

# The Prosody of Bisyllabic and Polysyllabic Words in Hong Kong Cantonese

Wai-Sum Lee<sup>+</sup>, Fangxin Chen<sup>#</sup>, K.K. Luke<sup>+</sup> & Liqin Shen<sup>#</sup>

<sup>+</sup>Department of Linguistics, University of Hong Kong, Hong Kong

<sup>#</sup>IBM China Research Lab, China

<sup>+</sup>wsleeba@hkucc.hku.hk; <sup>#</sup>chenfx@cn.ibm.com; <sup>+</sup>kkluke@hkusua.hku.hk; <sup>#</sup>shenlq@cn.ibm.com

## Abstract

This is a phonetic study of the prosody of bisyllabic and polysyllabic words in Hong Kong Cantonese. The values of the three prosodic properties, duration,  $F_0$ , and intensity, of the words are determined by the type of tones on the component syllables of words as well as the position of the syllables in the words. The  $F_0$  characteristic of the citation tones is maintained, where they are realized on bisyllabic or polysyllabic words. With the same tone on the component syllables of words, the duration of the rime is noticeably longer in the final syllable than in the non-final. However, the  $F_0$  and intensity are lower for the final syllable of words. The duration of the syllable-initial consonant is longer in the initial syllable, if the consonant is an aspirate or fricative. However, the duration of the syllable-initial sonorant is shorter in the initial syllable. The duration variation of the unaspirated initial consonants of the component syllables of words is not significant.

## 1. Introduction

Little work has been done on the prosody of larger linguistic units, such as words or phrases in Cantonese. This study is a phonetic investigation of duration,  $F_0$ , and intensity of bisyllabic, trisyllabic, and quadrisyllabic words in Hong Kong Cantonese (HKC). The results of this study are necessary information for synthesizing speech units longer than a monosyllable with high naturalness and for understanding the rhythmic structure of longer units in a tone language.

## 2. Method

A total of 300 test words, including 100 bisyllables, 100 trisyllables, and 100 quadrisyllables, in HKC were analyzed for their prosodic properties, i.e., duration,  $F_0$ , and intensity. 10 test words of each length are listed in Table 1 as samples. The test words, all frequently used common words in Hong Kong, were selected as representatives of 300 possible combinations of the component syllables in terms of the type of the citation tone on the syllables. The citation tones in HKC are [55, 33, 22, 21, 25, 23] (the long series) and [5, 3, 2] (the short series). For the short tones, usually, [5] is associated with a short vowel and [3] with a long vowel, but [2] with both. Thus, in this study, [2a] is used to represent the tone [2] with a longer vowel and [2b] the tone with a shorter vowel.

Two native speakers of Cantonese, both university students, one male and one female, were asked to utter all the 300 test words at a normal rate of speech. The recordings were performed in a sound-proof booth. The speech materials were analyzed, using the CSL (Computerized Speech Lab) 4100 speech analysis software by Kay Elemetrics of USA. The durations of the syllable-initial consonant and rime of each component syllable in the test words were measured directly from the speech waveforms. The synchronized wideband spectrograms of the speech signals were also referred to for verification where needed. The  $F_0$  contours and intensity curves of the test words were obtained, using the pitch synchronous method provided by the software.

Table 1: 30 of the 300 test words investigated in this study.

Bisyllables	Trisyllables	Quadrisyllables
[ts <sup>h</sup> a siu <sup>55-55</sup> ] 'barbecue pork'	[t <sup>h</sup> in fa pan <sup>55-55-25</sup> ] 'ceiling'	[ka t <sup>h</sup> ij tsy fu <sup>55-21-25-23</sup> ] 'housewife'
[sən fūŋ <sup>33-55</sup> ] 'envelope'	[fai ts <sup>h</sup> an tim <sup>33-55-33</sup> ] 'fast food shop'	[sei t <sup>h</sup> uŋ pat tat <sup>33-55-3-2a</sup> ] 'accessible'
[mou ts <sup>h</sup> uŋ <sup>22-55</sup> ] 'to impersonate'	[pan kŋ sət <sup>22-55-5</sup> ] 'office'	[pan kŋ si kan <sup>22-55-21-33</sup> ] 'office hours'
[p <sup>h</sup> un tsoi <sup>21-55</sup> ] 'pot plant'	[t <sup>h</sup> ou sy kun <sup>21-55-25</sup> ] 'library'	[sən kŋ kən tsəŋ <sup>21-55-25-55</sup> ] 'nervous'
[ts <sup>h</sup> øy siu <sup>25-55</sup> ] 'to cancel'	[mou ts <sup>h</sup> en tsit <sup>23-55-3</sup> ] 'Mother's Day'	[kən tsəŋ k <sup>w</sup> an t <sup>h</sup> eu <sup>25-55-55-21</sup> ] 'important moment'
[laŋ ts <sup>h</sup> ŋ <sup>23-55</sup> ] 'lonely'	[fɔ ts <sup>h</sup> e tsam <sup>25-55-22</sup> ] 'railway station'	[mei tsəŋ pət tsək <sup>23-55-3-5</sup> ] 'imperfect'
[kuk fa <sup>5-55</sup> ] 'chrysanthemum'	[kuk fa ts <sup>h</sup> a <sup>5-55-21</sup> ] 'chrysanthemum tea'	[pət ts <sup>h</sup> ət sɔ liu <sup>5-5-25-22</sup> ] 'expected'
[fat kun <sup>3-55</sup> ] 'judge'	[t <sup>h</sup> it si mŋ <sup>3-55-23</sup> ] 'steel net'	[kit fən tsŋ sy <sup>3-55-33-55</sup> ] 'marriage certificate'
[lɔk kun <sup>2a-55</sup> ] 'optimistic'	[sek kou tsəŋ <sup>2a-55-22</sup> ] 'plaster cast'	[tsap tsəŋ tsŋ sən <sup>2a-55-55-21</sup> ] 'to concentrate'
[fuk sou <sup>2b-55</sup> ] 'to revive'	[tsɪk sŋ kei <sup>2b-55-55</sup> ] 'helicopter'	[tət k <sup>h</sup> øy tsŋ fu <sup>2b-55-33-25</sup> ] 'special region government'

## 3. Results

### 3.1. Duration

#### 3.1.1. Duration of the rime

For the test words of any length, the durations of the rimes of the component syllables are related to the type of the citation tone on the syllables as well as the position of the syllables in the words. With regard to the component syllables occurring in the same position of words, the duration of the rime of the syllable is consistently longest when the syllables are of tone type T1 ([55, 33, 22, 25, 23], or [21]), longer with tone type T2 ([3] or [2a]), and shorter with tone type T3 ([5] or [2b]) for the two speakers (Tables 2a-b). For the male speaker, the durations of the rimes of the component syllables that occur in the same position of words are similar, when the syllables are all of the same type of tone, i.e., T1, T2, or T3 (Table 2a). This is also true for the female speaker, except when the final component syllable is of tone type T1. In this case, the duration of the rime of the final syllable is shorter with [21] than with [55, 33, 22, 25, 23] (Table 2b). In this study, T1a represents the tones [55, 33, 22, 25, 23] and T1b the tone [21] for the female speaker.

Tables 3a-b show the ratio between the duration of the rime of T1, T1a, T1b, or T2 syllables and the duration of the rime of T3 syllables for the two speakers. The ratio is calculated by using the duration of the rime of T3 syllables as the base. As shown in the two tables, for both speakers, the duration of the rime of T1, T1a, or T1b syllables is generally over 2 times longer than that of the rime of T3 syllables, and the duration of the rime of T2 syllables is about 1.5 times longer than that of the rime of T3 syllables, keeping position in the word constant. The difference in duration between the rimes of the syllables with different tone types is comparatively larger in the bisyllabic words than in the polysyllabic and larger when the syllables occurring in the final position of words than in the non-final.

Tables 2a-b: The mean durations (in ms) of the rimes of the syllables with the same type of tone for the two speakers. ('—' = non-applicable)

(a) Male Speaker

	1st Syllable			2nd Syllable			3rd Syllable			4th Syllable		
	T1	T2	T3	T1	T2	T3	T1	T2	T3	T1	T2	T3
Bi.	195.37	128.33	70.01	281.35	182.74	100.57	—	—	—	—	—	—
Tri.	158.84	101.38	63.98	169.64	110.24	69.69	274.83	174.94	100.42	—	—	—
Qua.	163.30	106.18	68.60	158.36	114.96	75.67	166.14	111.98	71.17	257.55	159.62	94.76

(b) Female Speaker

	1st Syllable				2nd Syllable				3rd Syllable				4th Syllable			
	T1a	T1b	T2	T3	T1a	T1b	T2	T3	T1a	T1b	T2	T3	T1a	T1b	T2	T3
Bi.	181.14	174.80	107.62	61.77	288.59	205.47	156.96	105.67	—	—	—	—	—	—	—	—
Tri.	112.36	112.57	71.42	49.51	122.66	122.55	80.79	56.30	251.69	183.82	134.39	93.43	—	—	—	—
Qua.	122.67	127.71	77.10	61.13	120.94	120.97	89.98	64.76	147.62	141.66	101.71	66.28	286.28	199.85	154.86	129.62

Tables 3a-b: The ratio between the duration of the rime of the syllables with T1, T1a, T1b, or T2 and the duration of the rime of the syllables with T3 for the two speakers. ('—' = non-applicable)

(a) Male Speaker

	1st Syllable			2nd Syllable			3rd Syllable			4th Syllable		
	T1	T2	T3	T1	T2	T3	T1	T2	T3	T1	T2	T3
Bi.	2.79	1.83	1	2.80	1.82	1	—	—	—	—	—	—
Tri.	2.48	1.58	1	2.43	1.58	1	2.74	1.74	1	—	—	—
Qua.	2.38	1.55	1	2.09	1.52	1	2.33	1.57	1	2.72	1.68	1

(b) Female Speaker

	1st Syllable				2nd Syllable				3rd Syllable				4th Syllable			
	T1a	T1b	T2	T3	T1a	T1b	T2	T3	T1a	T1b	T2	T3	T1a	T1b	T2	T3
Bi.	2.93	2.83	1.74	1	2.73	1.94	1.49	1	—	—	—	—	—	—	—	—
Tri.	2.27	2.27	1.44	1	2.18	2.18	1.44	1	2.69	1.97	1.44	1	—	—	—	—
Qua.	2.01	2.09	1.26	1	1.87	1.87	1.39	1	2.23	2.14	1.53	1	2.21	1.54	1.19	1

A comparison of the component syllables of the words all with the same tone type shows that the duration of the rime is noticeably longer for the final syllable than for the non-final (Tables 2a-b). This is true for the two speakers as well as test words of any length. Tables 4a-b show the ratios between the durations of the rimes of the final and non-final component syllables and between two non-final component syllables, keeping tone type constant, for the two speakers. As shown in

the tables, for both speakers, the duration of the rime of the final syllable is about 1.5-2 times that of the rime of the non-final syllable, regardless of whether the syllables are of T1, T1a, T1b, T2, or T3. The durations of the rimes of the non-final syllables of the same tone type are generally similar and the ratio is near to 1, although the duration of the rime of the initial syllable tends to be slightly shorter than that of the rime of the other non-final syllables in most cases.

Tables 4a-b: The ratios between the durations of the rimes of the final and non-final syllables and between two non-final syllables, where the syllables all with the same type of tone, for the two speakers. ('—' = non-applicable)

(a) Male Speaker

	Final Syllable/Non-Final Syllable								
	Final/1st			Final/2nd			Final/3rd		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
Bi.	1.44	1.42	1.44	—	—	—	—	—	—
Tri.	1.73	1.73	1.57	1.62	1.59	1.44	—	—	—
Qua.	1.58	1.50	1.38	1.63	1.39	1.25	1.55	1.43	1.33

	Non-Final Syllable/Non-Final Syllable								
	1st/2nd			1st/3rd			2nd/3rd		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
Bi.	—	—	—	—	—	—	—	—	—
Tri.	0.94	0.92	0.92	—	—	—	—	—	—
Qua.	1.03	0.92	0.91	0.98	0.95	0.96	0.95	1.03	1.06

(b) Female Speaker

	(b) Female Speaker											
	Final Syllable/Non-Final Syllable											
	Final/1st				Final/2nd				Final/3rd			
	T1a	T1b	T2	T3	T1a	T1b	T2	T3	T1a	T1b	T2	T3
Bi.	1.59	1.18	1.46	1.71	—	—	—	—	—	—	—	—
Tri.	2.24	1.63	1.88	1.89	2.05	1.50	1.66	1.66	—	—	—	—
Qua.	2.33	1.56	2.01	2.12	2.37	1.65	1.72	2.00	1.94	1.41	1.52	1.96
	Non-Final Syllable/Non-Final Syllable											
	1st/2nd				1st/3rd				2nd/3rd			
	T1a	T1b	T2	T3	T1a	T1b	T2	T3	T1a	T1b	T2	T3
Bi.	—	—	—	—	—	—	—	—	—	—	—	—
Tri.	0.92	0.92	0.88	0.88	—	—	—	—	—	—	—	—
Qua.	1.01	1.06	0.86	0.94	0.83	0.90	0.76	0.92	0.82	0.85	0.88	0.98

### 3.1.2. Duration of the syllable-initial consonant

For the syllable-initial consonant, six types that occur in the component syllables of the 300 test words were analyzed. These include (i) the unaspirated stops [p, t, k, k<sup>w</sup>], (ii) aspirated stops [p<sup>h</sup>, t<sup>h</sup>, k<sup>h</sup>], (iii) unaspirated affricate [ts], (iv) aspirated affricate [ts<sup>h</sup>], (v) fricatives [f, s, h], and (vi) sonorants [m, l]. The duration of the stop or affricate refers to the VOT in this study.

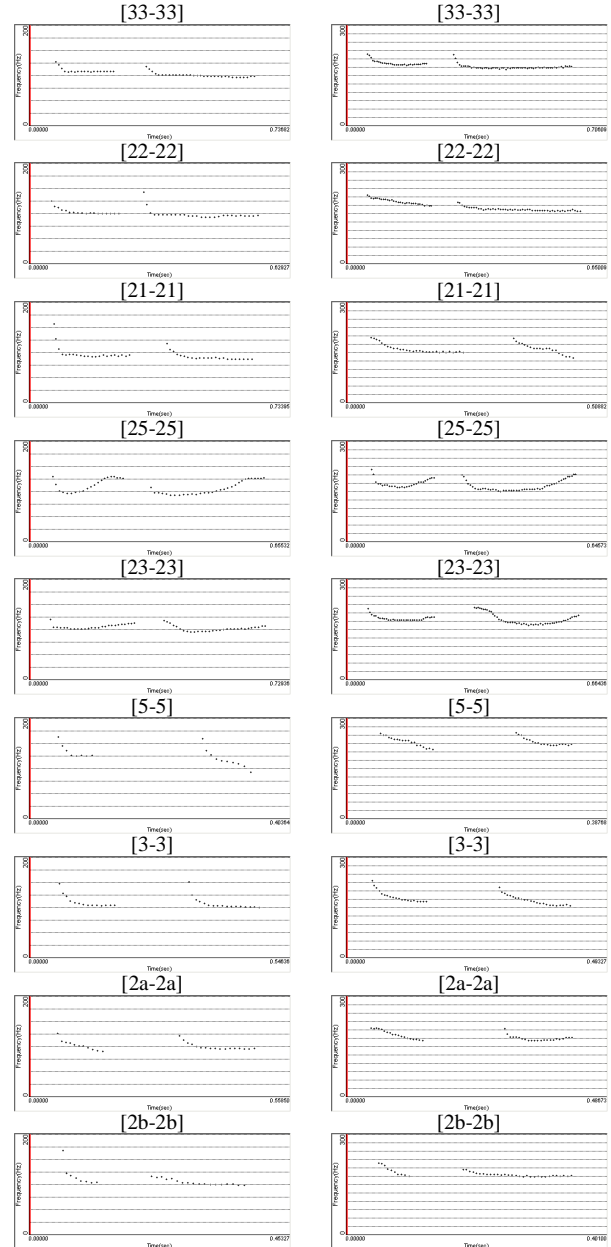
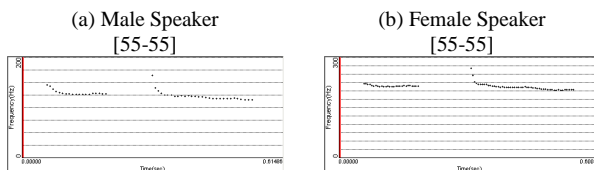
For both speakers, the duration of the syllable-initial unaspirated stop or affricate is similar in any component syllables, regardless of the position of the syllables in a word. However, the durations of the syllable-initial aspirated stops, aspirated affricate, fricatives, and sonorants vary in the component syllables depending on the position of the syllables in the words. For the two speakers, the durations of the syllable-initial aspirated stops, aspirated affricate, and fricatives tend to be shortened, when they occur in the final syllables of the bisyllabic words. In the trisyllabic and quadrisyllabic words, the durations of these syllable-initial consonants tend to be shortened in the non-initial or medial syllables. As for the syllable-initial sonorants, their durations tend to be lengthened in the non-initial syllables of the words of any length. The only exception is that the duration of the syllable-initial [l] is similar in the component syllables of the trisyllabic or quadrisyllabic words for the male speaker.

Besides the factor of the position of the component syllables in words, the duration of the syllable-initial consonant is also related to the type of the preceding or following sound. As shown in Tables 5a-b, for the two speakers, the durations of the affricates [ts] and [ts<sup>h</sup>] and the fricative [s] tend to be longer before a high front vowel [i] or [y] than the other types of vowel. For the fricatives [f] and [s], they tend to be longer after a stop than the other types of sound. Due to limited space, the mean duration of each type of the syllable-initial consonants analyzed in this study for the two speakers are not presented here.

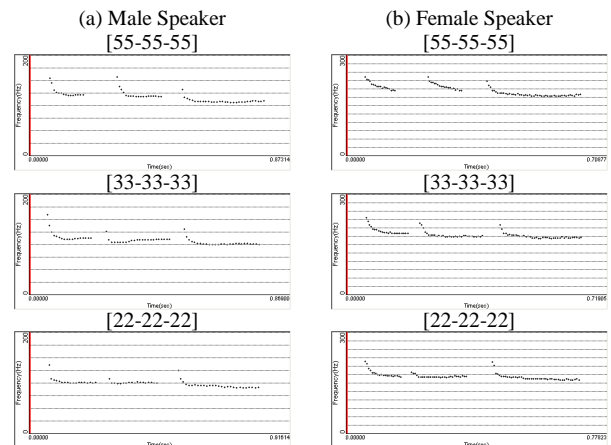
### 3.2. F<sub>0</sub>

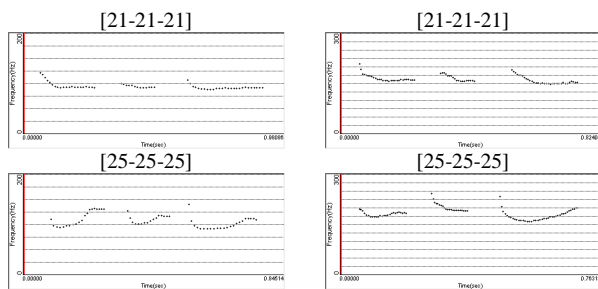
The F<sub>0</sub> characteristic of the component syllables in words of any length is related to the tone type. The F<sub>0</sub> values and contours of the different citation tones on the syllables of a word are basically maintained. The exception is that the [21] tone may be produced as low level, especially for the male speaker. The level tones, either the long series [55, 33, 22] or the short series [5, 3, 2a, 2b], may become slightly falling for the two speakers. And, for the female speaker, both [25] and [23] may become level, when they occur on non-final syllables of some trisyllabic and quadrisyllabic words.

In the case when the tones on the syllables of a word are the same, the F<sub>0</sub> is usually slightly lower for the final syllable. As shown in Figure 1a, for the male speaker, when the two syllables of a bisyllabic word of the same citation tone, the F<sub>0</sub> value is usually lower for the second syllable. This is also true for the female speaker, as shown in Figure 1b, except for the rising tone [25] or [23]. In this case, the value of the beginning portion of the F<sub>0</sub> contour is lower for the second syllable, but the value of the upward deflection of the F<sub>0</sub> contour is higher for the second syllable. The F<sub>0</sub> pattern for the bisyllabic words is also true for the trisyllabic (Figures 2a-b) and quadrisyllabic words (Figures 3a-b). There are only five trisyllabic and one quadrisyllabic words analyzed in this study, with all the component syllables being of the same citation tone.

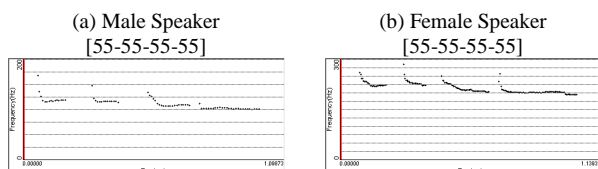


Figures 1a-b: F<sub>0</sub> contours of 10 bisyllabic words with all the component syllables being of the same citation tone for the two speakers.





Figures 2a-b: F<sub>0</sub> contours of 5 trisyllabic words with all the component syllables being of the same citation tone for the two speakers.



Figures 3a-b: F<sub>0</sub> contour of 1 quadrisyllabic word with all the component syllables being of the same citation tone for the two speakers.

### 3.3. Intensity

The variation in the intensity of the component syllables in the bisyllabic and polysyllabic words also depends on the tone type and the position of the syllables in the words. The intensity level of the syllables in words is generally positively correlated with the F<sub>0</sub> level of the syllables. That is, the intensity level is higher for syllables in a high tone or having a high F<sub>0</sub> value. For words consisting of syllables of the same tone, the intensity level tends to be lower for the final syllable than for the non-final. The intensity of the syllables of words is also related to the type of the sound segments contained in the syllables. For syllables with a nasal ending, the intensity is markedly lower toward the end of the syllables. Also, the intensity is relatively higher for the syllables containing a low vowel than the syllables containing a high vowel. Due to limited space, the intensity curves of the bisyllabic and polysyllabic words for the two speakers are not presented here.

## 4. Discussion

In HKC, the three prosodic properties of the component syllables in the bisyllabic and polysyllabic words are variables. The determining factors for the variation are basically (i) the types of the citation tones on the syllables and (ii) the position of the syllables in the words. The data indicate that the F<sub>0</sub> characteristic of the citation tones is maintained, where they are realized on bisyllabic or polysyllabic words. The duration of the rime is noticeably shorter in the non-final syllables than in the final. Conversely, the values of the F<sub>0</sub> and intensity are higher for the non-final syllables than for the final, although the difference is not great. A similar prosodic pattern has been reported on bisyllabic and trisyllabic words in Beijing Mandarin ([4], [5], [7]). In these words, the component syllables are associated with one of the four citation tones, but not the neutral tone, in Beijing Mandarin. For the bisyllabic and trisyllabic words with the neutral tone on one of the component syllables, the syllable associated with the neutral tone is noticeably shorter in duration, even though it occurs on the final syllable of the words ([2], [3]). The F<sub>0</sub> and intensity of the syllables of words with the neutral tone basically depend on the phonetic form of the neutral tone realized in each case ([2], [3]). The information indicates that the prosodic properties of the bisyllabic and polysyllabic words in HKC and Beijing Mandarin are related to the type of tones on the component syllables of words. Besides the factor of the tone, the position of the syllable in a word also affects the prosodic properties.

The final syllable's lower F<sub>0</sub> and intensity may be due to a physiologically-based universal speech pattern, namely the gradual reduction of subglottal pressure in the course of uttering a word or phrase group ([6]). As for the longer duration of the rime of the final syllable in a word, it is worth considering whether this is due to the effect of 'prepausal lengthening'. However, the durations of the syllable-initial consonants do not suggest that there is any significant 'prepausal lengthening'. In bisyllabic words, the durations of the syllable-initial unaspirates are generally similar in any component syllables of a word. And, for the aspirates or fricatives, the duration is longer for the consonant in the initial syllable than in the final. Furthermore, it is reported that for bisyllabic ([1]) and trisyllabic ([8]) words in Mongolian, the duration is markedly longer for the rime in the initial syllable than in the final. As Mongolian is a non-tonal language, the temporal characteristic of the component syllables of words in Mongolian is assumed to be related to the position of the syllables in the words. Since the temporal pattern of the bisyllabic and trisyllabic words in Mongolian is different from that of HKC or Beijing Mandarin, the longer duration for the rime in the final syllable of the words in HKC and Beijing Mandarin should be considered as a temporal characteristic of the language, rather than the result of 'prepausal lengthening'.

As for the syllable-initial consonant, it is interesting to find that in HKC the short consonants, such as the unaspirates vary less in duration; the long consonants, such as the aspirates and fricatives, tend to be shortened in the syllables that occur in the non-initial, especially the medial position of a word; and the sonorants, such as the nasal and lateral approximant, tend to be shortened in the initial syllable of a word. A similar tendency has been reported in relation to the studies which take the effect of speech rate into account with regard to the durations of the consonantal segments ([6]).

## 5. Conclusion

To conclude, in HKC, the prosodic characteristics of the bisyllabic, trisyllabic, and quadrisyllabic words are strongly affected by the characteristics of the tones on the component syllables of the words. The position of the component syllables in the words is also a factor in the variation.

## 6. References

- [1] Huhe H.; Zheng, Y.L.; Chen, J.Y., 2000. Research on stress in bisyllabic words of Mongolian. In the *Proceedings of ICSLP 2000*, 3, 171-174.
- [2] Lee, W.S., 2000. A phonetic and phonological analysis of the neutral tone in Beijing Mandarin. Paper presented at the 9th International Conference of Chinese Linguistics, National University of Singapore, 26-28 June, 2000.
- [3] Lin, M.C.; Yan J.Z. (林茂燦, 顏景助), 1980. 北京話輕聲的聲學性質. *Fangyan* (3), 166-178.
- [4] Lin, M.C.; Yan, J.Z.; Sun, G.H. (林茂燦, 顏景助, 孫國華), 1984. 北京話兩字組正常重音的初步實驗. *Fangyan* (1), 57-73.
- [5] Yan, J.Z.; Lin, M.C. (顏景助, 林茂燦), 1988. 北京話三字組重音的聲學表現. *Fangyan* (3), 227-237.
- [6] Werner, S.; Keller, E., 1994. Prosodic aspects of speech. In *Fundamentals of Speech Synthesis and Speech Recognition*, E. Keller (ed.). Chichester: Wiley, 23-40.
- [7] Wu, Z.J. (吳宗濟), 1984. 普通話三字組變調規律. *The Journal of Linguistic Society of China* (2), 70-92.
- [8] Zheng, Y.L.; Bao, H.Q. (鄭玉玲, 鮑懷翹), 2001. 蒙古語三音節詞韻律模式. In the *Proceedings of the 5th National Conference on Modern Phonetics*, 84-92.

[Acknowledgement: The authors would like to acknowledge Tam Wai-Lok's help in the processing of part of the phonetic data used in the study.]