Prosodic Characteristics of American English in School-Age Children

Katsura Aoyama[1], *Christina Akbari*[2], *James E. Flege*[3]

[1] University of North Texas, USA
[2] Arkansas State University, USA
[3] University of Alabama at Birmingham, USA

katsura.aoyama@unt.edu, cakbari@astate.edu, jeflege@uab.edu

Abstract

This study investigated prosodic characteristics of American English in school-age children. Previous studies reported that children's speech productions differed from those of adults in temporal and pitch aspects of speech prosody. The current study analyzed speech samples from 16 adults and 16 school-age children using both absolute measures (duration and fundamental frequency) and proportional measures (rhythm metrics and pitch range). The results showed differences between adults and children in absolute measures of temporal and pitch aspects of speech production, but these differences diminished in proportional measures. For temporal aspects of speech, absolute durations of children's utterances were longer than adults' utterances, whereas no statistically significant differences were found between adults' and children's rhythm metrics. Similarly, absolute fundamental frequency values were higher in children's speech than in adults' speech, but the pitch range did not differ between adults and children. These results suggest that children's speech may be slower in rate and higher in pitch, but their prosodic characteristics may be similar to those of adults in the temporal and pitch aspects of speech prosody by school age.

Index Terms: children, duration, pitch, American English

1. Introduction

It has been reported that children's speech prosody differs from that of adults in several different aspects. Previous studies have demonstrated that children's utterances tend to be longer in duration [1]–[4] and higher in fundamental frequency (F_0) [2, 5, 6]. In addition, speech rhythm may be different between children and adults in English [7]–[13]. Wells et al. [14] reported that the overall use of intonation continues to change into the school-age years, and Shport and Redford [15] reported that children's phrase-level prominence patterns are not yet adult-like in 6- and 7-year-olds. These studies indicate that the acquisition of prosody continues well into the school-age years in English.

For F_0, the absolute values steadily decrease with chronological age due to physiological growth [2, 6]. In addition, children learn how to use F_0 cues and intonation for other aspects of speech and language [14, 15]. Katz et al. [16] suggested that 7-year-old children were not able to use F_0 cues reliably for syntactic purposes in English.

For temporal aspects of speech, children's utterances are longer in duration than adults until approximately age six [1]–[3]. Children's articulation rate gradually increases from age 5 to 8, although there is a wide range of individual variations and task effects [17]. Moreover, some segments, such as the schwa, were longer in children's speech than in adults', while no age differences were found in other segments in [1, 18].

Grabe et al. [9] and Sirsa and Redford [13] suggested that the acquisition of rhythm in English takes longer than in other languages. Grabe et al. [9] analyzed speech samples from 4-year-olds using rhythm metrics. They found that the 4-year-olds' rhythm metric values differed from the adults' values in British English, suggesting that 4-year-olds have not yet acquired adult-like rhythm. Sirsa and Redford [13] reported that there are differences between 5-year-olds and 8-year-olds in speech rate and rhythm measures. Ordin and Polyanskaya [10] also reported changes in rhythm metrics in children between 4 and 11 years of age in British English.

Aoyama and Guion [19] examined prosodic aspects in school-age children and adults in American English. Although the focus of [19] was to compare non-native speakers and native speakers of American English, some adult-child differences were found across native and non-native speakers. For duration, some syllables were longer in children's speech than in adults' speech, while other syllables were shorter in children's speech than in adults' speech. Pitch range was also greater in adults' speech than in children's speech in one of the three utterances analyzed.

The aim of this study was to further explore the differences between adults and children that were found in [19]. The current study analyzed a larger set of speech samples from the same native English-speaking adults and children who participated in [19]. Additional measures of temporal and pitch aspects, both absolute and proportional, were employed to characterize the differences between adults and children.

2. Method

2.1. Participants

The data were collected as part of a larger study examining Japanese adults' and children's learning of American English [19]–[22]. Participants in the control groups of the above longitudinal study were 16 native English-speaking adults (7 males and 9 females) and 16 children (11 males and 5 females). Previous reports from the larger study primarily focused on segmental perception and production [20]–[22]. Although prosodic aspects of speech were reported in [19], the specific speech samples in the current study had not been analyzed previously (see 2.2 for details).

Speech samples were collected from each participant twice, approximately one year apart (T1 and T2). The adult participants' mean age at T1 was 40.26 years (*SD* = 4.73, range

doi: 10.21437/SpeechProsody.2016-117

33.9 to 49.9 years) and the child participants' mean age was 10.62 years (*SD* = 2.13, range 7.0 to 13.9 years). All of the participants lived in Alabama. The participants did not speak any language other than American English and reported no history of speech or hearing problems.

2.2. Materials and elicitation procedures

All participants were tested individually in a quiet room at their homes or at the University of Alabama at Birmingham (UAB). Ten phrases or sentences ("utterances") were elicited from each participant. The target utterances consisted of two to six syllables and included a variety of vowels and consonants. Each target utterance was elicited using the following format:

1) Q: *How are you today*? A: *I'm fine*.
2) Q: *Where do you live*? A: *In the United States*.
3) Q: *What time is it*? A: *Ten o'clock*.
4) Q: *How much does it cost*? A: *Five dollars*.
5) Q: *Where did the children go*? A: *They went to school*.
6) Q: *Where did the man go*? A: *He went to work*.
7) Q: *What did he drink*? A: *A glass of water*.
8) Q: *What did the girl eat*? A: *She ate a sandwich*.
9) Q: *What did you read*? A: *I read a good book*.
10) Q: *How old is Julie*? A: *She is eight years old*.

The participants first heard a recording of the question followed by the model (i.e., target) answer. The question was then repeated without the recording of the answer, and the participants answered. This question-answer format was used in order to elicit the same utterances from the participants without using written language.

The questions were spoken by a male native speaker of American English and the answers were spoken by a female native speaker of American English in the elicitation recording. The recorded questions and answers were digitized (22.05 kHz, 16-bit resolution) and normalized for peak intensity (50% of the full scale). The utterances were 500 milliseconds (ms) apart from one another in each sequence.

The question-answer sequences were presented to the participants using a laptop computer and loudspeakers. The participants wore a head-mounted Shure microphone (model SM 10A) connected to a Sony digital audio tape recorder (model TCD-D8). The order of presentation of the target utterances was fixed and the set of ten question-answer sequences was repeated twice.

The seven utterances analyzed in this study were: *In the United States* /ɪn ðə ˌjunaɪ̯təd ˈstets/, *Ten o'clock* /tɛn ɔˈklɑk/, *He went to work* /hi ˌwɛntu ˈwɝk/, *A glass of water* /ə ˌglæs əv ˈwɑtɚ/, *She ate a sandwich* /ʃi ˌet ə ˈsændwɪtʃ/, *I read a good book* /aɪ̯ ˌrɛd ə gʊd ˈbʊk/, and *She is eight years old* /ʃi ɪz ˈet jirz old/. The other three utterances (*I'm fine*, *Five dollars*, *They went to school*) were excluded because they were reported in [19]. A total of 448 utterances were analyzed in this study (7 utterances x 16 participants x 2 groups x 2 times).

2.3. Analysis

2.3.1. Duration measurements

Utterance durations, vowel durations of four vowels (/ɪ/, /æ/, /e$^{\text{ɪ}}$/, and /ʊ/), consonant intervals, and vocalic intervals were measured using wideband spectrograms produced by [23]. Standard segmentation criteria in [24], [25] as well as practices on rhythm measures in previous studies such as [26]–[28] were followed as closely as possible when measuring consonantal and vocalic intervals. The four vowels appeared in the following words: /ɪ/ (*in*, *sandwich*), /æ/ (*glass*, *sandwich*), /e$^{\text{ɪ}}$/ (*states*, *eight*), /ʊ/ (*good*, *book*). Utterance durations were calculated as the sum of all of the consonantal and vocalic intervals in the utterance. There were no pauses within the utterances.

Some additional considerations were made due to segmental characteristics and speaker variations. The glide /j/ between the vowel segments in "the" and "United" in utterance 2 was counted as part of the vocalic interval because it was difficult to segment it accurately from the neighboring vowel segments. Some speakers also produced approximants in place of a stop or in between vowels (e.g., /j/ in between "she" and "ate" in utterance 8). These approximants were also measured as part of the vocalic interval. In addition, speakers occasionally omitted a vowel or a consonant (e.g., /ə/ for the indefinite article "a"). As a result, these participants' values were based on fewer segments or intervals. Most importantly, the same measurement criteria were followed consistently for all of the utterances.

2.3.2. Rhythm metrics

As proportional measures of temporal aspects of speech, four different rhythm metrics were calculated for each utterance: the raw and normalized Pairwise Variability Index (PVI) (rPVI and nPVI) [27], [29], %V [30], and VarcoV [28]. Four rhythm metrics were used because one measure may be more sensitive to differences than others [26].

rPVI [27], [29] was calculated as follows:

$$rPVI = \left[\sum_{k=1}^{m-1} |d_k - d_{k+1}| / (m-1)\right] \qquad (1)$$

(where *m* is the number of intervals and *d* is the duration of the *k*th interval)

nPVI [27] was calculated as follows:

$$nPVI = 100 \text{ x } \left[\sum_{k=1}^{m-1} \left|\frac{d_k - d_{k+1}}{(d_k + d_{k+1})/2}\right| / (m-1)\right] \qquad (2)$$

%V was calculated as the sum of vocalic interval duration divided by the total duration of vocalic and consonant intervals [30]. VarcoV was calculated as standard deviation of vocalic interval duration divided by mean vocalic interval duration, multiplied by 100 [28].

A single value of each metric was calculated separately for each utterance for all participants first. The values for the seven utterances were then averaged for each metric and participant at T1 and T2.

2.3.3. F_0 measurements

The F_0 values were measured using the auto-correlation pitch tracker in [31]. The automatic pitch extraction function produced F_0 values in Hz at 10-ms intervals. The data points were then manually examined for any apparent errors. Six common F_0 measures, average speaking F_0, average vowel F_0, maximum F_0 (Max F_0), minimum F_0 (Min F_0), range in Hz, and range in semitones, were used following [2], [32], [33].

The average speaking F_0 was calculated in the same four vowels (eight tokens total) selected for duration measurements. The values for the first and last 10 ms of the vowels were excluded to avoid F_0 artifacts at segmental edges [2], [33]. Then the remaining values were averaged for each vowel. The average values for the eight vowel tokens were then averaged together to compute an overall speaking F_0 for each participant at T1 and T2. Both average F_0 for each vowel (averaged over two tokens) and average speaking F_0 (averaged over all vowel tokens) were analyzed.

Next, the highest F_0 value (Max F_0) and the lowest F_0 values (Min F_0) were identified in each utterance. The range in Hz was calculated as Min F_0 subtracted from Max F_0. The range in Hz was converted into semitones using the following formula [33]:

$$F_0 range\ in\ semitones = 39.863 \text{ x } \log(\frac{MaxF_0}{MinF_0}) \qquad (3)$$

Finally, the range in Hz and range in semitones were averaged over the seven utterances for each speaker at T1 and T2.

3. Results

3.1. Temporal aspects

3.1.1. Utterance and segmental durations

Table 1 shows the means and standard deviations for the utterance and vowel durations for the adults and children at T1 and T2. The utterance durations were analyzed using a two-way Analysis of Variance (ANOVA). Age (2 levels) was a between-subjects variable and Time (2 levels) was a within-subjects variable. This analysis yielded a significant main effect of Age, $F_{(1,30)} = 4.83, p = 0.036, \eta_p^2 = 0.14$. The main effect of Time, as well as the two-way interaction, were not significant, $F_{(1,30)}$ = 2.82 and 0.93, p = .10 and 0.86, η_p^2 = 0.09 and 0.03.

Table 1. *Duration measures.*

	Adult T1	Adult T2	Child T1	Child T2
Utterance	1026.33	999.06	1074.54	1067.18
(SD)	(71.64)	(80.27)	(89.92)	(78.44)
/ɪ/	68.13	64.17	56.88	59.64
(SD)	(9.39)	(12.69)	(14.25)	(14.38)
/e^{ɪ}/	182.13	174.97	179.25	173.08
(SD)	(22.57)	(22.46)	(40.93)	(13.23)
/æ/	129.25	125.28	124.88	119.62
(SD)	(14.13)	(17.47)	(18.58)	(22.57)
/ʊ/	112.29	109.35	108.02	106.55
(SD)	(19.26)	(18.24)	(17.12)	(21.87)

The vowel durations were analyzed using a three-way ANOVA. Age (2 levels) was a between-subjects variable, while Time (2 levels) and Vowel (4 levels) were within-subjects variables. The three-way ANOVA yielded a significant main effect of Vowel, $F_{(3,28)} = 99.32, p < 0.001, \eta_p^2 = 0.77$. The main effect for Time, Age, and all two-way and three-way interactions were not significant, $F_{(1,30)}$ = 0.03 to 2.08, p = 0.16 to 0.91, η_p^2 = 0.001 to 0.07. The post-hoc analysis indicated that the duration of the vowels differed significantly (from the longest to the shortest, /e^{ɪ}/ > /æ/ > /ʊ/ > /ɪ/) ($p < 0.05$).

Overall, the results indicated that the children's utterances were longer than the adults' both at T1 and T2. The durations of the vowels differed, both in adults and children, likely due to intrinsic vowel durations and stress patterns. The vowel duration patterns were consistent with previous reports on vowel duration such as [2], [5], [25].

3.1.2. Rhythm metrics

The means and standard deviations of the four rhythm metrics for the adults and children are shown in Table 2. The values of rPVI, nPVI, %V, and VarcoV were analyzed in a series of two-way ANOVAs. Age (2 levels) was a between-subjects variable and Time (2 levels) was a within-subjects variable. Out of all the statistical analysis, the only significant effect was the main effect of Time for %V, $F_{(1,30)} = 4.32, p = 0.046, \eta_p^2 = 0.13$. For all other statistical tests, the main effects of Time and Age, as well as the two-way interactions were not significant, $F_{(1,30)}$ = 0.03 to 2.34, p = 0.17 to 0.88, η_p^2 = 0.001 to 0.06. These results indicate that rhythm metric values did not differ between adults and children at either T1 or T2, except for the marginally significant difference between T1 and T2 for %V.

Table 2. *Rhythm measures.*

	Adult T1	Adult T2	Child T1	Child T2
rPVI	74.49	76.49	75.43	74.92
(SD)	(7.10)	(7.37)	(8.05)	(9.50)
nPVI	53.82	54.58	57.57	54.39
(SD)	(5.24)	(7.52)	(11.55)	(5.47)
%V	0.42	0.43	0.41	0.42
(SD)	(0.02)	(0.03)	(0.03)	(0.04)
VarcoV	43.67	44.91	47.77	45.14
(SD)	(2.79)	(5.22)	(7.12)	(5.94)

3.2. F_0 analysis

3.2.1. Absolute measures

The means and standard deviations for the adult and child average speaking F_0, Max F_0, Min F_0, and individual vowels are shown in Table 3. The values were differentiated between males and females for the F_0 analysis, because adult females have higher F_0 than adult males due to differences in membranous length of the vocal folds [34].

The average speaking F_0, Max F_0, and Min F_0 values were analyzed in a series of three-way ANOVAs. Age (2 levels) and Sex (2 levels) were between-subjects variables, and Time (2 levels) was a within-subjects variable. For all three measures (average speaking F_0, Max F_0, and Min F_0), ANOVAs yielded significant main effects of Age and Sex, and interaction between Age and Sex, $F_{(1,28)}$ = 4.74 to 55.33, p = 0.001 to .038, η_p^2 = 0.15 to 0.66. The main effect of Time as well as interactions involving Time (Time x Age, Time x Sex, and the three-way interaction) were not significant for these absolute F_0 measures, $F_{(1,28)}$ = 0.002 to 1.10, p = 0.30 to 0.97, η_p^2 = 0.001 to 0.17. The two-way interaction was significant because F_0 values were higher in adult females than adult males, but they were not different between male and female children. These statistical analyses indicate that the average speaking F_0, Max F_0, and Min F_0 were higher in children than in adults, and that they were higher in females than in males. They also indicated that average F_0, Max F_0, and Min F_0 were higher in the adult females than in the adult males, but they did not differ between the male and female children.

The average vowel F_0 values were analyzed using a four-way ANOVA. Age (2 levels) and Sex (2 levels) were between-subjects variables and Time (2 levels) and Vowel (4 levels) were within-subjects variables. The four-way ANOVA yielded a significant main effect for Age, Sex and Vowel, $F_{(3,26)}$ = 5.86

to 49.50, p = 0.001 to 0.04, η_p^2 = 0.15 to 0.67. The two-way interaction between Age and Sex was significant, $F_{(3,26)}$ = 28.37, p = 0.001, η_p^2 = 0.52. The main effect of Time, and all other two-way, three-way, and four-way interactions were not significant, $F_{(3,26)}$ = 0.03 to 1.36, p = 0.27 to 0.96, η_p^2 = 0.001 to 0.05.

These results indicated that the average vowel F_0 was higher in the adult females than in the adult males, but it did not differ between the male and female children. The post-hoc analysis indicated that the average F_0 of /ɪ/ was significantly higher than that of /e$^{\text{ɪ}}$/ and /ʊ/, and the average F_0 of /e$^{\text{ɪ}}$/ was significantly lower than that of /æ/.

Table 3. *Absolute F_0. T1 and T2 were averaged.*

	Adult males	Adult females	Child males	Child females
Ave F_0	104.33	176.97	196.23	189.48
(SD)	(19.07)	(26.25)	(33.13)	(33.09)
Max F_0	155.59	226.94	235.88	248.49
(SD)	(60.13)	(26.06)	(37.06)	(27.02)
Min F_0	86.11	128.70	150.58	143.23
(SD)	(9.01)	(33.20)	(37.17)	(42.41)
/ɪ/ F_0	109.55	181.34	203.68	200.70
(SD)	(17.62)	(34.71)	(35.82)	(41.91)
/e$^{\text{ɪ}}$/ F_0	100.34	174.16	194.97	189.47
(SD)	(15.55)	(24.88)	(30.96)	(39.85)
/æ/ F_0	103.35	177.60	199.21	189.92
(SD)	(18.98)	(16.79)	(32.63)	(25.26)
/ʊ/ F_0	105.91	175.15	188.80	189.47
(SD)	(22.83)	(26.92)	(32.98)	(17.57)

3.2.2. Pitch range

The means and standard deviations for the adult and child pitch range in Hz and semitones are shown in Table 4. The pitch range was analyzed using three-way ANOVAs. Age (2 levels) and Sex (2 levels) were between-subjects variables, and Time (2 levels) was a within-subjects variable. The only significant difference was the effect of Sex in the range in Hz, $F_{(1,28)}$ = 10.04, p = 0.004, η_p^2 = 0.26. All other main effects, two-way and three-way interactions were not significant for the range in Hz and range in semitones, $F_{(1,28)}$ = 0.01 to 3.68, p = 0.065 to 0.99, η_p^2 = 0.001 to 0.12. The significant effect of Sex for the pitch range in Hz was due to a greater pitch range in females than in males.

Table 4. *Pitch range. T1 and T2 were averaged.*

	Adult males	Adult females	Child males	Child females
Hz	69.48	98.24	84.48	108.54
(SD)	(35.03)	(24.07)	(17.24)	(30.10)
semitones	6.42	8.09	5.76	7.24
(SD)	(4.15)	(3.63)	(3.14)	(4.75)

In sum, the absolute average F_0 measures showed differences between adult males and adult females, and between adults and children. Pitch range in Hz was also greater in females than in males. However, no statistically significant differences were found between adults and children when pitch was analyzed using a proportional measure (i.e., pitch range in semitones).

4. Discussion

This study investigated speech prosody in school-age children in American English. Overall, the results showed that there were some differences between adults and children in absolute measures of temporal and pitch aspects of speech prosody, but these differences diminished in proportional measures. For temporal aspects of speech, the absolute durations of children's utterances were longer than the adults' utterances, whereas rhythm metrics showed no statistically significant differences between adults and children. Absolute F_0 values were higher in children than in adults, but the pitch range in semitones did not differ between adults and children.

The results on the absolute measures are generally comparable with those reported on temporal and pitch aspects of adults' and children's speech [1]–[6]. The proportional measures, rhythm metrics and pitch range, showed no statistically significant differences between adults and children, unlike those suggested by [9] and [19]. Grabe et al. [9] studied 4-year-olds, whereas the samples in this study were from school-age children. The acquisition of rhythm in English has been reported to take longer due to its complexity [8, 9, 10, 13], but the differences may be in the interaction of prosodic features and other domains such as syntax [14]–[16]. It is also possible that differences were not found between adults and children in this study because the participants repeated the target utterances after an auditory model. It is likely that the adults and children spoke in a more similar manner to each other than they would in a natural setting.

The children and adults in the current study were the participants in the control groups in [19]–[22]. [20]–[22] also reported some developmental changes and differences between adults and children at the segmental level in native English-speaking participants. In vowel production, [22] found that F2 frequencies decreased from T1 to T2 for /i/ and /ɛ / in children. In consonant production, the children's intelligibility scores for the production of /r/, /l/, and /w/ were high (> 94.1%) [20], whereas their scores for /f/, /s/, and /θ/ were lower than the adults' scores at both T1 and T2 [21]. In sum, the results from this study and the previous studies [19]–[22] suggest that some developmental changes still occur during the school-age years, but they may be limited to the fine tuning of segmental productions and absolute values of temporal and pitch aspects of speech prosody.

5. Acknowledgements

This study was supported by National Institutes of Health [DC00257] to UAB. Dr. Susan Guion-Anderson played a major role in the planning and implementation of the original longitudinal project at UAB. She designed the experiments with the third author and collected much of the data. She was also involved in many sub-projects that resulted from the original study until her death in 2011. The authors would like to thank the participants of this study. We also thank L. Legault, R. Froehlich, and K. Chapman for their assistance.

6. References

[1] R. D. Kent and L. L. Forner, “Speech segment durations in sentence recitations by children and adults,” *J. Phonetics*, vol. 8, pp. 157–168, 1980.

[2] S. Lee, A. Potamianos, and S. Narayanan, “Acoustics of children’s speech: Developmental changes of temporal and spectral parameters,” *J. Acoustical Soc. America*, vol. 105, no. 3, pp. 1455–1468, 1999.

[3] B. L. Smith, “Temporal aspects of English speech production: A developmental perspective,” *J. Phonetics*, vol. 6, pp. 37–67, 1978.

[4] B. L. Smith, “Effects of experimental manipulations and intrinsic contrasts on relationships between duration and temporal variability in children’s and adults’ speech,” *J. Phonetics*, vol. 22, pp. 155–175, 1994.

[5] J. Hillenbrand, L. A. Getty, M. J. Clark, and K. Wheeler, “Acoustic characteristics of American English vowels,” *J. Acoustical Soc. America*, vol. 97, no. 5, pp. 3099–3111, 1995.

[6] E. T. Stathopoulos, J. E. Huber, and J. E. Sussman, “Changes in acoustic characteristics of the voice across the life span: Measures from individuals 4–93 years of age,” *J. Speech, Language, and Hearing Research*, vol. 54, no. 4, pp. 1011–1021, 2011.

[7] G. D. Allen and S. Hawkins, “The development of phonological rhythm,” in *Syllables and Segments*, A. Bell and J. B. Hooper, Eds. Amsterdam, Netherlands: North-Holland, 1978, pp. 173–185.

[8] G. D. Allen and S. Hawkins, “Phonological rhythm: Definition and development," in *Child Phonology: Volume 1, Production*, G. H. Yeni-Komshian, J. F. Kavanagh and C. A. Ferguson, Eds. New York, NY: Academic Press, 1980, pp. 227–256.

[9] E. Grabe, B. Post and I. Watson, “The acquisition of rhythm in English and French,” in *Proc. of the International Congress of Phonetic Sciences, San Francisco, USA,* 1999, pp. 1201–1204.

[10] M. Ordin and L. Polyanskaya, “Acquisition of English speech rhythm by monolingual children,” in *Proc. INTERSPEECH-2015, 16th Annual Conference of the International Speech Communication Association, September 6–10, Dresden, Germany,* 2015, pp. 3120–3124.

[11] C. T. Ferrand and R. L. Bloom, “Gender differences in children’s intonational patterns,” *J. Voice,* vol. 10, no. 3, pp. 284–291, 1996.

[12] M. A. Redford, “The acquisition of temporal patterns,” in *The Handbook of Speech Production*, M. A. Redford, Ed. Boston: Wiley–Blackwell, 2015, pp. 379–403.

[13] H. Sirsa and M. A. Redford, “Towards understanding the protracted acquisition of English rhythm,” in *Proc. of the 17th International Congress of Phonetic Sciences, Hong Kong,* 2011, pp. 1862–1865.

[14] B. Wells, S. Peppé, and N. Goulandris, “Intonation development from five to thirteen,” *J. Child Language,* vol. 31, pp. 749–778, 2004.

[15] I. A. Shport and M. A. Redford, “Lexical and phrasal prominence patterns in school-aged children’s speech,” *J. Child Language,* vol. 41, pp. 890–912, 2014.

[16] W. F. Katz, C. M. Beach, K. Jenouri, and S. Verma, “Duration and fundamental frequency correlates of phrase boundaries in production by children and adults,” *J. Acoustical Soc. America*, vol. 99, no. 5, pp. 3179–3191, 1996.

[17] M. A. Redford, “The perceived clarity of children’s speech varies as a function of their default articulation rate,” *J. Acoustical Soc. America,* vol. 135, pp. 2952-2963, 2014.

[18] S. Nittrouer, “The emergence of mature gestural patterns is not uniform: Evidence from an acoustic study,” *J. Speech and Hearing Research*, vol. 36, no. 5, pp. 959–972, 1993.

[19] K. Aoyama and S. G. Guion, “Prosody in second language acquisition: An acoustic analysis on duration and F0 range,” in *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, O. Bohn and M. J. Munro, Eds. Amsterdam, Netherlands: John Benjamins, 2007, pp. 281–297.

[20] K. Aoyama, J. E. Flege, S. G. Guion, R. Akahane-Yamada, and T. Yamada, “Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/,” *J. Phonetics*, vol. 32, no. 2, pp. 233–250, 2004.

[21] K. Aoyama, S. G. Guion, J. E. Flege, T. Yamada, and R. Akahane-Yamada, “The first years in an L2-speaking environment: A comparison of Japanese children and adults learning American English,” *Int. Review Appl. Linguistics in Language Teaching*, vol. 46, pp. 61–90, 2008.

[22] G. E. Oh, S. Guion-Anderson, K. Aoyama, J. E. Flege, R. Akahane-Yamada, and T. Yamada, “A one-year longitudinal study of English and Japanese vowel production by Japanese adults and children in an English-speaking setting,” *J. Phonetics*, vol. 39, pp. 156–167, 2011.

[23] P. Milenkovic, *TF32* [Computer program]. 2002.

[24] R. D. Kent and C. Read, *Acoustic Analysis of Speech*, 2nd ed. Albany, NY: Delmar, 2002.

[25] G. E. Peterson and I. Lehiste, “Duration of syllable nuclei in English,” *J. Acoustical Soc. America*, vol. 32, pp. 693–703, 1960.

[26] A. Arvaniti, “The usefulness of metrics in the quantification of speech rhythm,” *J. Phonetics,* vol. 40, pp. 351–373, 2012.

[27] E. Grabe and E. Low, “Durational variability in speech and the rhythm class hypothesis,” in *Laboratory Phonology 7*, C. Gussenhoven and N. Warner, Eds. Berlin, Germany: Mouton de Gruyter, 2002, pp. 515–546.

[28] L. White and S. L. Mattys, “Rhythmic typology and variation in first and second languages,” in *Segmental and Prosodic Issues in Romance Phonology*, P. Prieto, J. Mascaró and M. Solé, Eds. Amsterdam, Netherlands: John Benjamins, 2007, pp. 237–257.

[29] E. Low, E. Grabe and F. Nolan, “Quantitative characterizations of speech rhythm: Syllable-timing in Singapore English,” *Language and Speech*, vol. 43, no. 4, pp. 377–401, 2000.

[30] F. Ramus, M. Nespor, and J. Mehler, “Correlates of linguistic rhythm in the speech signal,” *Cognition*, vol. 73, pp. 265–292, 1999.

[31] Scicon R&D, *Pitchworks* [Computer program], 2011.

[32] R. J. Baken and R. F. Orlikoff, *Clinical Measurement of Speech and Voice*, 2nd ed. San Diego, CA: Singular, 2000.

[33] P. Keating and G. Kuo, “Comparison of speaking fundamental frequency in English and Mandarin,” *J. Acoustical Soc. America*, vol. 132, no. 2, pp. 1050–1060, 2012.

[34] I. R. Titze, “Physiologic and acoustic differences between male and female voices,” *J. Acoustical Soc. America*, vol. 85, pp. 1699–1707, 1989.