# Temporal Organization of Syllable Production in the Speech of Cantonese-speaking Children

*Wai-Sum Lee and Eric Zee*

Department of Chinese, Translation and Linguistics, City University of Hong Kong, Hong Kong

w.s.lee@cityu.edu.hk; eric.zee@cityu.edu.hk

## Abstract

The paper investigates the temporal structures of three types of syllables, CV:, CVN, and CVV (C = [p $p^h$ ts $ts^h$ s], V: = [a:], VN = [an], VV = [ai]), in Cantonese. Results of a temporal analysis of the speech samples from two 9- to 10-year-old Cantonese-speaking children are as follows. First, the durations of the syllable-initial consonants and the following rime are not negatively correlated, as there is no temporal compensation involved within the sequence of the initial consonant and rime. It is assumed that the initial consonant and rime within a syllable are not programmed as a single articulatory unit (Lehiste [10]). The rime however has a tendency to be slightly shorter after the aspirated [$p^h$] and [$ts^h$] than after the unaspirated counterparts [p] and [ts]. Second, the duration of the vowel [a] of [an] and [ai] is about one half of the duration of the vowel [a:] of CV: syllables, suggesting the occurrence of compensatory shortening of the vowel duration when [a:] occurs in other syllable types. This in turn suggests that the rime [an] and the diphthong [ai] are programmed as a single articulatory unit. Third, the temporal data in this study support Lehiste's contention ([10]) that there is a closer connection between a vowel and a following consonant than between an initial consonant and a following vowel. Lastly, the patterns of temporal organization of syllable production for children are similar to those for adults, indicating that children of middle childhood at 9 to 10 years of age have acquired the adult-like temporal patterns.

**Index Terms**: temporal organization, syllable production, Cantonese, children of middle childhood

## 1. Introduction

There have been a number of studies of the temporal characteristics of children's speech ([1, 2, 3, 4, 5, 6, 7, 9, 12, 13]). Speech data examined in a large majority of the past studies were from English-speaking children. This study investigates the temporal organization of syllable production in the speech of Cantonese-speaking children, more specifically, the patterns of temporal structures of different types of syllables in Hong Kong Cantonese. It also determines whether temporal compensation takes place amongst the successive component segments of the syllables of different types, that is, whether the lengths of the component segments in a syllable are inversely correlated in order to maintain an invariant duration of the syllable ([8, 10, 11]).

## 2. Method

### 2.1. Test materials

The study investigates the temporal organization of three types of Cantonese syllables, CV:, CVN, and CVV, where C = [p $p^h$ ts $ts^h$ s], V: = [a:], VN = [an], and VV = [ai]. Each syllable type was associated with three different tones, namely a high level [55], a high rising [25], and a low falling [21]. Table 1 presents 35 test monosyllabic words used in this study. All the test syllables are meaningful monosyllabic words and commonly used in everyday speech. The monosyllables which are non-occurring in Cantonese (indicated with '—' in the table) were not included in the study. A randomized list of the 35 test words was prepared for eliciting speech samples from speakers.

Table 1. *Test monosyllabic words for this study ('—' = syllables non-occurring in Cantonese).*

| C | V: = [a:] | VN = [an] | VV = [ai] |
|---|---|---|---|
| [p] | [pa:$^{55}$] 'bus' | [pan$^{55}$] 'squad' | [pai$^{55}$] 'to bye' |
| | [pa:$^{25}$] 'to hold' | [pan$^{25}$] 'slab' | [pai$^{25}$] 'to shake' |
| | [pa:$^{21}$] 'father' | — | — |
| [$p^h$] | [$p^h$a:$^{55}$] 'to lie prone' | [$p^h$an$^{55}$] 'to climb' | [$p^h$ai$^{55}$] 'sect' |
| | [$p^h$a:$^{25}$] 'steak' | — | [$p^h$ai$^{25}$] 'tablet' |
| | [$p^h$a:$^{21}$] 'to crawl' | — | [$p^h$ai$^{21}$] 'to line up' |
| [ts] | [tsa:$^{55}$] 'residue' | — | [tsai$^{55}$] 'vegetarian' |
| | [tsa:$^{25}$] 'unskilled' | [tsan$^{25}$] 'a classifier' | — |
| | [tsa:$^{21}$] 'a particle' | — | — |
| [$ts^h$] | [$ts^h$a:$^{55}$] 'fork' | [$ts^h$an$^{55}$] 'meal' | [$ts^h$ai$^{55}$] 'to guess' |
| | — | [$ts^h$an$^{25}$] 'to produce' | [$ts^h$ai$^{25}$] 'to step' |
| | [$ts^h$a:$^{21}$] 'tea' | [$ts^h$an$^{21}$] 'withered' | [$ts^h$ai$^{21}$] 'brushwood' |
| [s] | [sa:$^{55}$] 'sand' | [san$^{55}$] 'hill' | [sai$^{55}$] 'to waste' |
| | [sa:$^{25}$] 'to sprinkle' | [san$^{25}$] 'scattered' | [sai$^{25}$] 'to move' |
| | [sa:$^{21}$] 'to steal' | [san$^{21}$] 'weak' | — |

### 2.2. Speakers

Speech data were provided by two 9- to 10-year-old male Cantonese-speaking children and also two male Cantonese-speaking adults in their early twenties for comparison. All the speakers were born and grew up in monolingual families in Hong Kong and none had history of speech or hearing problems.

### 2.3. Data collection and analysis

The speakers were instructed to utter the test monosyllabic words in the carrier sentence, [ŋɔ$^{23}$ jiu$^{33}$ tɔk$^{2}$ ___ pei$^{25}$ nei$^{23}$ $t^h$ɛŋ$^{55}$] 'I want (to) read ___ for you (to) listen', at a normal speech rate. Five repetitions of each of the 35 test monosyllables were recorded of each speaker. Digital audio recordings were performed in a sound-proof booth.

The speech analysis software CSL (Computerized Speech Lab) was used for measurements of the durations of the component elements of the test syllables. The measurements were directly made on the speech waveforms, making references to the synchronized LPC formant trajectories and wide-band spectrograms of the speech signals when necessary. For CV: syllables, the durations of the syllable-initial consonant and the following vowel were separately measured. For CVN syllables, duration measurements of the vowel-to-nasal transition and final nasal were also made. For CVV

syllables, the durations of the initial consonant, the two vowel elements, and the vowel-to-vowel transition were measured.

## 3. Results

In this section, the temporal structures of the test syllables CV:, CVN, and CVV are presented with bar charts for two male children and two male adults. For each speaker, the duration of a component segment in the syllables is the mean of five repetitions. On the charts, the durations of the initial consonants are represented in negative values, whereas the durations of the component segments in the rimes are in positive values.

### 3.1. CV: syllables

Fig. 1a to Fig. 1d show the temporal structures of CV: syllables, where C = [p pʰ ts tsʰ s] (in empty bars) and V: = [aː] (in dark bars), for four speakers, two male children (upper) and two male adults (lower).

Fig. 1a: Male Child 1

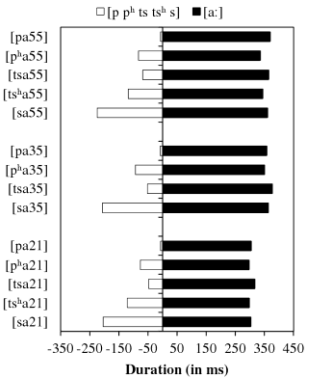


Fig. 1b: Male Child 2

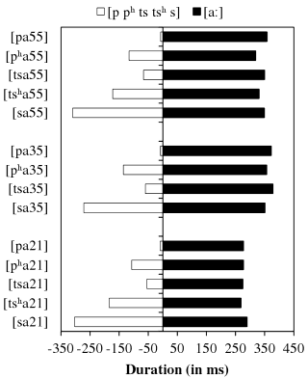


Fig. 1c: Male Adult 1

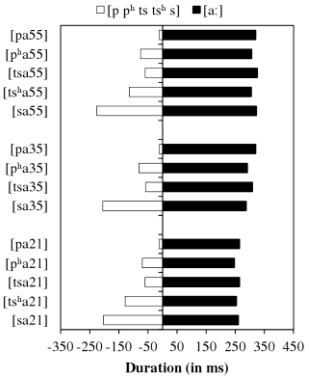


Fig. 1d: Male Adult 2

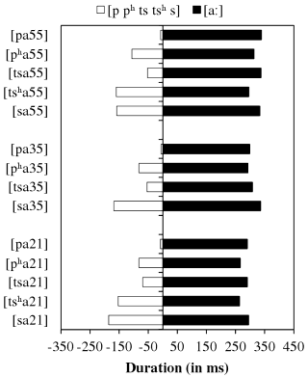


Fig. 1a-1d: *Temporal structures of CV: syllables, where C = [p pʰ ts tsʰ s] (in empty bars) and V: = [aː] (in dark bars), for two male children and two male adults.*

As shown in Fig. 1a to Fig. 1d for all the speakers, children and adults, the durations of the syllable-initial consonants do not vary according to the tone type. For both children and adults, the fricative [s] (253.62 ms and 191.47 ms) is longest, to be followed by the aspirated [tsʰ] (149.24 ms and 139.35 ms), aspirated [pʰ] (102.00 ms and 83.03 ms), unaspirated [ts] (58.47 ms and 59.4 ms), and unaspirated [p] (8.09 ms and 9.77 ms) in decreasing order.

As for the duration of the rime or the vowel [aː] of CV: syllables, variations are observed in different tonal contexts. For children (Fig. 1a and Fig. 1b) and adults (Fig. 1c and Fig. 1d), the duration of the vowel [aː] is shorter when associated with a falling [21] tone (290.62 ms and 269.34 ms) than associated with a high level [55] tone (348.05 ms and 319.72 ms) and a high rising [25] tone (363.35 ms and 305.20 ms).

For children and adults, there is a tendency for the duration of the vowel [aː] to be slightly shorter after the aspirated [pʰ] (322.67 ms and 285.80 ms) and [tsʰ] (310.16 ms and 279.45 ms) than after the unaspirated counterparts [p] (339.88 ms and 305.17 ms) and [ts] (326.36 ms and 304.40 ms). The reduction of the vowel duration after the aspirated [pʰ] and [tsʰ] is much smaller than the difference in duration between the aspirated [pʰ tsʰ] and the unaspirated [p ts], which suggests that in CV: syllables temporal compensation does not take place between the syllable-initial consonant and the following vowel. The duration of the entire syllable appears to be positively correlated with the duration of the syllable-initial consonant.

### 3.2. CVN syllables

Fig. 2a to Fig. 2d show the temporal structures of CVN syllables, where C = [p pʰ ts tsʰ s] (in empty bars) and VN = [an] consisting of a vowel [a] (in dark bars), a vowel-to-nasal (V-N) transition (in dotted bars), and a final nasal [n] (in grey bars), for two male children (upper) and two male adults (lower).

Fig. 2a: Male Child 1

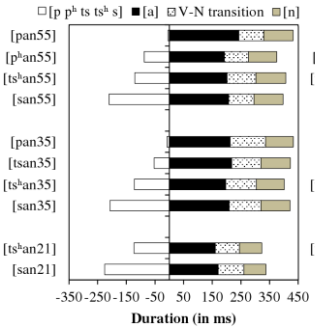


Fig. 2b: Male Child 2

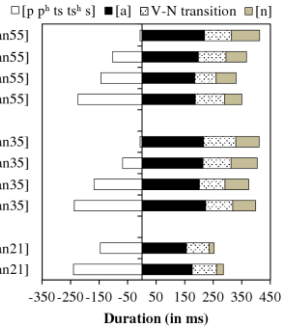


Fig. 2c: Male Adult 1

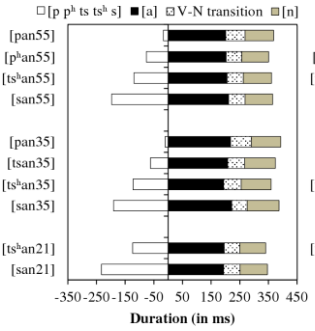


Fig. 2d: Male Adult 2

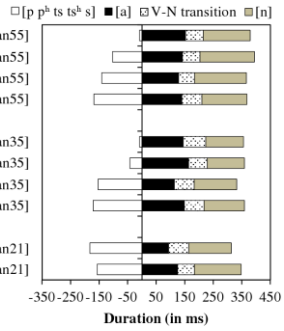


Fig. 2a-2d: *Temporal structures of CVN syllables, where C = [p pʰ ts tsʰ s] (in empty bars) and VN = [an] consisting of a vowel [a] (in dark bars), a vowel-to-nasal (V-N) transition (in dotted bars), and a final nasal [n] (in grey bars), for two male children and two male adults.*

As can be seen in Fig. 2a to Fig. 2d, similar to CV: syllables, there are no large variations in the durations of the syllable-initial consonants of CVN syllables associated with different tones. Regardless of the tone type, the duration is shortest for [p], to be followed by [ts], [pʰ], [tsʰ], and [s] in increasing order for both children (6.83 ms, 60.75 ms, 95.54 ms, 137.62 ms, and 224.49 ms, respectively) and adults (11.70 ms, 52.30 ms, 89.90 ms, 140.93 ms, and 186.80 ms, respectively).

As for the rime [an] of CVN syllables, it is similar to the vowel [a:] of CV: syllables, in that the rime duration varies with the tonal type. For both children (Fig. 2a and Fig. 2b) and adults (Fig. 2c and Fig. 2d), a shorter duration is observed for the rime [an] when associated with a [21] tone (299.31 ms and 336.55 ms) than with a [55] tone (383.39 ms and 368.50 ms) and a [25] tone (407.84 ms and 364.60 ms). The rime [an] also tends to be slightly shorter after the aspirated [pʰ tsʰ] (378.9 ms for children and 358.65 ms for adults) than after the unaspirated [p ts] (417.40 ms for children and 369.85 ms for adults). The difference is more pronounced for children than adults, and there is an exception for Male Adult 2 (Fig. 2d), in which the rime [an] when associated with a [55] tone is instead slightly longer after the aspirated [pʰ] than after the unaspirated [p]. In any case, no compensatory lengthening or shortening of the preceding initial consonant is observable.

As for the component segments of the rime [an], the vowel [a] and the final nasal [n] contribute more than the vowel-to-nasal transition to the duration reduction of the rime when associated with a [21] tone in the speech of children. For the two children (Fig. 2a and Fig. 2b), the durations of the vowel [a] and the final [n] of the rime when associated with a [21] tone (165.64 ms and 49.35 ms) are noticeably shorter than with a [55] tone (203.76 ms and 88.64 ms) and a [25] tone (210.87 ms and 92.13 ms). As for the vowel-to-nasal transition, it is only slightly shorter in [21] tone (84.32 ms) than in [55] tone (91.01 ms) and [25] tone (104.84 ms). For adults (Fig. 2c and Fig. 2d), the reduction of rime duration of CVN syllables with a [21] tone is mainly contributed by the vowel [a] of the rime with a duration of 151.30 ms in [21] tone which is shorter than [a] in [55] tone (175.53 ms) and [25] tone (176.05 ms). As for the vowel-to-nasal transition and the final [n] of the rime [an], the durations are similar in [21] tone (60.65 ms and 124.60 ms), [55] tone (60.45 ms and 135.53 ms), and [25] tone (66.08 ms and 122.48 ms).

In the speech of two children (Fig. 2a and Fig. 2b), the duration reduction of the rime [an] preceded by an aspirated initial consonant is also contributed by the shortening of the vowel [a] and the final [n] of the rime, with a larger contribution from [a] than [n]. The durations of [a] and [n] after the aspirated [pʰ tsʰ] are 196.82 ms and 87.77 ms which are shorter than [a] (223.06 ms) and [n] (98.99 ms) after the unaspirated [p ts]. The duration of the vowel-to-nasal transition of the rime [an] is similar after [pʰ tsʰ] (94.32 ms) and [p ts] (95.35 ms). For adults (Fig. 2c and Fig. 2d), the duration reduction of the rime after [pʰ tsʰ] is contributed solely by the vowel [a]. The respective durations of the vowel [a], vowel-to-nasal transition, and final [n] of the rime [an] are 162.35 ms, 61.60 ms, and 134.70 ms after [pʰ tsʰ] vs. 181.00 ms, 63.15 ms, and 125.70 ms after [p ts].

### 3.3. CVV syllables

Fig. 3a to Fig. 3d show the temporal structures of CVV syllables, where C = [p pʰ ts tsʰ s] (in empty bars) and VV = the diphthong [ai] consisting of a vowel [a] (in dark bars), a vowel-to-vowel (V-V) transition (in dotted bars), and a vowel [i] (in grey bars), for two children (upper) and two adults (lower).

Fig. 3a: Male Child 1 Fig. 3b: Male Child 2

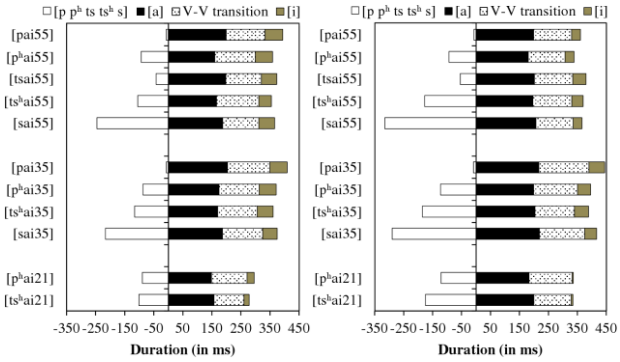


Fig. 3c: Male Adult 1 Fig. 3d: Male Adult 2

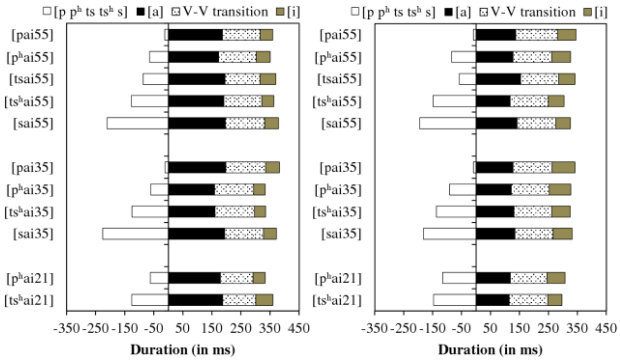


Fig. 3a-3d: *Temporal structures of CVV syllables, where C = [p pʰ ts tsʰ s] (in empty bars) and VV = [ai] consisting of a vowel [a] (in dark bars), a vowel-to-vowel (V-V) transition (in dotted bars), and a vowel [i] (in grey bars), for two male children and two male adults.*

A comparison of Fig. 1a to Fig. 1d, Fig. 2a to Fig. 2d, and Fig. 3a to Fig. 3d shows that the patterns of the temporal structures for CV:, CVN, and CVV syllables are similar. Firstly, for children and adults, the duration of the rime or the diphthong [ai] of CVV syllables is shorter when associated with a [21] tone (310.60 ms and 324.00 ms) than with a [55] tone (365.65 ms and 346.24 ms) and a [25] tone (394.53 ms and 343.25 ms). For children, the duration reduction of [ai] with a [21] tone is contributed by both the durations of the first vowel element [a] and the second vowel element [i], but not the vowel-to-vowel transition, of the diphthong. As for adults, however, the contribution to the reduction of the duration of the diphthong [ai] is mainly from the first vowel element [a].

Secondly, for children and adults, the duration of the diphthong [ai] is also slightly shorter after the aspirated [pʰ tsʰ] (363.66 ms and 333.80 ms) than after the unaspirated [p ts] (388.76 ms and 356.15 ms), and the reduction of the duration of [ai] after an aspirated consonant is mainly contributed by the first vowel element [a]. For children only, the duration of the second vowel element [i], rather than the vowel-to-vowel transition, of [ai], is also slightly shortened after an aspirated consonant.

Thirdly, the durations of the syllable-initial consonant the following rime or diphthong [ai] are not negatively correlated, as there is no temporal compensation between the initial consonant and [ai]. This is true for both children and adults.

### 3.4. A summary of the patterns of temporal structures of CV:, CVN, and CVV syllables

Based on a comparison of the temporal patterns pertaining to the three types of Cantonese syllables, CV: (Fig. 1a to Fig. 1d), CVN (Fig. 2a to Fig. 2d), and CVV (Fig. 3a to Fig. 3d), a number of observations are made as follows.

(i) For children and adults, the duration of each of the five initial consonants [p $p^h$ ts $ts^h$ s] is similar when occurring in any one of the three syllable types or associated with any one of the three tones [55, 25, 21]. Thus, the duration of the initial consonant is invariable, and it is longest during [s], to be followed by [$ts^h$], [$p^h$], [ts], and [p] in decreasing order.

(ii) The rime duration of any syllable type is shortened after the aspirated initial [$p^h$ $ts^h$] relative to the unaspirated counterparts [p ts] and when associated with a [21] falling tone. The reduction of rime duration is larger when associated with a [21] tone than after an aspirated initial consonant. For children, the duration reduction of the rimes of CVN and CVV syllables is contributed by the shortening of the durations of the vowel and final nasal or the two vowel elements, but not the vowel-to-nasal or vowel-to-vowel transition, of the rimes. For adults, the reduction of rime duration is chiefly contributed by the first vowel element of the rimes of CVN and CVV syllables. For both children and adults, the durations of the vowel-to-nasal transition of CVN syllables or the vowel-to-vowel transition of CVV syllables remain basically unchanged under any conditions.

(iii) There is no temporal compensation between the initial consonant and the following rime of any one of the three types of test syllables, which indicates that an invariant duration of the syllable of any type is not maintained. It also suggests that the initial consonant and the following rime are not programmed as a single articulatory unit (Lehiste [10])**.**

(iv) The rime duration varies slightly when occurring in the three types of syllables, CV:, CVN, and CVV. For both children and adults, the rime is slightly shorter (26.21 ms and 49.10 ms) in CV: syllables than CVN and CVV syllables.

(v) Of the rime [an] of CVN syllables and the diphthong [ai] of CVV syllables, the duration of the first vowel element [a] is about one half (53% for children and 46% for adults) of the duration of [an] or [ai]. Thus, the rime [an] or diphthong [ai] is divided into two parts generally equal in duration, in that the first part is the first vowel element and the second part includes the vowel-to nasal transition and the final nasal or the vowel-to-vowel transition and the second vowel element.

(vi) The durations of the vowel [a] of the rime [an] and the vowel [a] of the diphthong [ai] are similar. They are about one half (57% for children and 54% for adults) of the duration of the vowel [a:] of CV: syllables. This suggests the occurrence of compensatory shortening of the vowel duration when [a:] occurs in other syllable types, which in turn suggests that the rimes [an] and [ai] are programmed as a single articulatory unit.

## 4. Conclusion

The paper has presented the temporal structures of three types of syllables, CV:, CVN, and CVV, based on the speech from two 9- to 10-year-old Cantonese-speaking children. The following conclusions may be drawn. First, in Cantonese, the durations of the syllable-initial consonants and following rime are not negatively correlated, as there is no temporal compensation involved within the sequence of the initial consonant and rime. Following Lehiste [10], the sequence of the initial consonant and rime within a syllable are not programmed as a single articulatory unit. Second, the duration of the vowel [a] of [an] and [ai] is about one half of the duration of the vowel [a:] of CV: syllables, suggesting the occurrence of compensatory shortening of the vowel duration when [a:] occurs in other syllable types. This in turn suggests that the rimes [an] and [ai] are programmed as a single articulatory unit. Third, the temporal data in this study support Lehiste's [10] contention that there is a closer connection between a vowel and a following consonant than between an initial consonant and a following vowel. Lastly, the general patterns of temporal organization of syllable production for children are similar to those for adults, indicating that children of middle childhood at 9 to 10 years of age have acquired the adult-like temporal patterns. Similar developmental temporal data have also been reported for English-speaking children of middle childhood ([1, 2, 3, 4, 5, 6, 7, 9, 12, 13]).

## 5. Acknowledgements

Support by a SRG grant (#7002390) from the City University of Hong Kong for this research is gratefully acknowledged.

## 6. References

[1] DiSimoni, F.G., "Effect of vowel environment on the duration of consonants in the speech of three-, six-, and nine-year-old children", JASA, 55(2):360-361, 1974.

[2] DiSimoni, F.G., "Influence of consonant environment on duration of vowels in the speech of three-, six-, and nine-year-old children", JASA, 55(2):362-363, 1974.

[3] DiSimoni, F.G., "Influence of utterance length upon bilabial closure duration for /p/ in three-, six-, and nine-year-old children", JADA, 55(6):1353-1354, 1974.

[4] DiSimoni, F.G., "Some preliminary observations on temporal compensation in the speech of children", JASA, 56(2):697-699, 1974.

[5] Eguchi, S. and Hirsh, I.J., "Development of speech sounds in children", Acta Oto-Laryngologica, Supple. 257:5-51, 1969.

[6] Hawkins, S., "Temporal coordination of consonants in the speech of children: preliminary data", JP, 1:181-217, 1973.

[7] Hawkins, S., "Temporal co-ordination of consonants in the speech of children: further data", JP, 7:235-267, 1979.

[8] Kozhevnikov, V.A. and Chistovich, L.A., Speech: Articulation and Perception. Joint Publications Research Service, No.30, 543. Washington University: Moscow-Leningrad, 1965.

[9] Lee, S., Potamianos, A., and Narayanan, S., "Acoustics of children's speech: developmental changes of temporal and spectral parameters", JASA, 105(3):1455-1468, 1999.

[10] Lehiste, I., "Temporal organization of spoken language", in L.L. Hammerich, et al. [Ed.], Form & Substance: Phonetic and Linguistic Papers Presented to Eli Fischer-Jørgensen, 159-169. Copenhagen: Akademisk Forlag, 1971.

[11] Port, R.F., Al-Ani, S., and Maeda, S., "Temporal compensation and universal phonetics", Phonetica, 37:235-252, 1980.

[12] Singh, L., Shantisudha, P., and Singh, N.C., "Developmental patterns of speech production in children", Applied Acoustics, 68:260-269, 2007.

[13] Smith, B.L. and Kenney, M.K., "An assessment of several acoustic parameters in children's speech production development: longitudinal data", JP, 26:95-108, 1998.