det accounted for 29% of the variance in nPVI-V, and for 30% of the variance in Varco-V. Figure 7 shows the model results for nPVI-V.
Figure 7. The relationship between the rhythmic grouping of determiner noun phrases (N:Det) and a rate normalized interval based rhythm measure, nPVI-V, is shown as a function of age group and visit. Younger children were 5 years old at the first visit and 6 years old at the second. Older children were 8 years old at the first visit and 9 years old at the second.

As in the cross-sectional data, the rhythm measures were well accounted for by the noun-to-determiner vowel duration ratio. As before, higher values of the ratio corresponded to higher values of the rhythm measure. Again, in spite of the high correlation with the rhythm measures, mean N:Det did not vary significantly by visit or by age.

In contrast to N:Det, temporal patterns associated with lexical stress, V1:V2, and phrase-final lengthening, Ult:Pen, both varied significantly with visit (V1:V2, $F(1,18) = 6.58; p = 0.019$; Ult:Pen, $F(1,18) = 451.78; p < .001$). Children’s V1:V2 ratios were higher at the first visit ($M = 1.48$; s.d. = 0.42) compared to the second ($M = 1.26$; s.d. = 0.26). This effect did not vary significantly with age group.

Children’s Ult:Pen ratios were also higher at the first visit ($M = 3.26$; s.d. = 1.00) compared to the second ($M = 2.77$; s.d. = 0.81). This effect varied significantly with age group, $F(1,18) = 14.29; p = 0.001$, such that the mean difference in ratios between visits 1 and 2 was larger for older children (mean difference = 0.71) than for younger children (mean difference = 0.29). Children’s values of V1:V2 and Ult:Pen at visit 2 were smaller than at visit 1 and closer to the adult values shown in Table 2 above.
4. General Discussion

The present study confirms that younger children produce longer vowels and consonants than older children and adults, and that they speak at slower rates than older children and adults. The different temporal patterns of child and adult speech are assumed to reflect differences in control over articulatory timing; specifically, longer (and more variable) segmental durations, and slower articulation rates are associated with immature or poorer control than shorter segmental durations and faster articulation rates (Smith, 1978; Kent & Fomer, 1980; Tsao & Weismer, 1997; Lee, et al. 1999). We expected that developmental differences in articulatory timing would have consequences for language rhythm. This expectation was upheld in the present study. Mean comparisons of nPVI-V and Varco-V values indicated significant differences between younger and older children’s speech, and, in spite of the overlap between child and adult values in the cross-sectional data, the longitudinal results suggest that rhythm patterns continue to evolve in middle childhood.

Variance in the overall rhythm patterns produced by younger and older children was accounted for by a specific temporal pattern, N:Det, which captured the relative duration of the vowels in determiner noun phrases. Lower ratio values signal more equal vowel durations across the phrase. Higher N:Det values were associated with more adult-like rhythm patterns. These patterns reflect reduction in the determiner and cliticization to the following noun. Insofar as clitic groups are a recognized category within the prosodic hierarchy, the failure to reduce function words and cliticize them to adjacent content words could represent a failure to fully achieve a prosodic category.

Although there was no effect of age group on N:Det, results from subsequent analyses of the mean vowel durations for determiners and nouns indeed suggest that younger children failed to reduce the determiner to the same degree as older children. Younger children produced longer vowel durations in determiners than older children (Younger – Older = 23.21; Adjusted $p < 0.004$). However, younger children also produced somewhat longer vowel durations in nouns than older children (Younger – Older = 13.08; Adjusted $p = 0.053$). This result could indicate that some younger children may have achieved a long-short, noun-to-determiner pattern by lengthening the stressed content word vowel rather than by reducing the unstressed function word vowel. Extrapolating from this finding, we conjecture that in the absence of the skills necessary for reduction, children can achieve target prosodic category via
lengthening. The objective result of such a strategy is a more adult-like rhythm pattern, at least as defined by the interval-based measures.

The suggestion that children may achieve target prosodic temporal patterns via stressed vowel lengthening rather than via unstressed vowel reduction is consistent with the literature on lexical stress production in young children. Recall that Kehoe and colleagues (1995) found that lexically unstressed vowels are longer in 2-year-old speech compared to adult speech, even though the relative duration of lexically stressed to unstressed vowels is the same in child and adult speech. The present findings indicate that by school age, children’s lexically unstressed vowel durations are the same as older children and adults’ unstressed vowel durations. In spite of adult-like production of lexically unstressed syllables, we still find an effect of age on the temporal pattern associated with lexical stress. In particular, V1:V2 ratios were larger in children than in adults. This result was due to an age-related difference in stressed vowel durations. Children were found to produce longer stressed vowels than adults, and younger children produced longer stressed vowels than older children. When considered in conjunction with Kehoe et al.’s results, the present results suggest the following developmental progression. Early on children compensate for incomplete vowel reduction by over-lengthening an adjacent stressed vowel in order to achieve target prosodic patterns. Once children are able to reduce the unstressed vowel, they nonetheless continue to over-lengthen the stressed vowel, leading to exaggerated durational contrasts that nonetheless serve to cue prosodic categories.

The results on final lengthening are consistent with a development progression that leads to exaggerated durational contrasts in children’s speech. In particular, our data indicate greater final lengthening in older children’s speech compared to younger children’s speech. This difference was attributed to a difference in penultimate vowel durations, which were shorter in older children’s speech compared to younger children’s speech. There were no age-related differences in the duration of the ultimate vowel in children’s speech.

Note that the present results on prosodic temporal patterns conflict with one another with regards to unstressed vowel shortening. Younger children produced longer unstressed vowels than older children in determiner noun phrases, but younger and older children produced equally short and adult-like unstressed vowels in disyllabic nouns. The additional finding that younger children also produced longer vowels in phrase-penultimate words than older children suggests that the relevant distinction may be
between phrasal and lexical patterns rather than between stressed and unstressed vowels. It is possible that younger children master the motor skills required for lexical patterns prior to those required for phrasal patterns. This possibility is consistent with Goffman’s (2004) data on the production of unstressed syllables as a function of morpho-syntactic status. Recall that whereas children produced all unstressed syllables in the same way, adult reduced the unstressed syllables in determiner noun phrases more than those in disyllabic content words. Children in the present study certainly encoded all target durational contrasts, but younger children were unable to shorten vowels in a phrasal context to the same degree as older children or adults.

5. Conclusion

Studies on the acquisition of speech motor control indicate that articulatory timing continues to evolve until quite late in development. Our working hypothesis is that the prolonged acquisition of articulatory timing interacts with the acquisition of linguistic patterns, and specifically with the acquisition of prosody. The results from the current study confirm previous findings of longer durations and greater temporal variability in school-aged children’s segmental productions compared to adults’ productions. Further, the present results indicate that children’s language rhythm is distinct from adults’ rhythm. Age-related differences in rhythm were attributed to differences in the production of determiner noun phrases. More adult-like rhythm patterns were achieved when determiners were fully reduced and, presumably, cliticized to the following noun. Importantly, though, age-related differences in this pattern hinged on the degree to which the determiner was reduced rather than on whether or not it was reduced. Even the youngest school-aged children matched the target temporal pattern of short, then long vowels in determiner noun phrases. It is just that some young children were, presumably, unable to sufficiently reduce the determiner.

The results from analyses of the specific prosodic temporal patterns produced by children and adults suggested that children achieve adult-like patterns via a general strategy of vowel lengthening rather than vowel reduction. We assume that children adopt this strategy because they are sensitive to the perceptual contrast encoded in the input, but cannot attain this contrast via vowel reduction, which requires advanced motor skills. The results also suggest that the ability to reduce vowels is attained gradually. Whereas even the youngest school-aged children produced lexically unstressed vowels
similarly to adults, they did not produce the unstressed vowels in determiner noun phrases in the same way as adults, perhaps because these are reduced even further than lexically unstressed vowels. The gradual development of skills necessary to achieve vowel reduction may interact with the vowel lengthening strategy that children use to attain the prosodic patterns characteristic of the input. This interaction may give rise to an intermediate stage wherein contrasts are exaggerated. At this stage of development, children appropriately reduce vowels, but also continue to lengthen stressed vowels. Younger children appear to have reached this intermediate stage with regards to lexical stress production. Older children appear to have met this stage with regards to final lengthening.

**Acknowledgments**

This research was supported in part by Award Number R01HD061458 from the Eunice Kennedy Shriver National Institute of Child Health & Human Development (NICHD). The content is solely the responsibility of the authors and does not necessarily reflect the views of NICHD.
References


