Syllable isochrony and the prosodic features of stop syllables in Cantonese

Ling Zhang
The Education University of Hong Kong

Cantonese is a syllable-timed language: that is, the syllable is the isochronous unit of speech. However, in Cantonese, there is a type of closed syllable with the stop codas [-p], [-t], or [-k] (i.e. syllables with the so called "entering-tones") which sound much shorter than other syllables. On the surface, the shorter duration of stop syllables and the general prosodic feature of syllable-isochrony seem to conflict. This study conducted acoustic investigations of stop syllables in Cantonese in different contexts (i.e. in isolated form, in disyllabic words, and in disyllabic words located at the beginning, middle, and final positions of sentences). The results showed that stop syllables alone are shorter than non-stop syllables in various contexts. However, in disyllabic words or in sentences, there is a supplementary lengthening effect immediately after the stop syllables: there is more acoustic blank, and in some circumstances the initial of the following syllable is lengthened. Therefore, we propose that the phonetic realization of syllable isochrony is beyond the syllable itself in Cantonese. The results and discussions of this study may also shed light on the problem of the disappearance of "entering tones" from various Chinese dialects.

Keywords: syllable isochrony, stop syllable, entering tone, prosodic feature, Cantonese

1. Introduction

The syllable is an important concept in Cantonese, not only as a basic structure for phonological analysis but also as a basic isochronous unit in speech prosody, since it is a syllable-timed language (Mok 2008). According to the syllable isochrony hypothesis, each syllable in Cantonese should take up a similar amount of time. However, in Cantonese, there is a syllable type with entering tones, which end with the stop coda [-p], [-t], or [-k] and sound extraordinarily short. It seems that with different coda types (stop coda versus non-stop coda), syllables should have

https://doi.org/10.1075/lali.00098.zha | Published online: 15 December 2021
Language and Linguistics 23:1 (2022), pp. 20–46. ISSN 1606-822X | E-ISSN 2309-5067

distinctive lengths by nature, which seems contradictory to the syllable isochrony hypothesis in regard to Cantonese. This study conducted a series of acoustic experiments to investigate the prosodic features of stop syllables in Cantonese and tried to provide a unified framework to accommodate the general prosodic feature of syllable isochrony and the extraordinary shortness of stop syllables.

1.1 Rhythmic typology and syllable isochrony in Cantonese

Rhythm plays an important role in human behavior and activity. It is the structure of intervals in a succession of events (Allen 1975). The speech of human languages is generally characterized by a rhythmical component (Mariano 2011), although various typologies exist among different languages.

Early studies on rhythm typology proposed that all languages exhibit isochronous units of speech (i.e. the prominent and less prominent units in speech recur at equal intervals of time). Pike (1945) observed that stress seems to occur at regular temporal intervals in English, while syllables seem to have similar durations in Spanish. He was the first to introduce the two categories of stress-timed and syllable-timed languages. Abercrombie (1967: 97) further argued that every language in the world is spoken with one of these kinds of rhythm. Typical stress-timed languages include German, Dutch, and English, while typical syllable-timed languages include French, Italian, and Spanish.

However, early rhythm typology studies were mainly impressionistic. It is hard to find empirical support for isochrony in speech. Previous empirical studies found uneven length of inter-stress intervals, syllable, and mora durations (e.g. Allen 1975; Roach 1982; Dauer 1983, among many others). Some researchers suggested that speech rhythm is merely perceptual (e.g. Beckman 1992; Laver 1994). Roach (1982) and Dauer (1983) proposed that compared with syllable-timed languages, stress-timed languages have more variations and rules of word stress, more complex syllable structure and consonant clusters, and more frequent vowel reduction. On the basis of these features, researchers developed rhythmic measures based on consonantal and vocalic durations in the acoustic, including the interval measures of ΔC, ΔV, %V (Ramus et al. 1999), VarcoC, VarcoV (Dellwo 2006), and the pairwise variability index (e.g. rPVI_C and nPVI_V by Grabe & Low (2002)).

The rhythmic typological category of Chinese (including Beijing Mandarin and Cantonese) is not classified in early studies. Grabe & Low (2002) found that Mandarin Chinese had the lowest vocalic nPVI of all the languages investigated in their study, which suggested that Mandarin Chinese has the strongest syllable-timed feature. Several later studies (e.g. Jeng 2008, 2011; Lin & Wang 2007) also confirmed the syllable isochrony prosodic feature of Mandarin Chinese. Cao (2000; 2003) and Jeng (2005) investigated the effect of different contexts, including the positions of words within sentences and speech rate, on the syllable duration patterns of Chinese. These studies suggested that context plays an important role in syllable duration. In this study, the experimental design took the factor of context into consideration and investigated various contexts.

As far as Cantonese is concerned, Bauer (1995: 255–256) observed that the rhythm of Cantonese speech sounds like a series of recurring beats which correspond to peaks of sonority (i.e. syllables). He further claimed that although different types of Cantonese syllables may vary in their absolute durations, the syllables of an utterance recur at about equal intervals of time. Bauer's (1995) study was mainly based on impressionistic observation, and later empirical studies verified his argument. Peng (2006) conducted corpus-based study, and Mok (2008) carried out controlled acoustic experiments. Both of these studies compared the rhythmic patterns of Beijing Mandarin and Cantonese. Both studies supported that Cantonese is a syllable-timed language, and the syllable isochrony feature is even stronger in Cantonese than in Beijing Mandarin. It can be inferred that the prosodic feature of syllable isochrony is strong and robust in Cantonese. There is no doubt that Cantonese is a syllable-timed language.

1.2 Syllable structure and stop syllables in Cantonese

As illustrated in Light (1977: 75–76), the Chinese syllable is divided into three major constituents: initial, final, and tone. The final is further divided into the medial and the rhyme.[1] The rhyme is composed of a vowel nucleus and a syllable coda consonant or an offglide. Among the three major constituents of a syllable, both the initial and the final (as well as the sub-constituents of the nucleus and coda) are segmental constituents, while tone is a suprasegmental constituent which cannot be segmented along the timeline of a syllable. Tone is the feature of the whole syllable and is mainly carried by the final. With regard to Cantonese, it is generally accepted that there is no medial in Cantonese phonology (Hashimoto 1972; Cheung 1986; Bauer & Benedict 1997, among many others). The diagram in Figure 1 provides a depiction of the syllable structure in Cantonese. Without the medial, the final is the rhyme in Cantonese. In the present manuscript, the term *rhyme* is used henceforth.

1. In Light (1977), "rhyme" was written as "rime" and "vowel nucleus" was referred to as the "principal vowel".

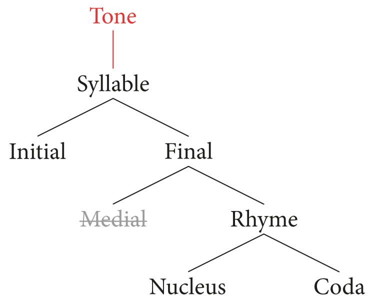


Figure 1. Diagram showing Cantonese syllable structure

Figure 1 depicts the maximum syllable structure of Cantonese. Some constituents can be missing in a legal syllable in Cantonese. The initial can be missing: this is also known as a "zero-initial" syllable. It should be noted that zero-initial and ŋ-initial are free allophones in present-day Cantonese (Cheung 2002). In our experiment, we used an ŋ-initial quantifier [ŋaːn^{23}] (眼), whose initial was pronounced as a ŋ-initial or zero-initial by our participants.

The rhyme in Cantonese can be a monophthong, or a diphthong consisting of a vowel nucleus and a consonant coda or an offglide. Previous studies reported that there are pairs of long and short vowel nuclei in Cantonese (Cheung 1994; Shi & Mai 2003; Cheung & Zhang 2010; Zhang 2010). Through acoustic experiment and perceptual test, Zhang (2010) proved that for the pair [aː] and [ɐ], the durational difference is the major distinction and the vowel height is a secondary distinction. A long vowel nucleus can appear in either monophthong or diphthong. A short vowel nucleus cannot be a monophthong. Kwan-hin Cheung (1994) and Zhang (2010) both suggested that the duration proportion of vowel nucleus and coda is complementary: a long vowel nucleus is followed by a short coda; while a short vowel nucleus is accompanied by a long coda. This complementary durational distribution is consistent with the feature of syllable isochrony in Cantonese: the long and short distinction of vowel nucleus does not make any great difference to the total duration of a rhyme, since there is a compensatory durational adjustment from the following codas.

There is a special kind of rhyme in Cantonese which only has one constituent of a syllabic nasal consonant [m̩] and [ŋ̩]. Kwan-hin Cheung (1986) proposed that [m̩] and [ŋ̩] should be analyzed as a rhyme coda rather than an initial consonant, which can allow the existence of [hm̩] and [hŋ̩] in terms of syllable structure. In our study, we used the numeral "5", which is an example of this rhyme, and can be freely pronounced as [m̩^{23}] or [ŋ̩^{23}] in present-day Cantonese. Sociolinguistic studies suggested that old people tend to use [ŋ̩] while the young generation prefer [m̩] (Bauer 1979; Cheung 2002). In our sample, since the participants were all young people, they all pronounced “5” as [m̩^{23}]. Kwan-hin Cheung’s (1986) view is adopted in the present study, and the duration of the nasal syllabic consonant [m̩] is counted as the duration of the rhyme.

According to coda types, Cantonese rhymes can be divided into three categories, the traditional Chinese linguistic terms for which are as follows: (1) Yinsheng rhymes: with no consonantal codas; (2) Yangsheng rhymes: with nasal codas; (3) Rusheng rhymes (or entering rhymes): with stop codas (i.e. [-p], [-t], or [-k]). A Rusheng rhyme stands out for its abrupt stopping at the end, which sounds extraordinarily short compared with the other two categories of rhymes. In this manuscript, syllables with a Rusheng rhyme are called stop syllables and syllables with a Yinsheng rhyme or a Yangsheng rhyme are called non-stop syllables.

Tone is a required constituent of Cantonese syllables. In traditional Chinese linguistics, the term *tone categories* not only involves the pitch categories that make lexicons distinctive but also takes categories of rhymes into consideration. Stop-syllables and non-stop syllables with the same pitch features are divided into separated tone categories: entering-tones and non-entering tones. Figure 2 shows the diagrams of the nine tones in Cantonese, with the five-point scale method for transcribing tones, which was proposed by Chao (2006[1930]). It can be observed from Figure 2 that T1 and T7, T3 and T8, and T6 and T9 are the same from the perspective of pitch shape and pitch height, being respectively high-level, mid-level, and mid-low-level tones.

According to the Western tradition of transcribing tones, which only considers the factor of pitch, each pair of the same pitch shape and height should be the same tone, and there are six tones in Cantonese. Comparing the six-tone system in the Western tradition and the nine-tone system in the Chinese tradition, the categorization of rhymes and tones are confounded in the Chinese tradition. The stop syllables mentioned here are also referred to as syllables with entering tones in Chinese. In addition to the canonical tones shown in Figure 2, there is a kind of morphological changed tone (變音) in Cantonese which denotes familiarity, intimacy, change of word class, or derived meaning. The changed tone is in the form of a high-rising tone which sounds the same as T2. When this changed tone is carried by a stop syllable, such as *aap2* ‘duck’, *zat2* ‘nephew’, or *zok2* ‘chisel’, it cannot get a position in the nine-tone system in Chinese but can be compatible in the six-tone system of Western linguistic tradition.

Wong & Chan (2018) disagreed that entering tones T7, T8, and T9 are shorter versions of the three level tones T1, T3, and T6. Their acoustic data graphs showed that entering tones have falling contours. However, from their data graphs, it can be observed that the so-called “level tones” also have slightly falling contours. This falling tendency can be explained by the declining intonation effect of speech, as suggested by Zhang (2016; 2017). In addition, entering tones and non-entering tones are in complementary distribution: they are carried by different types of rhymes (i.e. stop rhymes versus non-stop rhymes) and thus never appear in the same phonetic context. According to the Western linguistic tradition, although there is a slight difference, they should be regarded as allophones of the same tone rather than two distinctive tone categories (as proposed by Wong & Chan (2018)). The focus of Wong & Chan (2018) was on the f_0 patterns of entering tones and non-entering tones, while this study focuses on the durational patterns.

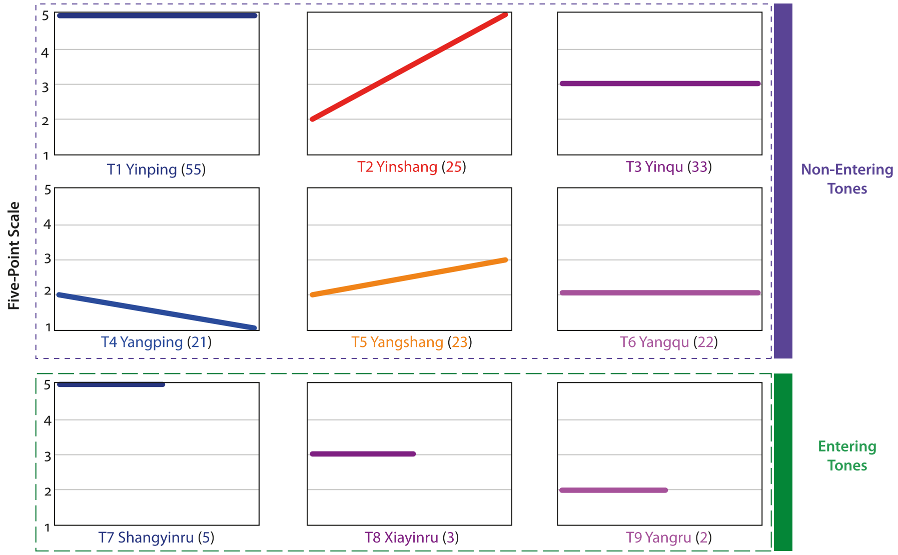


Figure 2. Diagrams showing Cantonese tones, including non-entering tones and entering tones

If there is an intrinsic distinctive difference in syllable length due to the nature of syllable coda types, this contradicts the syllable-isochrony hypothesis. The acoustic experiments conducted in this study were designed to figure out the mechanism for accomodating the two linguistic features of syllable-isochrony and the unbalanced syllable length between stop syllables and non-stop syllables.

2. Research method

2.1 Design of the reading list

The Chinese reading list for the acoustic experiments consisted of three parts. (See Appendix.) The first part consisted of numerals in isolated form, including the numbers from one to ten, plus the numerals for zero, hundred, and thousand, all of which are monosyllabic words in Cantonese. These numerals consist of different types of initials and rhymes, including both stop syllables and non-stop syllables. With the first part of the reading list, we can make a general observation on the duration pattern of isolated syllables in Cantonese.

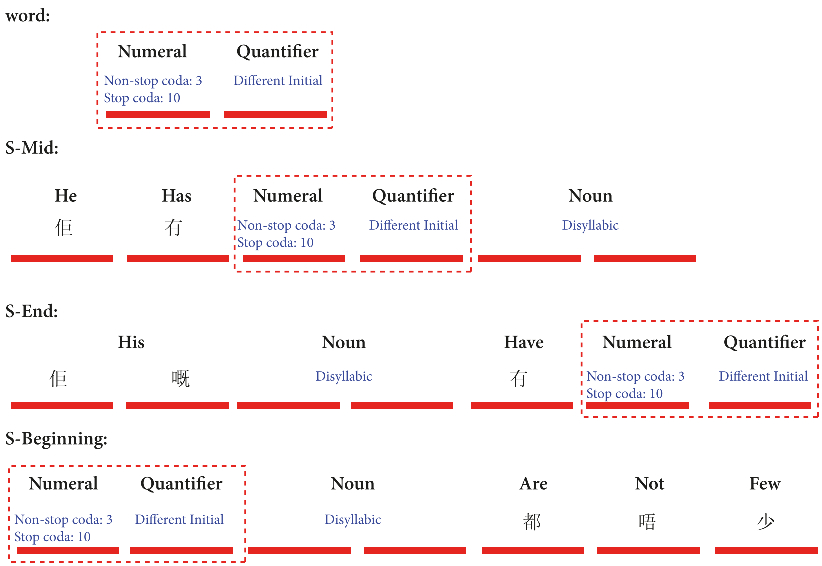


Figure 3. Diagrams showing the design of the second and third parts of the reading list

The second and third parts were designed to investigate the different duration patterns of the numerals "3 (*saam1*)" versus "10 (*sap6*)".[2] *Saam1* is an example of a non-stop syllable, while *sap6* is an example of a stop syllable. Both *saam1* and *sap6* have the same initial *s*- but have different rhymes (-*aam* and -*ap*). The rhyme -*aam* is composed of a long vowel nucleus *aa* and a bilabial nasal coda *m*, while the rhyme *ap* is composed of a short vowel nucleus *a* and a bilabial stop coda *p*.

2. *saam1* and *sap6* are the Cantonese transcription according to the Jyutping Scheme proposed by the Linguistic Society of Hong Kong (1993). This transcription method is alphanumeric, which is more convenient for denotation when plotting data graphs. The corresponding IPA transcription for *saam1* and *sap6* is [saːm^{55}] and [sɐp^{2}]. The Jyutping Scheme was also applied in TextGrid of Praat, as demonstrated in Figure 4.

In terms of the vowel height of the vowel nucleus and the place of articulation for the coda, they are similar, but in terms of the vowel length and the manner of articulation for the coda, which are directly related to duration patterns, they are contrastive.

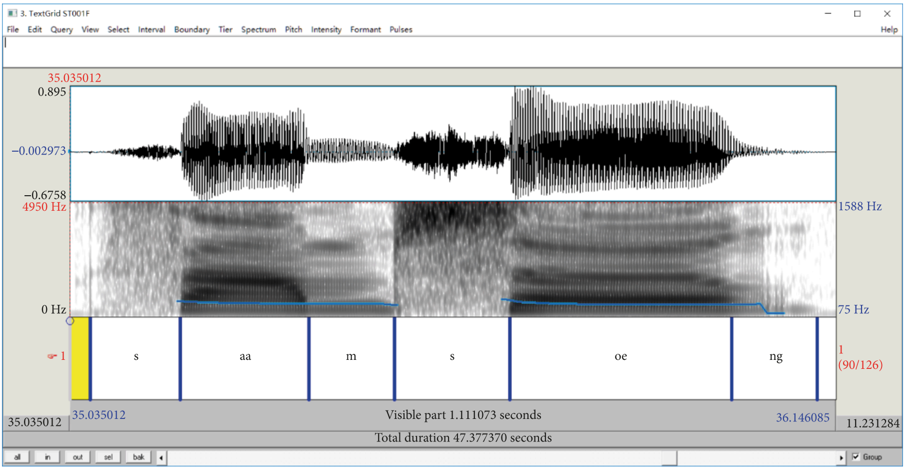
The various monosyllabic quantifiers in Cantonese can cover different possibilities of the composing initials, and thus the combination of a disyllabic word NUMERAL + QUANTIFIER can allow us to study the interaction between the rhyme of a numeral and the various types of following initial of a quantifier. In addition, placing NUMERAL + QUANTIFIER in various positions of a sentence, including at the beginning, end, and middle (abbreviated as S-BEGINNING, S-END, and S-MID in this paper), is a flexible approach. Figure 3 displays the diagrams showing the design of the second and third parts of the reading list. For the contexts of WORD and S-MID, the initials of QUANTIFIER covered the whole set of the 17 initials[3] of Cantonese, when we conducted a thorough investigation to study the interaction between the rhyme types of NUMERAL and the initial types of QUANTIFIER. For the contexts of S-END and S-BEGINNING, we conducted an exploratory investigation, and QUANTIFIER only covered three sets of initials (i.e. [n], [k], and [s]) as representatives of voiced, plosive, and fricative initials.

2.2 Participants, audio-recording settings, and acoustic measurement

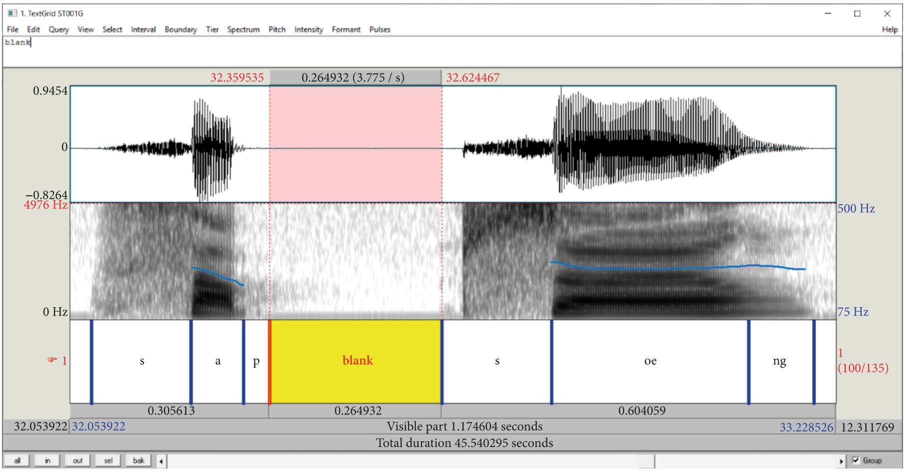
Twenty volunteer undergraduate students (10 male and 10 female) from the Education University of Hong Kong participated in our experiment. They were all native speakers of Hong Kong Cantonese. They were aged between 20 and 24 at the time they joined the experiment. No language, psychiatric, or psychological disorders were reported. The experiment was conducted in a quiet office at the Education University of Hong Kong. The procedures of the experiment were explained to participants first, and written consent forms were signed before the experiment.

The participants were asked to read aloud the single syllables, disyllabic words, and sentences listed in the Appendix. If the participants made a mistake during their reading, they were asked to go ahead and provide a correct reading. They were told that the wrong version would be ignored and only the correct version would be used for later data analysis. Audio recordings were made with a Marantz PMD620 Professional Handheld Digital Audio Recorder and a Sennheiser e845-s microphone. The microphone was set at mono channel, the sampling rate was 44,100 Hz, the resolution was 16-bit, and the saving format was WAV file.

3. [k^{w}] and [k^{wh}] are not included.



a.



b.

Figure 4. Screenshots (*saam1 soeng1* in the upper panel and *sap6 soeng1* in the lower panel) of the TextGrid method in Praat, with segmental boundaries marked by blue lines, Jyutping transcription denoting targeted segments, and "blank" representing acoustic blank

The audio recordings were further processed and divided into several sound files. The acoustic measurement was carried out by the free Praat software (Boersma & Weenink 2018). First, the boundaries of the segments (i.e. the initial, vowel nucleus, and coda of each targeted syllable as well as the acoustic blank between NUMERAL and QUANTIFIER) were marked by the TextGrid function in Praat.

Screenshots of two examples of this procedure are displayed in Figure 4, with the upper panel showing *saam1 soeng1* and the lower panel showing *sap6 soeng1*, both of which are disyllabic words of "NUMERAL (3 or 10) + QUANTIFIER (*soeng1*)". It can be observed that there is an obvious acoustic blank between *sap6* and *soeng*, which is marked as "blank" in the lower panel.

Second, a Praat script written by the author was applied to measure the duration of each segment with TextGrid boundaries. The data were exported to Microsoft Excel (Office 365) and SPSS 25 Network Version (with EdUHK license) for further statistical analyses.

3. Results

3.1 Duration patterns of isolated syllables

Figure 5 shows the duration data from Part I of the reading list (i.e. the numerals in isolated syllables). The stacked columns represent the mean values of the composing initials and rhymes, and the error bars indicate ±1 standard deviation. The *y* axis is the duration data in seconds (abbreviated as "s" henceforth). The *x* axis is the numerals (which are denoted by Arabic numbers). Under the Arabic numbers, there is a line of Jyutping transcription (Linguistic Society of Hong Kong 1993).

The data displayed in Figure 5 indicate that there is great variance among these numerals in terms of duration of rhymes, duration of initials, and total duration of the syllables.

The duration of rhymes data show a clear distinction between stop syllables and non-stop syllables: all rhymes with a *-p, -t*, or *-k* coda (i.e. *jat1, luk6, cat1, baat3, sap6, baak3*) are much shorter than those without a stop coda (i.e. *ji6, saam1, sei3, m5, gau2, ling4, cin1*). On the basis of the average rhyme duration, the numerals can be listed in order from the shortest to the longest: *sap6* (0.151s) < *luk6* (0.158s) < *jat1* (0.164s) < *cat1* (0.167s) < *baak3* (0.212s) < *baat3* (0.237s) < *ling4* (0.290s) < *ji6* (0.349s) < *gau2* (0.366s) < *cin1* (0.380s) < *sei* (0.417s) < *saam1* (0.435s) < *m5* (0.460s). Thus, there is a clear cut-off between stop rhymes and non-stop rhymes. In addition, to avoid the influence from different initials, those numerals with the same initial but with stop or non-stop rhymes can be paired for observation (i.e. the rhymes of *jat1* vs. *ji6; luk6* vs. *ling4; cat1* vs. *cin1; sap6* vs. *saam1*; and *sap6* vs. *sei3*). For each pair, the first syllable is a stop syllable and the second syllable is a non-stop syllable. It can be observed that the rhymes of stop syllables are only about half the length of their non-stop counterparts, or even shorter. An independent-samples *t*-test was carried out to examine whether the difference in duration of rhymes is significant between stop and non-stop syllables. A Levene's test for equality of variances was conducted first, with $F = 142.406$ and $p < 0.001$. Thus, equal variances were not assumed. The *t*-test result is $t = 17.174$, $df = 543.926$, $p < 0.001$; this verified that the rhymes of stop syllables are significantly shorter than the rhymes of non-stop syllables.

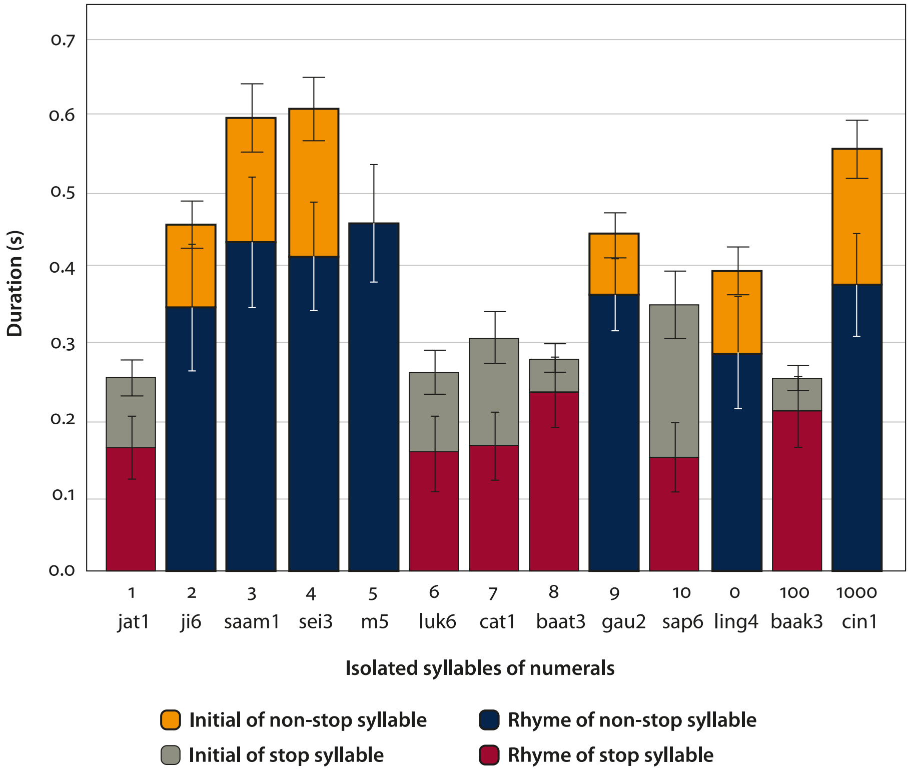


Figure 5. Stacked mean duration columns of the composing initials and rhymes of numerals in isolated syllables, with error bars indicating ±1 standard deviation

The duration of initials also varies. On the basis of the average duration of initials, the numerals can be listed in order from the shortest to the longest: *m5* (0s) < *baak3* (0.043s) < *baat3* (0.045s) < *gau2* (0.079s) < *jat1* (0.092s) < *luk6* (0.106s) < *ling4* (0.108s) < *ji6* (0.109s) < *cat1* (0.141s) < *saam1* (0.165s) < *cin1* (0.179s) < *sei3* (0.195s) < *sap6* (0.202s). The duration patterns of initials are closely related to their articulation manners. The unaspirated stops of *b-* and *g-* are the shortest; the approximants of *j-* and *l-* are a little longer; and the affricate *c-* and the fricative *s-* are the longest. Regarding the difference in the duration of initials between stop syllables and non-stop syllables, there is no fixed pattern: *jat1* < *ji6*, *luk6* < *ling4*, *cat1* < *cin1*; while *sap6* > *saam1*, and *sap6* > *sei3*.

The total duration of a syllable, that is, the sum of the initial and the rhyme, is decided by both components. On the basis of the average total duration values, the numerals can be listed in order from the shortest to the longest: *baak3* (0.255s) < *jat1* (0.256s) < *luk6* (0.264s) < *baat3* (0.282s) < *cat1* (0.309s) < *sap6* (0.353s) < *ling4* (0.398s) < *gau2* (0.445s) < *ji6* (0.459s) < *m5* (0.460s) < *cin1* (0.559s) < *saam1* (0.600s) < *sei* (0.612s). Although the order based on total syllable duration is not the same as that based on total rhyme duration, there is a clear cut-off between the stop and non-stop syllables: all the stop syllables are shorter than the non-stop syllables. An independent-samples *t*-test was carried out to examine whether total syllable duration is significantly different between stop and non-stop syllables. A Levene's test for equality of variances was conducted first, with $F = 62.693$ and $p < 0.001$. Thus, equal variances were not assumed. The *t*-test result was $t = 32.948$, $df = 706.386$, $p < 0.001$. This result confirmed that the total duration of stop syllables is significantly shorter than that of non-stop syllables.

Therefore, the duration data for isolated numeric syllables presented here indicate that syllables in isolated form are not isochronous phonetically. Compared with non-stop syllables, stop syllables have significantly shorter rhymes and total syllable duration. The duration of initials also varies, depending on the articulation manner, but this variation does not exhibit a distinction between stop and non-stop syllables.

3.2 Duration patterns of the syllables *saam1* and *sap6* in different contexts

As observed in § 3.1, in terms of total syllable duration, there is a significant difference between stop syllables and non-stop syllables in the context of isolated syllables. This section reports the findings of a further analysis conducted to investigate the difference in duration between stop and non-stop syllables (*sap6* and *saam1* as representatives) in other contexts, such as in a disyllabic word alone or at the beginning, end, and middle of sentences (abbreviated as S-BEGINNING, S-END, and S-MID, respectively). The design of the experiments was illustrated in detail in § 2.1. The duration data are displayed as a bar-chart in Figure 6, and the corresponding statistical analysis results are listed in Table 1. The "difference" column in Table 1 represents the duration of saam1 minus the duration of *sap6*, and the "ratio" column represents the saam1 to *sap6* duration ratio.

It can be observed from Figure 6 that the stop syllable *sap6* is significantly shorter than the non-stop syllable saam1 in all contexts. The paired comparisons between *saam1* and *sap6* in Table 1 show that the mean duration difference ranges from 0.11s (at S-BEGINNING) to 0.247s (in ISOLATION). The duration ratio of *saam1* to *sap6* ranges from 1.502 (S-BEGINNING) to 1.745 (S-END): that is, the non-stop syllables are at least 50% longer than the stop syllables. The paired-samples *t*-tests results listed in Table 1 further confirm that in all the contexts (ISOLATION, WORD, S-MID, S-END, S-BEGINNING), the syllable duration of *sap6* is significantly shorter than saam1, with all *p*-values smaller than 0.001. Therefore, the duration data in this experiment verify that stop syllables are much shorter than non-stop syllables in various contexts. This result again contradicts the hypothesis of syllable-isochrony in Cantonese.

The data shown in Figure 6 and Table 1 also display the duration features of different contexts. For the both types of syllables (either stop syllable *sap6* or non-stop syllable *saam1*), they reach the longest duration in the context of ISOLATION, and the second longest in the context of WORD. In the contexts of sentences (S-MID, S-END, and S-BEGINNING), the duration data are shorter than in the contexts of ISOLATION and WORD. Thus, the speech rates differ in different contexts. Syllable isochrony should be discussed with the condition of a certain context.

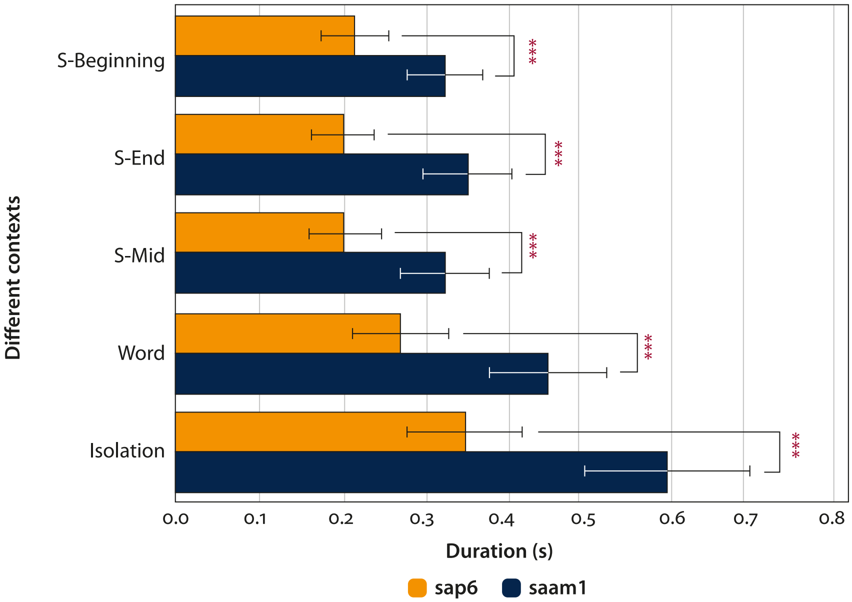


Figure 6. Mean duration bars of *sap6* and *saam1* in different contexts, with error bars indicating ±1 standard deviation, and "***" indicating $p < 0.001$ in the paired-samples *t*-test

Table 1. Results of paired comparison and paired-samples *t*-tests for syllable duration data between *saam1* and *sap6*

Context	Mean of *saam*1 (s)	Mean of *sap*6 (s)	Difference (s)	Ratio	*T*	*df*	*p*
ISOLATION	0.6	0.353	0.247	1.700	19.745	59	<0.001
WORD	0.454	0.274	0.18	1.657	57.997	339	<0.001
S-MID	0.329	0.206	0.123	1.600	55.705	339	<0.001
S-END	0.356	0.204	0.152	1.745	27.319	59	<0.001
S-BEGINNING	0.329	0.219	0.11	1.502	21.381	59	<0.001

3.3 Duration patterns of the disyllabic words "*saam1/sap6* + QUANTIFIER" in different contexts

In this section, the analysis will go beyond the domain of the syllable (as in § 3.2) and expand into the domain of the disyllabic word. The disyllabic words are in the format of "*saam1/sap6* + QUANTIFIER". The testing contexts include the disyllabic words alone and the disyllabic words located at S-MID, S-END, and S-BEGINNING. The durations of the whole disyllabic words, which embraced the acoustic blank (if there had been one) between *saam1/sap6* and the QUANTIFIER, were measured. The results are displayed in Figure 7, and the corresponding statistical analysis results are listed in Table 2.

Different from the patterns displayed in Figure 6, which show significant differences between the paired duration bars, there is much less difference between the paired duration bars in Figure 7; in particular, each pair in sentences (S-MID, S-END, and S-BEGINNING) has almost the same length. Although the pair in WORD has the greatest and observable duration difference, the mean difference is only 0.045s. In terms of duration ratio (i.e. the duration of "*saam1* + QUANTIFIER" divided by the duration of "*sap6* + QUANTIFIER"), the data are very close to 1 in all the contexts, ranging from 0.999 to 1.056, indicating that they have very similar durations.

Paired-samples *t*-tests were also conducted. The results listed in Table 2 verify that in the contexts of S-MID and S-END, there are no significant differences in duration, with *p* values much higher than 0.05. The duration difference is still statistically significant in the context of WORD ($p<0.001$) and marginally significant in the context of S-BEGINNING ($p=0.034$). However, comparison of the duration bars between Figures 6 and 7 indicates that the duration difference within pairs is obviously reduced in Figure 7 compared with Figure 6. The WORD duration comparisons within pairs in Figure 7 indicate that the disyllabic words are approaching isochrony within a certain context.

For context comparisons, it can be observed from Figure 7 that the disyllabic words alone (WORD) are the longest; and the disyllabic words located in sentences (S-MID, S-END, and S-BEGINNING) are much shorter. This is consistent with the observation from Figure 6. Among the different locations in sentences, the disyllabic words located at S-END are visibly longer than those located at S-MID and S-BEGINNING. This can be explained by the final lengthening effect (Beckman & Edwards 1987; Cao 2003, among many others). In Figure 6, *saam1* is also longer in S-END than in S-MID and S-BEGINNING, but *sap6* in S-END is shorter than in S-BEGINNING. In fact, *saam1* and *sap6* are not the ultimate syllable in S-END: "*saam1*/*sap6* + QUANTIFIER" is the final disyllabic word, and *saam1*/*sap6* is the penultimate syllable. Thus, the final lengthening effect is obvious on the final word as a whole, but not consistent for the first syllable of the final word: if it is a stop syllable like sap6, it may not be lengthened.

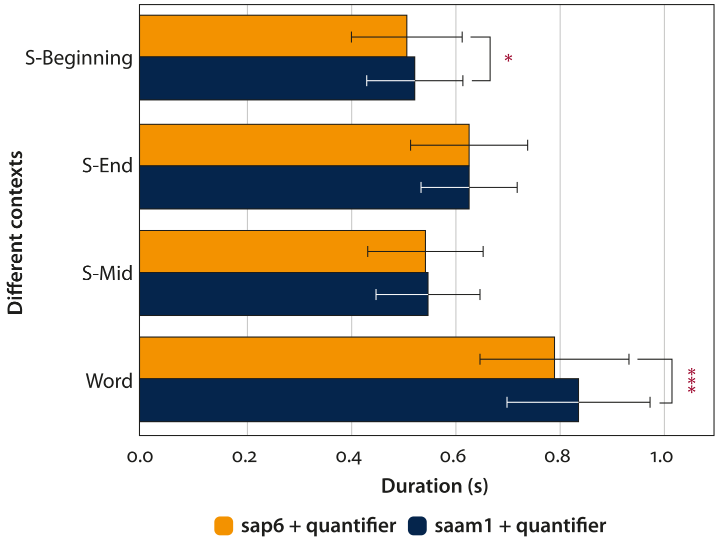


Figure 7. Mean duration bars of the words of 3x ("*saam1*+QUANTIFIER") and 10x ("*sap6*+QUANTIFIER") in different contexts, with error bars indicating ±1 standard deviation and *** and * indicating $p < 0.001$ and $p < 0.05$ in paired-samples *t*-tests, respectively

Table 2. Results of paired comparison and paired-samples *t*-tests for word duration data between "*saam1*+QUANTIFIER" and "*sap6*+QUANTIFIER"

Context	*saam1* + Quantifier (s)	*sap6* + Quantifier (s)	Difference (s)	Ratio	*T*	*df*	*p*
WORD	0.843	0.798	0.045	1.056	9.339	339	< 0.001
S-MID	0.556	0.55	0.006	1.010	1.605	339	0.109
S-END	0.6328	0.6329	−0.0001	0.999	−0.007	59	0.995
S-BEGINNING	0.529	0.516	0.013	1.025	2.175	59	0.034

3.4 Duration structure of the initials, rhymes, and acoustic blank in the words of "*saam1*/*sap6* + QUANTIFIER"

Comparing the results in § 3.2 and § 3.3, it should be noted that the syllables *saam1* and *sap6* have significantly different duration patterns, with the non-stop syllable *saam1* being much longer than the stop syllable *sap6* (§ 3.2). However, the disyllabic words "*saam1* + QUANTIFIER" and "*sap6* + QUANTIFIER" are approaching isochrony (§ 3.3), which implies that certain duration restructuring happens within the disyllabic word. In this section, we report the findings of further analyses of the duration composition of the disyllabic words that were carried out to investigate the restructuring. As introduced in § 2.1, the initials of QUANTIFIER covered the whole set of the 17 initials in Cantonese, which was embedded in the contexts of WORD and S-MID. The duration of the components of "*saam1*/*sap6* + QUANTIFIER" were measured and analyzed, including the initial of the NUMERALS (*s* for both *saam1* and *sap6*), the rhyme of the NUMERALS (*aam* for *saam1* and *ap* for *sap6*), the acoustic blank between the NUMERAL and the QUANTIFIER, the initial of the QUANTIFIER, and the rhyme of the QUANTIFIER. The stacked duration bars of the above components are displayed in Figures 8 and 9, and the corresponding statistical test results are listed in Tables 3 and 4, which show the data in the contexts of WORD and S-MID, respectively. In both Figures 8 and 9, the data are divided into three groups according to the articulation manners of the initials of the QUANTIFIER: QUANTIFIER with plosive initials ([p], [t], [k], [pʰ], [tʰ], and [kʰ]), QUANTIFIER with affricative/fricative initials ([ts], [tsʰ], [s], [f], and [h]), and QUANTIFIER with voiced initials ([j], [w], [l], [m], [n], and [ng]).

Generally speaking, Figures 8 and 9 share similar patterns, but each component in the context of WORD (Figure 8) is a little longer than its counterpart in the context of S-MID (Figure 9). The common patterns within *saam1*-*sap6* pairs and within the same context can be summarized as follows. First, the initial [s] of numerals shows only a little fluctuation within pairs; usually, the *s* of *sap6* is slightly longer than the *s* of *saam1*. Second, the rhymes of numerals show a distinct pattern:

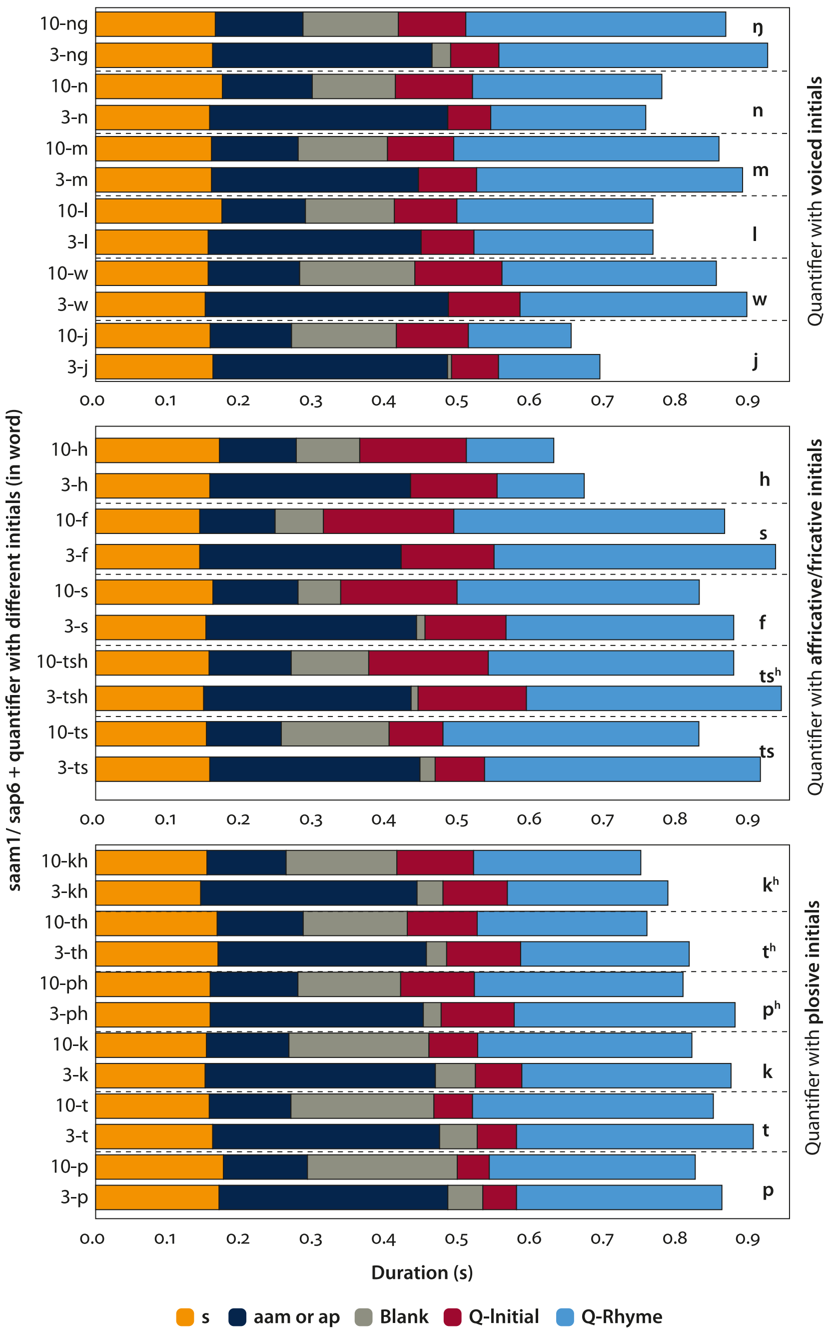


Figure 8. Stacked duration bars of composing initials, rhymes, and acoustic blank of "*saam1*/*sap6* + QUANTIFIER" in Word

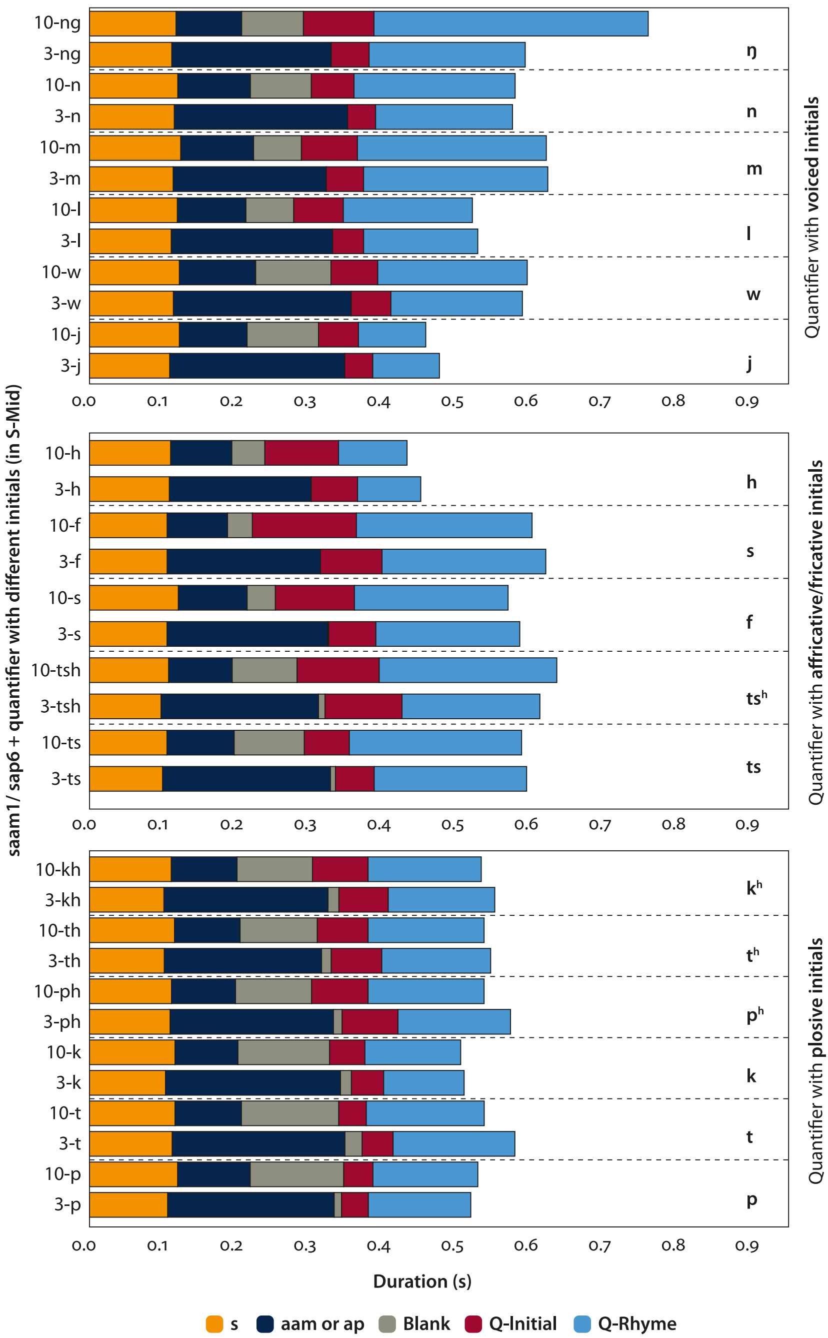


Figure 9. Stacked duration bars of composing initials, rhymes, and acoustic blanks of "*saam1*/*sap6* + QUANTIFIER" in S-Mid

ap is much shorter than *aam*, being one third to half the length of *aam*. Third, there is very little or even no acoustic blank between *saam1* and the following QUANTIFIER, but there is usually an obvious acoustic blank between *sap6* and the QUANTIFIER. Fourth, the duration of the initial of the QUANTIFIER varies according to the manner of articulation as well as the type of rhyme (stop or non-stop) of the preceding numerals. A more detailed analysis is provided below. Fifth, there does not seem to be an obvious difference in the rhymes of the QUANTIFIER between the groups "*saam1* + QUANTIFIER" and "*sap6* + QUANTIFIER".

Table 3. Results of paired comparison and paired-samples *t*-tests for comparison of duration structure of "*saam1* + QUANTIFIER" and "*sap6* + QUANTIFIER" in WORD

	saam1 (s)	*sap6* (s)	Difference (s)	*T*	*df*	*p*
Initial of NUMERAL	0.156	0.161	−0.005	−2.655	339	0.008
Rhyme of NUMERAL	0.299	0.114	0.185	72.048	339	0.000
Acoustic Blank + Initial of QUANTIFIER	0.105	0.239	−0.134	−40.382	339	0.000
Rhyme of QUANTIFIER	0.284	0.284	−0.001	−0.215	339	0.830

Table 4. Results of paired comparison and paired-samples *t*-tests for comparison of duration structure of "*saam1* + QUANTIFIER" and "*sap6*+QUANTIFIER" in S-MID

	saam1 (s)	*sap6* (s)	Difference (s)	*T*	*df*	*p*
Initial of NUMERAL	0.107	0.115	−0.009	−9.425	339	0.000
Rhyme of NUMERAL	0.222	0.091	0.131	64.067	339	0.000
Acoustic Blank + Initial of QUANTIFIER	0.064	0.163	−0.100	−42.847	339	0.000
Rhyme of QUANTIFIER	0.167	0.190	−0.023	4.162	339	0.000

From the above general observations, it can be inferred from the distinctive syllable duration between *saam1* and *sap6* to the nearly isochronous disyllabic word duration that the restructuring mainly happens among the middle components, including the Rhyme of the NUMERAL, the Acoustic Blank in-between, and the Initial of the QUANTIFIER (Q-Initial). According to the observation from Figures 8 and 9, compared with the non-stop numeral *saam1*, the stop numeral *sap6* has a much shorter rhyme, but there is much longer Acoustic Blank before the QUANTIFIER and/or the Q-Initial is lengthened. Regarding the plosive Q-Initials, they are very short (especially the unaspirated plosives) and have a limited lengthening effect after *sap6* by themselves, but the Acoustic Blank in between NUMERAL and

QUANTIFIER is long. As regards affricates and voiced consonants, their duration is a little longer, and they exhibit a slight lengthening effect after stop syllable *sap6*, together with an observable Acoustic Blank. As for fricatives, their duration is long, and they show an evident lengthening effect after *sap6*; there is also a little additional Acoustic Blank before, though not as long as before other types of consonants.

Summarizing the patterns observed from different types of initials, it seems that the Acoustic Blank and the Q-Initial are complementarily distributed (especially after the stop syllable *sap6*): a shorter Q-Initial is usually accompanied by a longer Acoustic Blank before; or longer Q-Initial has shorter or even no Acoustic Blank before. To test this hypothesis, Pearson correlation coefficient data were calculated, and the results are listed in Table 5. The data support that both in Word and in S-MID, after the stop syllable *sap6*, the Acoustic Blank and Q-Initial are significantly and negatively correlated (i.e. longer Acoustic Blank associates with shorter Q-Initial, or vice versa). Regarding the non-stop syllable *saam1*, significant negative correlations between Acoustic Blank and Q-Initial also exist in the context of WORD, though with a smaller amplitude of correlation. However, in the context of S-MID, there is no significant correlation between Acoustic Blank and Q-Initial. Probably, the complementary distribution of Acoustic Blank and the following initial mainly happen when the context allows a longer duration. In S-MID, after the non-stop syllable saam1, the space is too limited for this complementary distribution.

Table 5. Pearson correlation coefficient results for Acoustic Blank and Q-Initial

	Acoustic Blank (s)	Q-Initial (s)	Pearson's *r*	N	*p*
WORD-*saam1*	0.018942	0.085896	−0.337	340	< 0.001
WORD-*sap6*	0.134211	0.104556	−0.465	340	< 0.001
S-MID-*saam1*	0.00646	0.057182	−0.092	340	0.09
S-MID-*sap6*	0.087509	0.075693	−0.428	340	< 0.001

Taking Acoustic Blank and Q-Initial together as a whole, it seems to be in a complementary relationship with the duration of the Rhyme of the NUMERAL. To test this hypothesis, all the duration data of both "*saam1*+QUANTIFIER" and "*sap6*+QUANTIFIER" were put together and the Pearson correlation coefficient between the Rhyme of the NUMERAL and the "Acoustic Blank + Q-Initial" was calculated. In the context of WORD, Pearson's $r = -0.622$, $N = 680$, $p < 0.001$. In the context of S-MID, Pearson's $r = -0.676$, $N = 680$, $p < 0.001$. The correlation analyses confirmed that there is significant negative correlation between the Rhyme of the NUMERAL and the "Acoustic Blank + Q-Initial" (i.e. the shorter is the Rhyme of

NUMERAL, the longer the "Acoustic Blank + Q-Initial"). Thus, "Acoustic Blank + Q-Initial" can play a compensatory role when the preceding rhyme is a short stop syllable.

4. Discussions and conclusions

Through the acoustic experiments conducted in this study, we identified the prosodic features of Cantonese syllables in various contexts; this can help to interpret the mechanism behind the surface contradiction between syllable isochrony and the shorter length of stop syllables. The results indicate that although both initials and rhymes play important roles in deciding the total duration of a syllable, the stop rhyme and non-stop rhyme distinction is more decisive. The data also confirm that in all the tested contexts, stop syllables are much shorter than non-stop syllables. To phonetically realize isochrony in a disyllabic word, there is a supplementary lengthening effect following the stop rhymes: there is more acoustic blank, and in some circumstances the initial of the following syllable is lengthened. Through the supplementary lengthening mechanism, the shortness of the stop syllable can be compensated for and the disyllabic word can approximate the status of isochrony.

Perry et al. (2009) examined the relationship between the acoustic duration of syllables and the silent pauses (the equivalent of acoustic blank in our study) that follow them at syntactic junctures in Cantonese. They found a compensatory relationship between the acoustic duration of syllables and the duration of the silent pauses after them. Their studies focused on syntactic juncture and did not pay special attention to the difference between stop and non-stop syllables, but the underlying mechanism should be the same as that in the present study: the silent pause, or the acoustic blank after the syllable, can serve as a durational buffer for the syllable which is adaptable according to the preceding syllable. If the preceding syllable is long, the acoustic blank is short, and vice versa. Regarding the stop syllables carrying the morphological changed tone (e.g. *aap2* 'duck') mentioned in § 1.2, the general impression is that they are a little longer than the stop syllables carrying T7, T8, and T9. It can be inferred that the buffer after them is shorter as a result. This post-syllabic buffer effect can minimize the intrinsic durational differences decided by syllable structure and thus help to keep a balanced tempo in the speech rhythm.

This mechanism sounds like a *compensatory lengthening* effect. However, in the literature (De Chene & Anderson 1979; Hayes 1989; Gess 1998; Kiparsky 2011; Kavitskaya 2014, among many others), the term *compensatory lengthening* refers to the lengthening of a vowel that happens upon the loss of a following syllable coda or the loss of a vowel in an adjacent syllable. This means the so-called *compensatory lengthening* is the lengthening of the vowel nucleus to compensate for the loss of the following segment, while the lengthening effect investigated in this study refers to the lengthening of the following acoustic blank or segment to compensate for the shortness of the vowel nucleus. Therefore, they operate in opposite directions. To show the difference, the lengthening effect investigated in this study is named *supplementary lengthening*.

Hayes (1989) argued that compensatory lengthening may arise from speakers' attempts to preserve a word's moraic count. Moraic phonology can also be applied to explain the lengthening effect investigated in this study. Although the visible acoustic signal of a stop coda is short, phonologically it occupies a mora's length, and thus acoustically it is realized as an invisible signal as an additional acoustic blank, or as "borrowing" a certain length from the initial of the following syllable. In other words, the supplementary lengthening is to fulfill the mora of the stop coda, which is cross-syllabic.

It is well known that entering tones are vanishing, or indeed have vanished, in many Chinese dialects. Many previous studies discussed the processes, stages, and mechanisms of this phenomenon (Zhu et al. 2008; Song 2009, among many others). The disappearance of entering tones involves the loss of stop codas, the lengthening of the rhymes, and the merger of entering and non-entering tone categories. Previous studies on this topic mainly focused on syllables in isolation and seldom investigated the process in continuous speech. Cross-syllabic factors and speech rhythm reasons were not discussed. However, this sound change should be happening in continuous speech, the most frequent context in our daily speech. Our study indicates that stop syllables must depend on the cross-syllabic supplementary lengthening to realize the mora of stop coda, and thus to approach prosodic feature of the syllable isochrony. Supplementary lengthening or borrowing length from the following segment are expedient processes and do not seem to be solid and steady. Thus, the stop coda is being weakened and the preceding vowel nucleus is being lengthened gradually to adjust the durational structure to fulfill the rhythmic feature of syllable isochrony (i.e. the intra-syllabic compensatory lengthening process takes place to substitute for the cross-syllabic supplementary lengthening process). This hypothesis should be further verified in future studies; and this is the target for the next step of our research.

To summarize, our experiment in this study confirmed that stop syllables alone are shorter than non-stop syllables in various contexts, while in continuous speech (including in disyllabic words alone or at S-BEGINNING, S-MID, or S-END) there is a supplementary lengthening effect: either there is additional acoustic blank or the initial of the following syllable is lengthened. This supplementary lengthening effect can serve as a buffer for maintaining the syllable-timed prosodic feature. We further propose that this cross-syllabic supplementary lengthening process is an expedient process and is apt to be substituted by another phonological process, namely compensatory lengthening: with weakening stop codas, the preceding vowel nucleus is lengthening to adjust the syllabic internal durational structure. This proposal should be further tested in future research.

Acknowledgements

This study was supported by the Start-up Research Grant (RG 101/2016-2017R) and Internal Research Grants (RG 59/2017-2018R, RG 88/2018-2019R) from the Education University of Hong Kong (EdUHK). The author would like to thank EdUHK for its generous financial support. The author would also like to express her gratitude to Mr Tam, Ka Ming for his assistance in the acoustic measurement and to the two anonymous reviewers for their useful comments and suggestions. Finally, the writing process of this manuscript was completed in Hong Kong, Japan, and the USA when the author was on various journeys. The author would like to thank her family and colleagues for their support.

Abbreviations

S-MID	In the middle position of a sentence
S-BEGINNING	At the beginning of a sentence
S-END	At the end of a sentence
Q-Initial	The initial of a quantifier
Q-Rhyme	The rhyme of a quantifier

Appendix. Chinese reading list for the acoustic experiments

一、請讀下列數詞，每個讀三次:
一 二 三 四 五 六 七 八 九 十 零 百 千
二、請讀下列詞語，每個讀一次:
（一）1、三部 2、三樖 3、三萬 4、三份 5、三碗 6、三袋 7、三條 8、三年 9、三籃 10、三頁 11、三張 12、三串 13、三箱 14、三個 15、三期 16、三眼 17、三盒

（二）1、十部 2、十樖 3、十萬 4、十份 5、十碗 6、十袋 7、十條 8、十年 9、十籃 10、十頁 11、十張 12、十串 13、十箱 14、十個 15、十期 16、十眼 17、十盒

三、請讀下列句子，每個句子讀一次:
1、 佢有三部手機。佢有十部手機。
2、 佢有三樖生菜。佢有十樖生菜。

3、 佢有三萬港幣。佢有十萬港幣。
4、 佢有三份資料。佢有十份資料。
5、 佢有三碗雞湯。佢有十碗雞湯。
6、 佢有三袋生果。佢有十袋生果。
7、 佢有三條鯽魚。佢有十條鯽魚。
8、 佢有三年工齡。佢有十年工齡。
9、 佢有三籃生果。佢有十籃生果。
10、佢有三頁草稿。佢有十頁草稿。
11、佢有三張海報。佢有十張海報。
12、佢有三串珍珠。佢有十串珍珠。
13、佢有三箱餅乾。佢有十箱餅乾。
14、佢有三個問題。佢有十個問題。
15、佢有三期報紙。佢有十期報紙。
16、佢有三眼鐵釘。佢有十眼鐵釘。
17、佢有三盒餅乾。佢有十盒餅乾。
18、佢嘅工齡有三年。佢嘅工齡有十年。
19、三年工齡都唔少。十年工齡都唔少。
20、佢嘅問題有三個。佢嘅問題有十個。
21、三個問題都唔少。十個問題都唔少。
22、佢嘅餅乾有三箱。佢嘅餅乾有十箱。
23、三箱餅乾都唔少。十箱餅乾都唔少。

References

Abercrombie, David. 1967. *Elements of general phonetics*. Edinburgh: Edinburgh University Press.

Allen, George D. 1975. Speech rhythm: Its relation to performance universals and articulatory timing. *Journal of Phonetics* 3(2). 75–86. https://doi.org/10.1016/S0095-4470(19)31351-8

Bauer, Robert S. 1979. Alveolarization in Cantonese: A case of lexical diffusion. *Journal of Chinese Linguistics* 7(1). 132–141.

Bauer, Robert S. 1995. Syllable and word in Cantonese. *Journal of Asian Pacific Communication* 6(4). 245–306.

Bauer, Robert S. & Benedict, Paul K. 1997. *Modern Cantonese phonology*. New York: Mouton de Gruyter. https://doi.org/10.1515/9783110823707

Beckman, Mary E. 1992. Evidence for speech rhythms across languages. In Tohkura, Yohichi & Vatikiotis-Bateson, Eric & Sagisaka, Yoshinori (eds.), *Speech perception, production and linguistic structure*, 457–463. Tokyo: IOS Press.

Beckman, Mary E. & Edwards, Jan. 1987. The phonological domains of final lengthening. *The Journal of the Acoustical Society of America* 81(S1). S67–S67. https://doi.org/10.1121/1.2024348

Boersma, Paul & Weenink, David. 2018. *Praat: Doing phonetics by computer* (Version 6.0.37). (http://www.praat.org/) (Accessed 2018-03-14.)

Cao, Jianfen. 2000. Rhythm of spoken Chinese: Linguistic and paralinguistic evidences. In Editorial Committee of ICSLP 2000 (ed.), *Sixth International Conference on Spoken Language Processing (ICSLP 2000)*, vol. 2, 357–360. Beijing: ISCA Archive.

Cao, Jianfen. 2003. Jiyu yufa xinxi de hanyu yunlv jiegou yuce [Prediction of prosodic organization based on grammatical information]. *Zhongwen Xinxi Xuebao* [Journal of Chinese Information Processing] 17(3). 41–46.

Chao, Yuen Ren. 2006[1930]. A system of "tone-letters". In Wu, Zongji & Chao, Xinna (eds.), *Linguistic essays by Yuenren Chao*, vol. 30, 98–103. Beijing: The Commercial Press. (First published in 1930.)

Cheung, Hung-nin Samuel. 2002. 21 shiji de xianggang yueyu: Yige xin yuyin xitong de xingcheng [Cantonese of the 21st century: Formation of a new phonological system in the Hong Kong language]. *Jinan Xuebao (Zhexue Shehui Kexue Ban)* [Jinan Journal (Philosophy & Social Science Edition)] 24(2). 25–40.

Cheung, Kwan-hin. 1986. *The phonology of present-day Cantonese*. London: University College London. (Doctoral dissertation.)

Cheung, Kwan-hin. 1994. Yueyu yuyin de yinchang danwei [The duration unit of Cantonese sounds], In Shan, Zhouyao (ed.), *Diyijie Guoji Yuefangyan Yantaohui lunwenji* [Procceedings of the First Yue Conference], 79–85. Hong Kong: Modern Education Research Society.

Cheung, Kwan-hin & Zhang, Ling. 2010. Yueyu yinjie yanchang dui yunfu yunwei shichang de yinxiang [The effect of syllable lengthening: On the duration of syllable nucleus and coda in Cantonese]. In Cheung, Hung-nin Samuel & Chang, Song Hing (eds.), *Diachronic change and language contact dialects in Southeast China* (Journal of Chinese Linguistics Monograph Series 24), 145–161. Berkeley: Project on Linguistic Analysis.

Dauer, R.M. 1983. Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics* 11(1). 51–62. https://doi.org/10.1016/S0095-4470(19)30776-4

De Chene, Brent & Anderson, Stephen R. 1979. Compensatory lengthening. *Language* 55(3). 505–535. https://doi.org/10.2307/413316

Dellwo, Volker. 2006. Rhythm and speech rate: A variation coefficient for ΔC. In Karnowski, Pawel & Szigeti, Imre (eds.), *Language and language-processing*, 231–241. Frankfurt: Peter Lang.

Gess, Randall. 1998. Compensatory lengthening and structure preservation revisited. *Phonology* 15(3). 353–366. https://doi.org/10.1017/S0952675799003656

Grabe, Esther & Low, Ee Ling. 2002. Durational variability in speech and the rhythm class hypothesis. In Gussenhoven, Carlos & Warner, Natasha (eds.), *Laboratory phonology VII*, 515–546. Berlin: Mouton de Gruyter.

Hashimoto, Oi-kan Yue. 1972. *Phonology of Cantonese*. Cambridge: Cambridge University Press.

Hayes, Bruce. 1989. Compensatory lengthening in moraic phonology. *Linguistic inquiry* 20(2). 253–306.

Jeng, Jing-Yi. 2005. Butong yanyu sudu, fayu danwei he fayu weizhi dui guoyu yinduan shichang de yingxiang [The influence of speaking rate, utterance unit and position on segmental duration of Mandarin], *Nanda Xuebao (Renwen yu Shehui Lei)* [Journal of National University of Tainan (Humanities and Social Sciences)] 39(1). 161–185.

Jeng, Jing-Yi. 2008. Butong yanyu sudu xia guoyu yunlü de jishi dengshixing [The timing isochrony of spoken Mandarin at different speaking rates] *Zhonghua Xinli Xuekan* [Chinese Journal of Psychology] 50(1). 1–19.

Jeng, Jing-Yi. 2011. Huayu yuju chansheng yu zhijue de jishi texing [The speech timing characteristics of Mandarin sentence production and perception]. *Zhonghua Xinli Xuekan* [Chinese Journal of Psychology] 53(3). 323–347.

Kavitskaya, Darya. 2014. *Compensatory lengthening: Phonetics, phonology, diachrony*. New York: Routledge. https://doi.org/10.4324/9781315024141

Kiparsky, Paul. 2011. Compensatory lengthening. In Cairns, Charles E. & Raimy, Eric (eds.), *Handbook of the syllable*, 31–69. Leiden: Brill.

Laver, John. 1994. *Principles of phonetics*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139166621

Light, Timothy. 1977. The Cantonese final: An exercise in indigenous analysis. *Journal of Chinese Linguistics* 5(1). 75–102.

Lin, Hua & Wang, Qian. 2007. Mandarin rhythm: An acoustic study. *Journal of Chinese Language and Computing* 17(3). 127–140.

Mariano, Paolo. 2011. *Rhythm typology: Acoustic and perceptive studies*. Torino: University of Turin. (Doctoral dissertation.)

Mok, Peggy Pik-Ki. 2008. On the syllable-timing of Cantonese and Beijing Mandarin. In Bao, Huaiqiao & Shen, Jiaxuan (eds.), *Proceedings of the 8th Phonetics Conference of China (PCC 2008) and the International Symposium on Phonetic Frontiers (ISPF 2008)*, 148–154. Beijing: Institute of Linguistics, Chinese Academy of Social Sciences.

Peng, Gang. 2006. Temporal and tonal aspects of Chinese syllables: A corpus-based comparative study of Mandarin and Cantonese. *Journal of Chinese Linguisitcs* 34(1). 134–154.

Perry, Conrad & Wong, Richard Kwok-Shing & Matthews, Stephen. 2009. Syllable timing and pausing: Evidence from Cantonese. *Language and Speech* 52(1). 29–53. https://doi.org/10.1177/0023830908099882

Pike, Kenneth L. 1945. *The intonation of American English*. Ann Arbor: University of Michigan Press.

Ramus, Franck & Nespor, Marina & Mehler, Jacques. 1999. Correlates of linguistic rhythm in the speech signal. *Cognition* 73(3). 265–292. https://doi.org/10.1016/S0010-0277(99)00058-X

Roach, Peter. 1982. On the distinction between stress-timed languages and syllable-timed languages. In Crystal, David (ed.), *Linguistic controversies: Essays in linguistic theory and practice in honour of F.R. Palmer*, 73–79. London: Edward Arnold.

Shi, Feng & Mai, Yun. 2003. Guangzhouhua chang [a] he duan [ɐ] yuanyin de tingbian shiyan 廣州話長[a]和短[ɐ]元音的聽辨實驗 [Perceptual test on long [a] and short[ɐ] in Cantonese]. *Zhongguo Yuwen Yanjiu* [Studies in Chinese Linguistics] 2. 50–59.

Song, Yidan. 2009. Nanjing fangyan zhong de rusheng housewei shiyan yanjiu [On experiment of entering tone of Nanjing dialect]. *Nanjing Shifan Daxue Wenxueyuan Xuebao* [Jounal of School of Chinese Language and Culture, Nanjing Normal University] 2009(2). 166–172.

The Linguistic Society of Hong Kong. 1993. *The Jyutping scheme*. (https://web.archive.org/web/20130426050642/http://www.lshk.org/node/47) (Accessed 2019-02-14.)

Wong, Puisan & Chan, Hoi-Yin. 2018. Acoustic characteristics of highly distinguishable Cantonese entering and non-entering tones. *The Journal of the Acoustical Society of America* 143(2). 765–779. https://doi.org/10.1121/1.5021251

Zhang, Ling. 2010. Guangzhouhua changduan yuanyin de yuyin shiyan xin tan [Long and short vowels in Cantonese: A revisit with new experiments]. *Fangyan* [Dialect] 2010(2). 134–144.

Zhang, Ling. 2016. *Intonation effects on Cantonese lexical tones in speaking and singing*. Munich: LINCOM GmbH.

Zhang, Ling. 2017. Yueyu shengdiao yu jiangshi yingao [Hong Kong Cantonese lexical tones and declining intonation]. *Yuyan Kexue* [Linguistic Sciences] 16(2). 182–192.

Zhu, Xiaonong & Jiao, Lei & Yan, Zhicheng & Hong Ying. 2008. Rusheng yanhua santu [Three ways of Rusheng sound change]. *Zhongguo Yuwen* [Studies of the Chinese Language] 2008(4). 324–338.

Author's address

Ling Zhang
Department of Chinese Language
Education University of Hong Kong
B3-2/F-13, Dept. of CHL, EdUHK
10, Lo Ping Road
Tai Po
Hong Kong
zhangl@eduhk.hk

Publication history

Date received: 29 April 2019
Date accepted: 19 December 2019
Published online: 15 December 2021