Bilingualism: Language and Cognition 14 (4), 2011, 458–472 © Cambridge University Press 2011 doi:10.1017/S1366728910000453

The acquisition of speech rhythm by three-year-old bilingual and monolingual children: Cantonese and English*

PEGGY P. K. MOK
Chinese University of Hong Kong

(Received: September 14, 2009; final revision received: August 5, 2010; accepted: September 9, 2010; First published online 19 July 2011)

This study investigates the acquisition of speech rhythm by Cantonese–English bilingual children and their age-matched monolingual peers. Languages can be classified in terms of rhythmic characteristics that define English as stress-timed and Cantonese as syllable-timed. Few studies have examined the concurrent acquisition of rhythmically different languages in bilingual children. This study uses data of six Cantonese–English bilingual children around age 3;0 and compares them with six monolingual children in each language using recently developed acoustic rhythmic metrics on consonantal, vocalic and syllabic intervals. Qualitative data on syllable structure complexity and vowel quality are also included. Results on syllable duration show that monolingual children display distinct rhythmic patterns while the differences between the two languages of the bilingual children are less distinct. Bilingual English has less durational variability than monolingual English. Bilingual children have a distinct phonological developmental trajectory from monolingual children, which is manifested in acquisition delay and is influenced by language dominance. This shows that the two phonologies interact at the prosodic level.

Keywords: speech rhythm, bilingual acquisition, Cantonese, English

1. Introduction

This paper investigates the acquisition of speech rhythm by three-year-old Cantonese–English simultaneous bilingual children, and compares them with their age-matched monolingual peers. Research on early bilingual phonological acquisition often focuses on segmental aspects, e.g. phonemic inventory and error patterns, while prosodic aspects receive relatively little attention. This study broadens the research into bilingual phonological acquisition by examining an important prosodic feature, SPEECH RHYTHM, using recently developed acoustic metrics supplemented by some qualitative data on syllable structure complexity and vowel reduction. The inclusion of developmental data on both quantitative and qualitative measures can give us useful insight into the early acquisition of prosody.

1.1 Speech rhythm

Speech rhythm, although descriptively rather elusive, is perceptually quite salient. Rhythm can be roughly defined as the repeating pattern of occurrence in time of relatively strong and weak events. In speech, these events generally relate to syllables and stress. Other prosodic characteristics like tone and intonation also contribute to the impression of speech rhythm. All these features are inter-related in an ill-defined way which contributes to the elusiveness of speech rhythm. Despite this, the durational aspect of speech rhythm has received much research attention. It also forms the focus of the present study. Linguists have long argued that languages fall into distinct rhythmic classes (Pike, 1945; Abercrombie, 1967). Germanic languages like English, German and Dutch are typical examples of stress-timed languages while Romance languages like Spanish, Italian and French are examples of syllable-timed languages. A third type of rhythm class has also been suggested: MORA-TIMING, of which Japanese is a typical example (Ladefoged, 1975). The rhythm class hypothesis was based on the notion of isochrony, i.e. there are units of equal or near-equal duration in the speech signal for such classification: syllables for syllable-timed languages, inter-stress intervals (feet) for stress-timed languages and mora for mora-timed languages. A mora is a sub-syllabic

* The author would like to thank the parents of several monolingual children for the monolingual data. She is also grateful to Eunice Wong, Iris Luk, Wai Kit Kwok and Tintus Chan for data collection and segmentation. She thanks Angel Chan and Donald White for helpful discussions, and Lawrence White, Ruth Tincoff and three anonymous reviewers for constructive comments on earlier versions of the manuscript. This study was funded by the Direct Grant for Research 2007–2008, Second Round, Chinese University of Hong Kong.

Address for correspondence:
Department of Linguistics and Modern Languages, Leung Kau Kui Building, Chinese University of Hong Kong, Shatin, Hong Kong
peggymok@cuhk.edu.hk

timing unit relating to syllable weight. Each mora is believed to take about the same length of time to say. However, decades of experimental studies have failed to find concrete evidence for these isochronous units in the speech signal to support the impression of different rhythm classes (see Dauer, 1983; Grabe & Low, 2002; Warner & Arai, 2001 for review).

Despite the lack of experimental support, Dauer (1983) and Roach (1982) pointed out three important phonological features that differentiate stress-timed and syllable-timed languages: syllable structure, vowel reduction and stress. Stress-timed languages have greater variation in syllable length and structure, more reduced unstressed syllables, more variation in the phonetic realisation of stress and more stress-related rules than syllable-timed languages. For example, the stress patterns in English are much more complex and variable than in Italian. The coalescence of these phonological differences results in a perceptual distinction between stress-timing and syllable-timing. In addition, contrary to the early assumption that this distinction was categorical, Dauer (1983) suggested that languages can be more or less stress-timed or syllable-timed, with a continuum between the two.

These insights of Dauer (1983) and Roach (1982) created a new paradigm in speech rhythm research. Several acoustic metrics that quantified the auditory impression of different rhythm classes were developed based on the phonological differences between syllable-timed and stress-timed languages. Instead of searching for isochrony in the speech signal, these acoustic measurements gauge the durational variability of speech. According to the various phonological differences mentioned above, stress-timed languages should have a higher variability of consonant and vowel duration than syllable-timed languages. Two main types of rhythmic metrics were proposed: (i) global durational variability metrics of ΔC (standard deviation of consonantal duration) and $\%V$ (percentage of vocalic duration in speech) proposed by Ramus, Nespor and Mehler (1999); and (ii) local variability metrics of Pairwise Variability Index (PVI) of vocalic and intervocalic (consonantal) intervals proposed by Grabe and Low (2002). The Δ metric gauges durational variability of the whole utterance or discourse, while PVI captures the local differences in duration between successive units. There are two versions of PVI: raw and normalised (normalised for speech rate). Raw PVI is used for consonantal intervals while normalised PVI is used for vocalic intervals (see Equation 1). Stress-timed languages will have a lower $\%V$ and a higher ΔC, and higher PVIs for both vocalic and consonantal intervals than syllable-timed languages. The results by Ramus et al. (1999) and Grabe and Low (2002) show promising groupings of languages according to their traditional rhythm classes, while languages having less typical or unknown rhythm may fall between these groupings, e.g. Czech and Estonian. Subsequent studies found that the rhythmic metrics of vocalic intervals are more robust than those on consonantal intervals (e.g. Bunta & Ingram, 2007; White & Mattys, 2007).

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

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

(where m = number of items; d = duration of the kth interval)

Refinements to the rhythmic metrics were subsequently proposed. First of all, speech rate is an important factor contributing to durational variability which can affect the results of the rhythmic metrics (Barry, Andreeva, Russo, Dimitrova & Kostadinova, 2003; Dellwo & Wagner, 2003). VarcoX, normalising ΔX by dividing it by the mean X duration (where X stands for any interval), was proposed by Dellwo (2006). Normalised versions of the rhythmic metrics (both VarcoX and normalised PVI) were found to be more robust than the unnormalised versions in calibrating speech rhythm (White & Mattys, 2007; Wiget, White, Schuppler, Grenon, Rauch & Mattys, 2010). Besides, in addition to the variability of consonant and vowel duration, the duration of larger linguistic units, especially syllables, can also give reliable distinction between languages with different rhythmic patterns (Deterding, 2001; Mok & Dellwo, 2008; Nolan & Asu, 2009). These revisions to the original Δ and PVI metrics are adopted in the present study. More detail will be given in Section 2.3 below.

1.2 Acquisition of speech rhythm

Despite its elusiveness, speech rhythm forms a prosodic cornerstone in early language acquisition. Several studies have demonstrated that this rhythmic foundation allows newborn infants to distinguish languages from each other (e.g. Bosch & Sebastián-Gallés, 2001; Nazzi et al., 1998; Nazzi et al., 2000). Many studies have examined speech rhythm in adults using acoustic rhythmic metrics, especially acquisition of rhythm in adult L2 learners (e.g. Jian, 2004; Low, Grabe & Nolan, 2000; Mok & Dellwo, 2008; White & Mattys, 2007). What has received surprisingly less attention is the long process through which an infant achieves native rhythmic competence. In fact, only a few studies have investigated the acquisition of speech rhythm by children. Before the introduction of the rhythmic metrics, Allen and Hawkins (1980) suggested that children begin with a more syllable-timed rhythm regardless of their target languages, because two important elements of stress-timing, consonant clusters and vowel reduction, are difficult for children to master.

This hypothesis was empirically supported by Grabe, Post and Watson (1999), who compared three monolingual English and three monolingual French children using normalised vocalic PVI. They found that at age four, these children displayed significantly different rhythmic patterns. However, the patterns of the French children were similar to their mothers while the patterns of the English children were different from their mothers, tending towards syllable-timing. Their results suggest that stress-timing is more difficult to acquire, echoing Allen and Hawkins (1980). Their results also show that at age four, monolingual children learning rhythmically different languages already exhibit corresponding patterns in their speech production.

If monolingual children exhibit different rhythmic patterns at an early age, a logical question to ask is: How would bilingual children acquire rhythmically different languages? Whitworth (2002) was the first to investigate bilingual acquisition of speech rhythm using acoustic rhythmic metrics. She studied six German–English bilingual children from age five to 13 using raw consonantal and normalised vocalic PVI. Since both German and English are stress-timed languages, it is not surprising that she found no significant difference between the durational variability of the two languages in both the bilingual children and their parents. She suggested that speech rhythm is not completely acquired at around age 11. Such a late age of acquisition was proposed presumably because the author was referring to the subtle differences between the two languages belonging to the same rhythmic category. Based on the results of younger monolingual children by Grabe et al. (1999) above, it can be expected that children acquiring rhythmically different languages could master their patterns at an earlier age.

To-date, only two studies have specifically examined the acquisition of speech rhythm by bilingual children exposed to rhythmically different languages. Using raw consonantal and normalised vocalic PVI, Lleó, Rakow and Kehoe (2007) compared three German–Spanish bilingual children at age three with an equal number of their monolingual counterparts. The monolingual children displayed significantly different rhythmic patterns as reflected in both consonantal and vocalic PVIs, while the bilingual children tended to have similar patterns in both of their German and Spanish. These results suggest that the two languages of the bilingual children did not remain separate but interacted. Bunta and Ingram (2007) examined two groups of Spanish–English bilingual children (3;9–4;5 and 4;6–5;2) with their monolingual peers and adults using normalised PVI for both consonantal and vocalic intervals. There were five children in each group. The authors found that both younger and older bilingual children displayed distinct rhythmic patterns for their two languages using normalised vocalic PVI. The children also differed from their monolingual English peers, but not their Spanish peers. However, the comparisons using normalised consonantal PVI did not yield consistent significant differences. Their results on vocalic intervals suggest that bilingual children show a bias toward syllable-timing, but they can also separate the speech rhythm of Spanish and English at around four years of age. Taken together, results of the several studies mentioned above suggest that, at the earliest, monolingual children acquiring rhythmically different languages already display distinct rhythm at age three, while bilingual children may only separate the rhythm of their two languages at around age four. Nevertheless, these studies examined bilingual children acquiring different pairs of languages at different ages. Also, there were only three children in each group in Lleó et al. (2007) while the two age groups span many months in Bunta and Ingram (2007), so the above conclusion should be treated as tentative only. Clearly, more studies using recently developed acoustic rhythmic metrics, especially those examining younger children with more participants, are needed to confirm whether and how bilingual children differ from monolingual children in their acquisition of speech rhythm.

1.3 The present study

The present study aims to investigate the acquisition patterns of speech rhythm of three-year-old bilingual and monolingual children learning two rhythmically and typologically different languages: Cantonese (syllable-timed, Sinitic) and English (stress-timed, Germanic). Grabe et al. (1999), Lleó et al. (2007), and Bunta and Ingram (2007) compared pairs of European languages which, although different rhythmically, are nevertheless similar in other respects. Cantonese is a Chinese language which is tonal and isolating. It has a very simple syllable structure with no lexical stress and no phonological vowel reduction. Each syllable, which roughly corresponds to a morpheme, carries a lexical tone. In emotionally neutral sentences, each syllable receives nearly equal emphasis (Bauer & Benedict, 1997). Impressionistically, Cantonese is a typical syllable-timed language. Acoustic rhythmic metrics also confirm the syllable-timing of Cantonese (Mok, 2009; Mok & Dellwo, 2008).

One aspect of Cantonese phonology needs further explanation regarding variability of vowel duration. Duration is a relevant and primary feature of the Cantonese vowel system. Kao's (1971) data showed that (i) long vowels in open syllables are longer than either long or short vowels in closed syllables; (ii) long vowels are longer than short vowels in the same closed syllables; and (iii) for both long and short vowels in closed syllables, vowels ending with /j w m n ŋ/ are longer than those ending in /p t k/. Thus, Kao (1971) concluded that there are three factors affecting vowel duration in Cantonese: syllable structure (open or closed), vowel quantity (long

Table 1. *Background of the monolingual children.*

Cantonese				British English			
Child	Age of recording	Sex	Source	Child	Age of recording	Sex	Source
TTC	2;11.16	F	HKU-70	Ella	3;1	F	Forrester
LCN	2;11.29	F	HKU-70	Molly	3;1	F	York
CWT	3;0.29	M	HKU-70	Angus	2,11	M	HK
DGC	3;1.05	M	HKU-70	Jessica	3;1	F	HK
LCL	3;1.06	F	HKU-70	Madi	3;0	F	Twin
LTH	3;0.07	M	Local	Fletchy	3;0	M	Twin

or short) and coda endings (/j w m n ŋ/ vs. /p t k/). However, since vowel duration is dependent on three factors, vowel duration in Cantonese is not just a matter of vowel length difference. Most phonemic analyses treated length as an allophonic feature because except for the low vowel pair [a] and [ɐ], other short vowels are in complementary distribution with their long counterparts (Chao, 1947; Hashimoto, 1972; Yuan, 1960). To the best knowledge of the present author, there is no published study investigating the acquisition of vowel length by Cantonese children. Given that vowel duration depends on several factors and is largely complementary, the variability of vowel duration in natural utterances should not be biased by vowel length difference alone.

Previous studies of Cantonese–English simultaneous bilingual children show strong evidence of syntactic transfer and interaction (Yip & Matthews, 2007). It is quite likely that such interaction is also observable in the children's phonological development as well, but thus far there have been very few studies that examined the phonological development of Cantonese–English bilingual children (e.g. Holm & Dodd, 1999), especially in the area of prosodic development. The present study fills this gap by examining speech rhythm using recent acoustic metrics.

Furthermore, the existing research into bilingual speech rhythm development has not utilised the full battery of available measurements. Previous studies into the acquisition of speech rhythm have solely focused on the measurement of PVI of consonant and vowel duration. None of them explained why %V and ΔC proposed by Ramus et al. (1999) were not used in their studies. Although both types of metrics concern durational variability, Δ and PVI capture variability differently (globally vs. locally). In addition, Section 1.1 has shown that rhythmic metrics of syllable duration also distinguish rhythmic patterns robustly. Since there are different types of acoustic rhythmic metrics available (measuring variability globally or locally; raw or normalised), the present study uses a variety of metrics of all three intervals (consonant, vowel and syllable) in order to obtain a more comprehensive understanding of the rhythmic patterns of the children. The conclusions drawn will be based on consistent patterns of the full spectrum of available measurements. In addition, the present study compares more children in each group (N = 6) than Lleó et al. (2007) (N = 3) and Bunta and Ingram (2007) (N = 5), and analyzes a large number of utterances per child (see details in Section 2, "Method", below). The age range of both monolingual and bilingual data is also carefully controlled (from 2;10.15 to 3;1.30). Some qualitative data regarding syllable structure complexity and vowel reduction in the children's production are also included to illustrate the patterns obtained using the rhythmic metrics. It is hoped that the extra data can provide reliable evidence of how three-year-old children acquire speech rhythm bilingually and monolingually. The following three main research questions are asked:

1. Do monolingual children acquiring Cantonese and English display distinct rhythmic patterns at age three?
2. Do bilingual children have the same patterns as the monolinguals at the same chronological age?
3. What can affect the rhythmic development of bilingual children?

2. Method

2.1 Participants

Data from six monolingual Cantonese, six monolingual British English and six Cantonese–English simultaneous bilingual children were used in this study. Data of the monolingual children came from various sources. Table 1 shows the details of these children. Five Cantonese children are featured in the HKU-Cantonese-70 corpus available in CHILDES (Fletcher, Leung, Stokes & Weizman, 2000). The other Cantonese child was recruited locally through word of mouth. Data of one English child came from the Forrester corpus in CHILDES (Forrester,

Table 2. *Background of the six bilingual children.*

Child	Native language of mother	Native language of father	Sex	Language dominance	Age range for data used
Timmy	Cantonese	British English	M	Cantonese	2;11.18–3;0.16
Sophie	Cantonese	British English	F	Cantonese	3;0.02–3;01.30
Alicia	Cantonese	British English	F	Cantonese	3;0.10–3;0.24
Llywelyn	Cantonese	British English	M	Cantonese	2;11.29–3;01.04
Charlotte	Cantonese	British English	F	English	2;10.15–3;0.03
Janet	Cantonese	British English	F	Cantonese	2;10.30–3;01.01

2002). Two of the English children were recruited in an English-speaking kindergarten in Hong Kong for children of expatriate families, and one English child was from a kindergarten in York, England. The other two English children were fraternal twins from an expatriate family in Hong Kong recruited through word of mouth. All the English children based in Hong Kong were monolingual. The contents of the monolingual recordings were similar to the bilingual ones: natural conversations recorded in unstructured sessions.

The six Cantonese–English bilingual children (two boys and four girls) are featured in the Hong Kong Bilingual Child Language Corpus which is available through the YipMatthews corpus in CHILDES (http://childes.psy.cmu.edu/media/Biling/YipMatthews/). Yip and Matthews (2007) give detailed background of five of these children. Table 2 shows the language background of all bilingual children. They were offspring of mixed marriages who were exposed to Cantonese and English from birth, and grew up in a 'one parent one language' environment. All but one child were Cantonese-dominant. They were recorded longitudinally at weekly or bi-weekly intervals in two unstructured play situations, one for Cantonese and one for English, although some language mixing can be found in the recordings, especially the earlier ones.

2.2 Materials

Around 20 to 30 utterances of between four and nine syllables within the same breath group with minimal pausing were used for each child (see Table 3). Shorter utterances were not used because they are unsuitable for calculating durational variability, while longer utterances were quite rare in the speech of the children. The choice of utterances depends on factors like the quality and length of the original recordings and how talkative the child was in the recordings. Utterances with unnatural intonation, excessive stress on a particular syllable or word, excessive final lengthening or too much background noise were excluded. For example, utterances spoken while the children were shouting or spoken in a sing-song manner during play were excluded. Only interpretable natural utterances showing clear formant structure in the spectrograms were used. This explains why the number of utterances varies from child to child. Nevertheless, this will not affect the results because the rhythmic metrics calculated for each child were averaged across all utterances produced by the child. The statistical comparisons were conducted using the averaged scores.

Each utterance was segmented into intervals at three levels (consonantal, vocalic and syllabic intervals in milliseconds) using Praat (Boersma & Weenink, 2009). Final syllables were not excluded because Bunta and Ingram (2007) showed clearly that the inclusion or exclusion of the final syllables did not affect their results using normalised rhythmic metrics. Also, utterances with four syllables will be too short for calculating variability if the final syllables were excluded. Segmentation criteria for consonantal and vocalic intervals follow those in Grabe and Low (2002) and Peterson and Lehiste (1960), i.e. mainly acoustic. The closure phase of utterance-initial stops and utterance-final unreleased stops were excluded since there is no reliable cue for marking them. Liquids and glides were considered consonantal if they were clearly distinguishable from the vowel, otherwise they were included as part of the vocalic portion. Only portions with clear vowel formants were considered vocalic. A stretch of consonantal or vocalic portion can span across word/syllable boundaries, as long as it is continuous. Pauses and hesitations were minimal as they were deliberately avoided when the utterances were chosen. Any pause or hesitation within an utterance was excluded.

Since the boundaries between syllables are often a contentious issue, the segmentation of syllabic intervals inevitably requires phonological judgments. Syllable structure in Cantonese is very simple: only /p t k m n ŋ/ can appear syllable-finally (coda /j w/ appear only in diphthongs), and consonant clusters and resyllabification are prohibited. The segmentation of syllables in the Cantonese data was, therefore, quite straightforward. In instances where two very similar consonants abutted, e.g. a syllable-final unreleased stop followed by a syllable-initial unaspirated stop, the syllable boundary was placed in the

Table 3. *Number of utterances, averaged number of syllables per utterance, and speech rate (syllables/second) for each child; standard deviations in brackets.*

	Cantonese				English			
	Child	Number of utterances	Averaged syllable number	Speech rate	Child	Number of utterances	Averaged syllable number	Speech rate
Monolingual	TTC	27	6.1 (1.3)	3.8 (0.4)	Ella	22	6.4 (1.3)	3.3 (0.8)
	LCN	31	6.3 (1.4)	3.9 (0.6)	Molly	20	5.2 (1.4)	3.3 (0.6)
	CWT	33	5.8 (1.0)	3.8 (0.5)	Angus	32	5.8 (1.0)	4.1 (0.7)
	DGC	27	6.9 (1.5)	3.9 (0.6)	Jessica	26	6.7 (1.1)	4.0 (0.7)
	LCL	27	6.3 (1.1)	3.7 (0.5)	Madi	29	6.3 (1.2)	3.9 (0.7)
	LTH	29	5.8 (1.1)	4.4 (1.0)	Fletchy	33	5.3 (1.2)	3.3 (0.6)
	Mean	**29 (2.5)**	**6.2 (0.4)**	**3.9 (0.3)**	**Mean**	**27 (5.3)**	**5.9 (0.6)**	**3.6 (0.4)**
Bilingual	Timmy	31	6.1 (1.0)	3.8 (0.7)	Timmy	29	6.5 (1.2)	3.5 (0.6)
	Sophie	30	5.9 (1.3)	4.1 (0.6)	Sophie	20	6.6 (1.6)	3.7 (0.7)
	Alicia	42	6.2 (1.2)	3.7 (0.6)	Alicia	28	6.0 (1.3)	3.2 (0.5)
	Llywelyn	20	6.2 (1.4)	3.9 (0.4)	Llywelyn	20	6.7 (1.5)	3.8 (0.5)
	Charlotte	16	5.3 (0.9)	4.0 (0.7)	Charlotte	23	6.2 (1.4)	4.0 (0.7)
	Janet	31	6.4 (1.2)	4.1 (0.8)	Janet	19	4.3 (0.5)	4.1 (0.8)
	Mean	**28.3 (9.2)**	**6.0 (0.4)**	**3.9 (0.2)**	**Mean**	**23.2 (4.4)**	**6.0 (0.9)**	**3.7 (0.3)**

middle of the long consonantal portion. Syllabification in English is more problematic since it allows more complex consonant sequences. Nevertheless, complex consonant sequences were quite rare in the children's data because words with complex clusters were used infrequently and were often simplified in realisation (more detail to follow in Section 3, "Results", below). The segmentation principles for syllables in Deterding (2001) were adopted. The maximal onset principle was followed as long as it produced phonotactically permissible onsets. The contentious ambisyllabic intervocalic consonants, e.g. the /m/ in *lemon*, were treated as the onsets of the second syllables. Three phonetically-trained observers segmented the utterances based on the above criteria. Each of them segmented a separate portion of the data. All segmentation was cross-checked by the author. About 20% of the data was slightly modified by the author for consistence across observers.

In addition to rhythmic metrics, two types of qualitative data are used to illustrate the patterns obtained using the rhythmic metrics: occurrence of different syllable structures and vowel quality. Details about these data will be given in the "Results" section below for easy reference.

2.3 Rhythmic metrics

Unlike previous studies which used only PVI, both the metrics proposed by Ramus et al. (1999; ΔC, %V) and Grabe and Low (2002; PVI) were used on three intervals (consonantal (C), vocalic (V) and syllabic (S)) in this study. Both the Δ and PVI metrics gauge durational variability, but they have different perspectives. The Δ measure captures durational variability globally (of the whole utterance) while the PVI measure captures durational variability locally (between successive units). Although it is likely that both metrics of the same interval will be highly correlated, it is theoretically possible to have an utterance scoring high in global variability but low in local variability (see Low et al., 2000). Therefore, both types of metrics were adopted in this study for a more comprehensive investigation. Previous studies have shown that raw metrics for vowel intervals are highly affected by speech rate so normalisation is necessary. However, normalisation for consonantal intervals is more problematic since it may eliminate language differences in syllable structure (Bunta & Ingram, 2007; Grabe & Low, 2002; White & Mattys, 2007). Therefore, both raw and normalised metrics for consonantal intervals were used, while only normalised metrics were adopted for vocalic and syllabic intervals. This results in nine rhythmic metrics: ΔC, VarcoC, rPVI-C, nPVI-C, VarcoV, nPVI-V, %V, VarcoS and nPVI-S. "Varco" is the normalised version of Δ metrics. For PVI, "r" refers to the raw version of the measure while "n" means normalised. These rhythmic metrics were first calculated for each utterance and then averaged across utterances for each child. Therefore, the data for each child were an averaged score based on all utterances.

3. Results

Since the speech materials were natural utterances, an important question is whether they are of comparable complexity. Table 3 shows the averaged number of syllables per utterance and the speech rate in terms of number of syllables per second (excluding pauses) for each child. Independent t-tests confirm that the comparisons between monolingual Cantonese and monolingual English, and between monolingual and bilingual Cantonese and English are all non-significant for both metrics. Although paired samples t-tests comparing the speech rate of the two languages of the bilingual children are significant [$t(5) = 2.968$, $p = .031$], the difference is very small: Cantonese (3.9 syllables/s) and English (3.7 syllables/s). Therefore, we can conclude that the speech materials are of comparable complexity, and speech rate should not be a major factor affecting the following results.

The mean values and standard deviations of all the rhythmic metrics in Cantonese and English of the monolingual and bilingual children can be found in the Online Supplementary Materials, available at the Journal's website. Just to recapitulate, a higher value of the eight rhythmic metrics (except %V) shows more durational variability, which indicates characteristics of stress-timing. Due to frequent vowel reduction and the preponderance of consonant clusters, stress-timed languages will have a lower %V than syllable-timed languages.

The first research question (Section 1.3 above) is whether monolingual children acquiring rhythmically different languages display distinct patterns at age three. Figure 1 shows the values of the rhythmic metrics for consonantal, vocalic and syllabic intervals. As expected, monolingual English has higher values for the eight rhythmic metrics and a lower %V than monolingual Cantonese, and the difference is consistently significant for both metrics of syllabic intervals (VarcoS [$t(10) = -2.534$, $p = .030$]; nPVI-S [$t(10) = -2.233$, $p = .050$]). There is also a non-significant difference for VarcoC [$t(10) = -2.007$, $p = .073$]. The consistent patterns of syllabic intervals suggest that monolingual English children have a higher durational variability in their production than monolingual Cantonese children, i.e. monolingual English is more stress-timed than monolingual Cantonese. The two groups of children display distinct rhythmic patterns at age three.

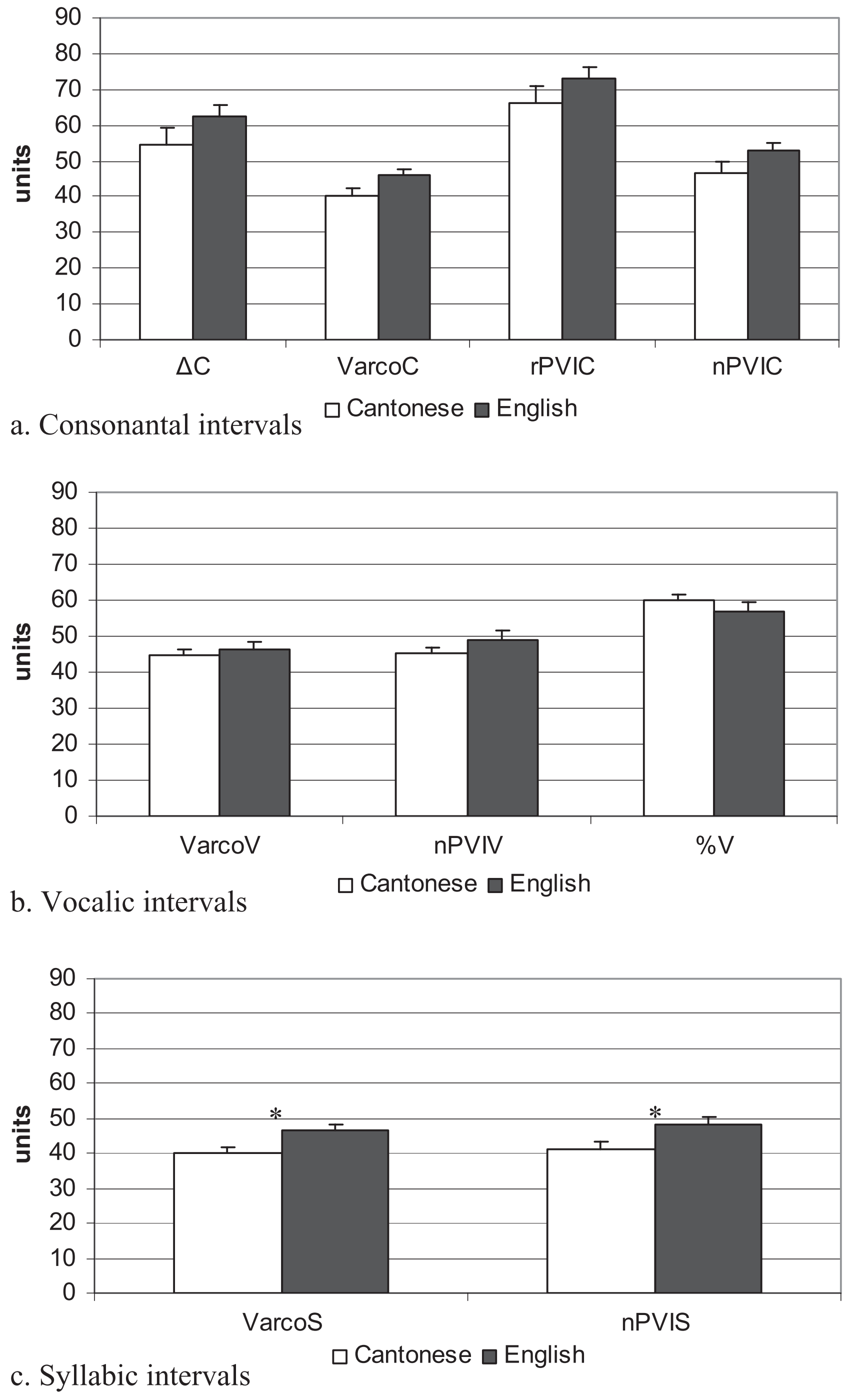
The second question is whether bilingual children have the same rhythmic patterns as their age-matched monolingual peers. We can compare how the bilingual and monolingual children differ in the two languages. The Cantonese comparisons show that there are only two significant differences between bilingual and monolingual Cantonese: %V [$t(10) = 2.278$, $p = .046$] and raw consonantal PVI [$t(10) = -2.305$, $p = .044$] but no consistent difference is evident in any interval. However, bilingual and monolingual English are consistently different for the two metrics of syllabic intervals (see Figure 2): VarcoS [$t(10) = -3.748$, $p = .004$]; nPVI-S [$t(10) = -3.112$, $p = .011$]. This consistent pattern in syllabic intervals between monolingual and bilingual English is particularly worth noting because it is similar to the differences between monolingual Cantonese and monolingual English observed above (see Figure 1).

Since the metrics on syllabic intervals are consistently significantly different for monolingual Cantonese and monolingual English, and for monolingual and bilingual English, we can see the direction of influence clearly if we plot the values of the metrics on syllabic intervals in a scatterplot. The Varco measure gauges global durational variability of the whole utterance while the PVI gauges local variability between successive units. We can see in Figure 3 that monolingual English is noticeably distinct from the other groups for both metrics, while bilingual English clusters with monolingual and bilingual Cantonese. We can safely conclude that, therefore, while monolingual and bilingual children are similar in their acquisition of syllable-timing (Cantonese), the acquisition of stress-timing (English) by bilingual children is strongly influenced by their other language. This conclusion also matches the auditory impression of their English quite well.[1]

It will be useful to compare the two languages of the bilingual children to confirm the patterns observed above. Figure 4 shows the values of the rhythmic metrics for consonantal, vocalic and syllabic intervals of the two languages for the bilingual children. In general, the bilingual patterns are similar to those of monolingual children as shown in Figure 1, but the differences between the two languages are attenuated, especially for vocalic and syllabic intervals. While the two rhythmic metrics with syllabic intervals are all significantly different for the monolingual children, the two metrics are all non-significant for the bilingual children ($p > .05$). Only raw consonantal PVI is significantly different [$t(5) = -3.678$, $p = .014$], while there is a non-significant difference for normalised consonantal PVI (nPVI-C [$t(5) = -2.386$, $p = .063$]) and ΔC [$t(5) = -2.402$, $p = .061$]. The results indicate that although bilingual children generally follow the monolingual patterns, the rhythm of their two languages are less distinct than their monolingual counterparts, echoing the patterns observed above.

Results of the quantitative metrics reveal a general picture of the rhythmic patterns of the target children. The metrics on syllabic intervals show consistent differences

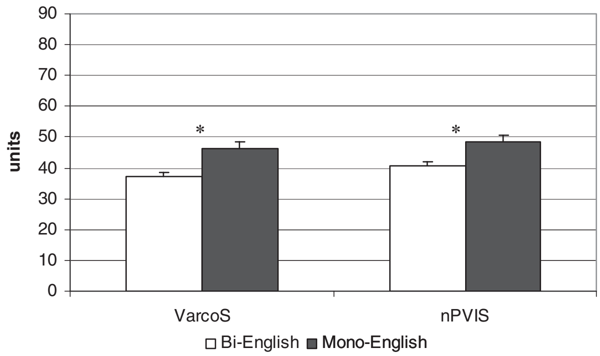
[1] Recordings of the bilingual children are available online at http://childes.psy.cmu.edu/media/Biling/YipMatthews/. Interested readers can listen to the rhythm of their English utterances.



* = p < .05. The y-axis is relative to each rhythmic metric.

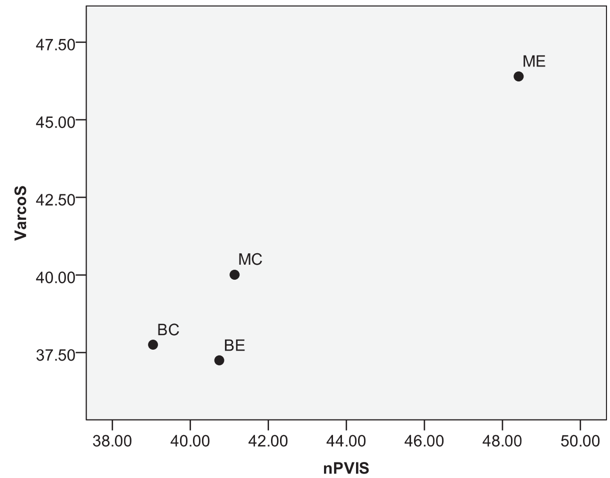
Figure 1. Rhythmic metrics of (a) consonantal, (b) vocalic and (c) syllabic intervals of monolingual Cantonese and monolingual English children.

between the languages. Difference in syllable structure is an important feature contributing to the impression of different speech rhythm. In order to further investigate the phonological difference between monolingual and bilingual children, a detailed inspection of the structure of all syllables produced by three children was conducted to confirm the observed patterns (one monolingual Cantonese (DGC), one monolingual English (Jessica) and one bilingual (Sophie)). They are typical examples of the children they represent. All syllables produced by them were counted and checked auditorily. Each syllable was transcribed for its canonical syllable structure and



* = p < .05. The y-axis is relative to each rhythmic metric.

Figure 2. Mean values of two rhythmic metrics on syllabic intervals using data from bilingual English and monolingual English.



B = bilingual; M = monolingual; C = Cantonese; E = English

Figure 3. Scatterplot of syllabic intervals using VarcoS and nPVI-S.

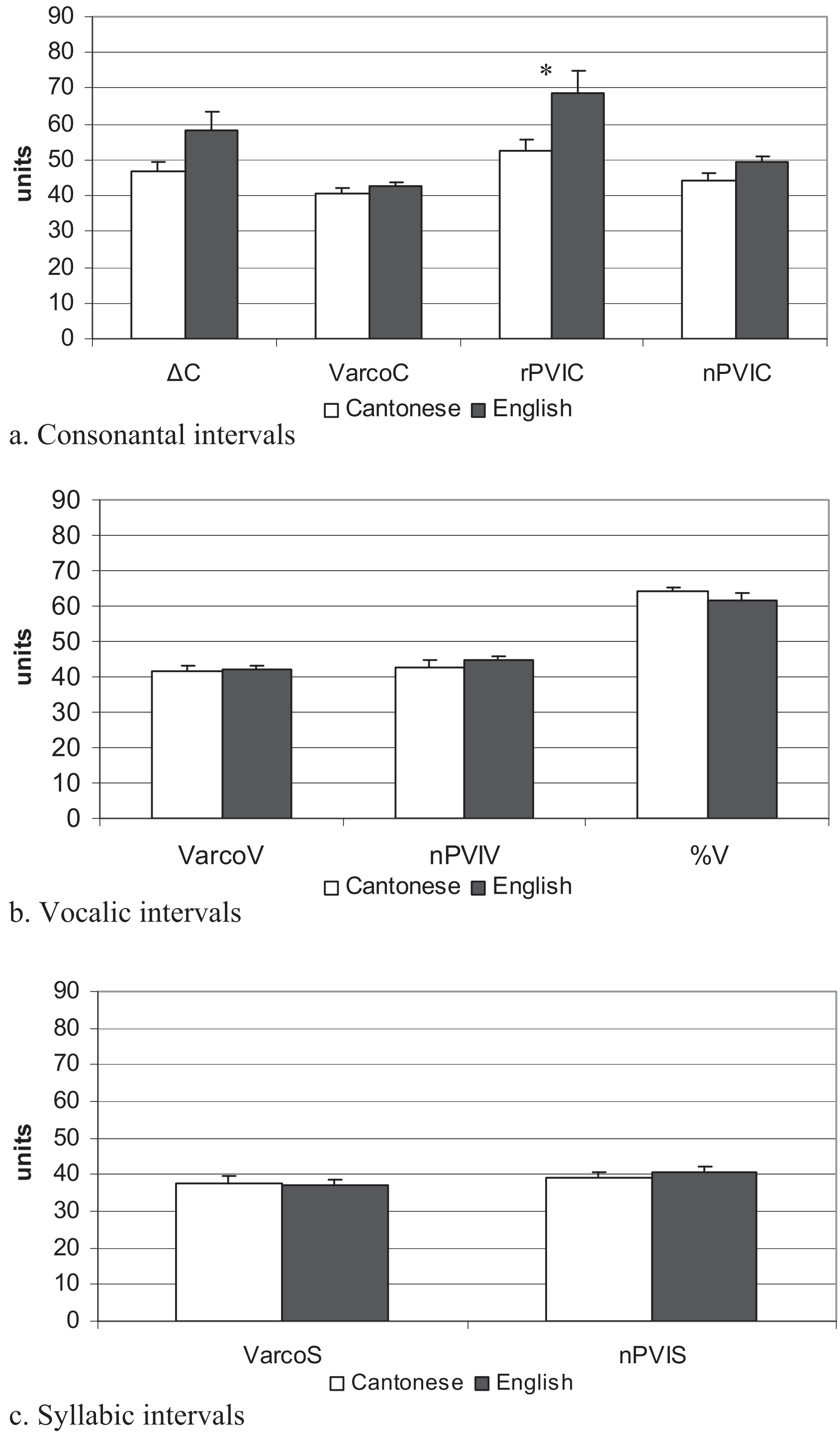
how it was actually produced by the children in the recording. Table 4 shows the frequency of syllables (%) in various structures in both canonical form and in actual realisation. A number of observations can be made. First, majority of the syllables in both canonical form and actual realisation are of simple structure (V, CV, CVC) in both Cantonese and English for both monolingual and bilingual children. Second, CV is the most frequent syllable type for Cantonese (monolingual 73%, bilingual 63%), followed by CVC (monolingual 18%, bilingual 16%), while for English, both CV and CVC occur most frequently (CV: monolingual 39%, bilingual 52%; CVC: monolingual 32%, bilingual 24%). Third, syllables with complex structure (involving consonant clusters) in English are dramatically reduced in actual realisation, and the difference is more pronounced in bilingual than monolingual English. There is also a clear tendency towards more CV (52%) than CVC (24%) structures in bilingual English while there is a more equivalent use of CV (39%) and CVC (32%) structures in monolingual English. Finally, the total frequency of syllables having simple structure (V, CV, CVC) in actual realisation is very similar in monolingual (98%) and bilingual (97%) Cantonese, while there is a slightly larger difference between monolingual (81%) and bilingual (86%) English. Taken together, these observations suggest that the bilingual child's Cantonese is comparable to the monolingual counterpart, but her English has simpler syllable structure than the monolingual child. The qualitative data concur with and illustrate the patterns obtained using the rhythmic metrics well.

In addition to syllable structure complexity, vowel reduction is another important feature contributing to speech rhythm. Data on vowel quality can help us verify the conclusions drawn above based on the rhythmic metrics and syllable structure count. Formant frequency data (F1 and F2) of the vowel in the English word *this* (/ɪ/) in non-focus-bearing positions were taken from each bilingual and monolingual English children. This vowel was chosen for several reasons. First and foremost, it is important to use the same word across all children for a fair comparison. Since the materials were natural utterances, only very few words were produced by all children in the recordings. The word *this* is one of them. Second, the vowel in *this* is both preceded and followed a consonant. This can ensure that the formant measurements will not be affected by the uncontrolled adjacent contexts. Third, the word *this* often appears in non-focus-bearing positions, which allows us to investigate vowel reduction more directly. F1 and F2 frequencies of the vowel in the word *this* were taken at around the vowel midpoint. The frequency data were based on three tokens of the word *this* spoken by each child in different utterances, except for one bilingual (Charlotte) and two monolingual (Ella and Molly) children because not enough usable tokens are available. The formant frequency data of these three children were based on one token each only.

Figure 5 shows the formant frequency data of the bilingual and monolingual children. Each data point represents data from one child. It is obvious that data from the monolingual children cluster tightly together while data from the bilingual children are more scattered in both F1 and F2 dimensions. The standard deviations of the bilingual children are much larger than the monolingual children. The data show that the bilingual children produced the vowel /ɪ/ with more variable quality and less reduction. The limited data on formant frequency agree with the conclusion based on rhythmic metrics very well.

4. Discussion

The results indicate that bilingual and monolingual children differ in their prosodic development of speech



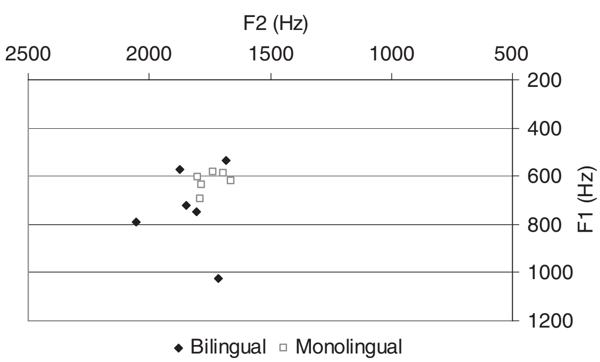
* = p < .05. The y-axis is relative to each rhythmic metric.

Figure 4. Rhythmic metrics of (a) consonantal, (b) vocalic and (c) syllabic intervals in bilingual Cantonese and bilingual English.

rhythm as reflected by the durational rhythmic metrics. While monolingual children acquiring Cantonese and English already display distinct rhythmic patterns at age three, the differences between the two languages of the bilingual children are less distinct. These results could arguably be interpreted as supporting a unitary undifferentiated phonological system for the bilingual children, but such a conclusion is premature for two reasons. First, syllable-timing is prevalent in early phonology regardless of the target languages. Features

Table 4. *Occurrence (%) of different syllable structures in actual realisation (and canonical form) produced by three children (one monolingual Cantonese, one monolingual English, and one bilingual Cantonese–English).*

	Monolingual Cantonese	Monolingual English	Bilingual Cantonese	Bilingual English
Total number of syllables	191 (194)	170 (173)	178 (182)	131 (132)
Syllable structure	%	%	%	%
V	7 (3)	10 (9)	18 (7)	10 (7)
CV	73 (73)	39 (30)	63 (72)	52 (47)
CVC	18 (22)	32 (37)	16 (18)	24 (24)
VC	0 (0)	9 (10)	1 (0)	11 (8)
C	2 (2)	3 (0)	2 (3)	0 (0)
CCV		1 (1)		2 (3)
CCVC		3 (3)		1 (3)
VCC		0 (3)		0 (3)
CVCC		1 (6)		0 (5)



Bilingual mean (and standard deviation) – F1: 732 (176); F2: 1830 (132)
Monolingual mean (and standard deviation) – F1: 621 (40); F2: 1744 (57)

Figure 5. Formant frequency data of the vowel in the word *this* (/ɪ/) produced by bilingual and monolingual children.

of stress-timing are acquired later (Allen & Hawkins, 1980; Grabe et al., 1999). It is difficult, therefore, to tell whether there is a unitary bilingual system or whether the bilingual system is simply at an earlier developmental stage. Kehoe, Lleó and Rakow (2004) and Paradis (2001) have pointed out that great care must be taken when interpreting developmental results such as these. Second, a more detailed inspection of the data also suggests otherwise. Although there were no consistent significant differences between the two languages of the bilingual children, the rhythmic metrics indicate that their English generally has a higher durational variability than their Cantonese, following the same patterns of the monolingual children. Bilingual English is less stress-timed than monolingual English, but it is nevertheless developing in the expected direction. This suggests that bilingual English and Cantonese do not simply form one undifferentiated unitary system. Instead, the results seem more consistent with the view that bilingual children have a distinct phonological developmental trajectory from monolingual children, and bilingual children are not simply two monolinguals in one (Grosjean, 1989; Vihman, 1996). The subtle differences in the rhythmic patterns of the bilingual children illustrates that the two phonological systems of bilingual children do not develop independently but interact.

How might the two phonological systems of the Cantonese–English bilingual children interact? One obvious manifestation is acquisition delay. The bilingual English data follow the patterns of monolingual English but with less durational variability, but it is developing in the expected direction. Lleó et al. (2007) found similar results at age three that the rhythmic patterns of the two languages of German–Spanish bilingual children are similar while they are different for monolingual children. The results in Bunta and Ingram (2007) also demonstrate that Spanish–English bilingual children show a bias toward syllable-timing, although the children can also distinguish the two rhythmic patterns in their production at around age four. As mentioned above, a less variable timing pattern, i.e. pertaining to syllable-timing, is easier to master than a more variable pattern, i.e. stress-timing. Monolingual English and German children can develop a more variable pattern earlier than bilingual children learning the same languages. Similar delay in bilingual acquisition of other marked phonological features is also reported by Lleó (2002) and Kehoe (2002). Since the Cantonese–English bilingual children in this study also exhibit many examples of delay in syntactic acquisition (Yip & Matthews, 2007), their less distinct rhythmic patterns at age three can also be regarded as a form of delay in phonological acquisition. Given the above findings, we may conclude that in general, monolingual children can

 https://doi.org/10.1017/S1366728910000453

distinguish different rhythmic patterns at age three while bilingual children may only distinguish them at around age four, i.e. there is a delay of approximately one year.

The time course of delay probably depends on the nature of the phonological differences between the language pairs. The delay of rhythmic development of bilingual children may also be shorter than one year. Unfortunately, thus far there is no study investigating bilingual acquisition of speech rhythm between age three and four. Moreover, the recordings of most of the Cantonese–English bilingual children in this study do not go beyond 3;4. It is difficult, therefore, to assess a more precise age of divergence of rhythmic patterns for the bilingual children. This difficulty notwithstanding, there are related questions regarding early prosodic development that can be addressed. Do monolingual children distinguish speech rhythm before age three, if syllable-timing is prevalent in early childhood? And how do their rhythmic patterns compare with bilingual children? Lleó and Kehoe (2002) pointed out that it is important to access bilingual phonological acquisition during the second year of life, especially in the area of prosody, when syntax is still very limited. Paradis (2001) showed that both monolingual and bilingual two-and-a-half-year-olds acquiring French and English display biases in word truncation according to language-specific stress patterns. Her findings are particularly relevant to our questions about speech rhythm development, since stress is an important feature of speech rhythm. French, like Cantonese, is also a syllable-timed language. If two-and-a-half-year-old children are sensitive to language-specific stress patterns in word truncation, it is possible that they may show sensitivity to different rhythms too. In order to address these issues, additional data at age 2;6 from five of the Cantonese–English bilingual children and their monolingual peers are being analysed. Speech data earlier than 2;6 are not considered because they are usually quite short, and thus unsuitable for calculating durational variability. Impressionistic observation suggests that monolingual English at 2;6 sounds more regular and is closer to syllable-timing. If quantitative data using rhythmic metrics confirm the impressionistic observation, then we may be able to pinpoint the age of rhythm differentiation in monolingual children. Such information can complement more detailed bilingual phonological comparisons about syllable structure and vowel reduction patterns.

Besides delay, language dominance and cross-linguistic transfer may also affect the different developmental patterns of bilingual and monolingual speech rhythm. Figure 3 shows clearly that bilingual English does not simply lie somewhere between monolingual Cantonese and monolingual English, as what one might expect if rhythmic development of bilingual English is only delayed. Bilingual English clusters with monolingual and bilingual Cantonese, indicating that Cantonese exerts a much stronger influence than English. Five of the six bilingual children in this study are Cantonese-dominant (see Table 2 above). Their language dominance was measured objectively using mean length of utterance (MLU) and MLU differentials (the difference between MLU values for a child's two languages at a given time point), and ample evidence of Cantonese influencing their English syntactically can be found (see Yip & Matthews, 2007 for detail). The observed pattern in their speech rhythm is what is predicted by their language dominance. Nevertheless, since most of the bilingual children in this study are Cantonese-dominant, it is impossible to distinguish the separate effects of delay and language dominance based on the present data only. It is quite likely that both factors contribute to the observed pattern. A detailed case study of another, balanced Cantonese–English bilingual child at age three (not included in the present study) using both quantitative and qualitative data reveal that language dominance does affect bilingual speech rhythm development (White, 2009; White & Mok, 2008). Compared with the monolingual peers, the balanced bilingual child also shows a delay in her rhythm development, but her two languages generally fall between the monolingual peers with less bias towards Cantonese. Kehoe et al. (2004) also found that the dominant language tends to influence the other language in early bilingual phonological acquisition of voice onset time. These results suggest that there is a close relationship between language dominance and the direction of cross-linguistic influence. Of course, more comparisons with more bilingual children being dominant in both languages and being balanced are needed to further explore the influence of language dominance, but the present data are consistent with the prediction based on language dominance.

Similar effects of Cantonese (syllable-timing) dominance in early bilingual acquisition can be found in adult second language acquisition as well. Various studies of the speech rhythm of adult second language learners who speak a syllable-timed first language show that the rhythmic pattern of their second language, notably English, is often less stress-timed than monolingual English speakers even if they are competent speakers in the second language (e.g. Gibbon & Gut, 2001; Jeon, 2006; Jian, 2004; Low et al., 2000; Mok & Dellwo, 2008; Setter, 2006). Plenty of evidence in early and adult bilingual acquisition points to the conclusion that it is usually the rhythm of the stress-timed language that is being affected. The cross-linguistic influence is largely asymmetrical. To the best knowledge of the present author, there is no study clearly demonstrating stronger influence of the stress-timed language on the syllable-timed one (but see White and Mattys (2007), who found that VarcoV scores for Spanish and English of Spanish–English bilingual speakers are both intermediate between monolingual realisations of the two languages). Perhaps syllable-timing is not only linguistically less marked and easier to acquire for children, but also exerts stronger influence when adults acquire another language.

Previous studies show that rhythmic metrics on vocalic intervals are more robust than those on consonantal intervals (e.g. Bunta & Ingram, 2007; White & Mattys, 2007). The present results show that while the differences in rhythmic metrics on consonantal and vocalic intervals between the two languages follow the expected patterns, only the metrics on syllabic intervals show consistent significant differences. One may question why the variability of syllable duration shows the most robust language distinction, and what lies behind such distinction. There are a number of possible answers. First, the qualitative data in the "Results" section (Section 3) above show clearly that syllables with consonant clusters are rare in children's production of both monolingual and bilingual English, and the children also simplified the clusters in actual realisation. The most common syllable forms are V, CV and CVC in both Cantonese and English (although CV is the most frequent in Cantonese while there is a more even division between CV and CVC in English). The actual difference of consonant duration between the two languages is therefore much smaller than what is predicted based on adult phonology in which there is a much larger difference between the syllable structures of the two languages.

In addition, segmentation of consonantal and vocalic intervals ignores syllable boundaries. Abutting vowels (and consonants) were considered as one large interval. This resulted in long stretches of consonantal and vocalic intervals which do not correspond to any linguistic unit in production. In addition, vowel reduction in English may not always be effectively reflected in vocalic intervals because of abutting vocalic intervals, and also because initial glides were considered vocalic if they could not be clearly separated in the spectrograms, which is not unusual. Therefore, variability of vocalic and consonantal intervals is not equivalent, or directly translatable, to the patterns in syllable duration. In contrast, both syllable structure complexity and vowel reduction can be captured by syllabic intervals. These issues may explain why the phonological differences between Cantonese and English are not as robustly reflected in consonantal and vocalic intervals as in syllabic intervals. Various studies demonstrated that the syllable is an important unit in child phonology, and it is also perceptually more salient to young children than segments (e.g. Liberman, Shankweiler, Fischer & Carter, 1974; Treiman, 1985; Treiman & Zukowski, 1996). We can assume that the syllable has an important role in their production too. Francis Nolan, who initiated the use of PVI for rhythm research (Low et al., 2000), argued that despite the difficulties of defining syllable boundaries, the syllable is an important phonological unit which should not be ignored in calibrating speech rhythm in favour of units more convenient for acoustic analysis (Nolan & Asu, 2009). This suggestion and the present results concur with other studies showing that variability of syllable duration can also reliably distinguish languages with different rhythmic patterns (Deterding, 2001; Mok & Dellwo, 2008). Because of the phonological modifications in actual realisation and segmentation issues mentioned above, one could also argue that variability of syllable duration is more suitable than variability of consonant and vowel duration in studies using rhythmic metrics with young children. Clearly, more studies comparing these metrics on consonantal, vocalic and syllabic intervals with data on syllable structure complexity and vowel reduction are needed to confirm this suggestion.

Incorporating all of the above discussion, since the results indicate that there is a delay in the development of English rhythm of the bilingual children, one may question which phonological features are being delayed. Syllable structure, stress and vowel reduction are three important features contributing to differences in speech rhythm (Dauer, 1983). The qualitative data show that syllable structure in bilingual English is indeed simpler than in monolingual English, but the difference is not too extensive. Also, the Cantonese and English syllables are of similar structure. It is, therefore, quite possible that stress and vowel reduction contribute more to the rhythmic delay. The results show that the English of the bilingual children is strongly influenced by their Cantonese because of language dominance. In Cantonese, there is no lexical stress and no phonological vowel reduction. These phonological features of Cantonese concur well with the delayed English rhythm of the bilingual children. Monolingual English scores higher than bilingual English in VarcoV and nPVI-V, and also has a lower %V (see Online Supplementary Materials), although the differences are not significant. Due to the segmentation issues mentioned above, vowel reduction may not always be effectively reflected in the metrics of vocalic intervals. Nevertheless, their patterns do agree with expectation. The limited data on formant frequency also confirm that the bilingual children produced the vowel in the word *this* with more variable quality and less reduction. Since both syllable structure complexity and vowel reduction can be more effectively captured by syllabic intervals, the consistent syllabic differences between monolingual and bilingual English, and between monolingual Cantonese and monolingual English strongly suggest that the English of the bilingual children is delayed in terms of vowel reduction and stress difference. Nonetheless, more controlled data on vowel reduction are needed to verify this conclusion, as the present data are all based on natural utterances.

Although bilingual Cantonese and English generally do not differ in terms of rhythmic metrics, bilingual Cantonese (3.9 syllables/s) has a significantly faster speech rate than bilingual English (3.7 syllables/s) albeit with only a small difference. This could be due to the difference in syllable structure. There is a higher percentage of syllables with simpler structures (V and CV) in bilingual Cantonese than bilingual English, which may result in more syllables per second in bilingual Cantonese. In addition, most of the bilingual children were Cantonese-dominant and they were living in a predominantly Cantonese environment, thus they probably would find it easier to speak in Cantonese than English. The slightly higher speech rate in bilingual Cantonese matches their dominance in Cantonese well. These two factors can explain the small but significant difference in speech rate between the two languages of the bilingual children.

Finally, there are some subtle but statistically significant differences between monolingual and bilingual Cantonese. Bilingual Cantonese has a higher %V and a lower raw consonantal PVI than monolingual Cantonese. The qualitative data on syllable structure frequency can help explain this pattern. Table 4 above shows that bilingual Cantonese has a higher percentage of V syllables than monolingual Cantonese, and a lower percentage of CV syllables than monolingual Cantonese. This results in more vocalic portion and less consonantal portion in the speech signal, which is reflected by the significant differences in %V and raw consonantal PVI. This could mean that, in addition to English, the bilingual children were also displaying a tendency towards simpler syllable structures in Cantonese. Developmentally, the bilingual children were capable of producing a range of syllable structures and consonants, but they had a tendency towards simplification in both languages. This possibility echoes the discussion of acquisition delay above.

Most of the research on early bilingual phonological acquisition focuses on segmental properties. The present study provides new data using recently developed acoustic metrics on bilingual acquisition of speech rhythm in a language pair that is less well-understood in terms of bilingual phonological development. The data confirm that the two languages of bilingual children are not autonomous and they interact. The interaction between the two languages is manifested in acquisition delay of stress-timing by the bilingual children compared to their monolingual counterparts. Language dominance also results in an asymmetrical influence of prosodic development. The findings on this under-researched prosodic aspect provide useful insights into early bilingual phonological acquisition. More studies on the phonological development of Cantonese–English bilingual children are needed as the present study on speech rhythm only depicts a general picture of their phonological development. Detailed inspection of their acquisition of phonological properties like consonant clusters and vowel reduction can further strengthen the conclusions drawn here. Future studies on various language pairs and at different ages can further advance our understanding of prosodic development of both monolingual and bilingual children.

References

Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh: Edinburgh University Press.

Allen, G., & Hawkins, S. (1980). Phonological rhythm: Definition and development. In G. Yeni-Komshian, J. Kavanagh & C. Ferguson (eds.), *Child phonology* (vol. 1): *Production*, pp. 227–256. New York: Academic Press.

Barry, W. J., Andreeva, B., Russo, M., Dimitrova, S., & Kostadinova, T. (2003). Do rhythm measures tell us anything about language type? In Solé et al. (eds.), pp. 2693–2696.

Bauer, R. S., & Benedict, P. K. (1997). *Modern Cantonese phonology*. Berlin: Mouton de Gruyter.

Boersma, P., & Weenink, D. (2009). *Praat: Doing phonetics by computer* (Version 5.1.12) [computer program]. http://www.praat.org/Boula de.

Bosch, L., & Sebastián-Gallés, N. (2001). Early language differentiation in bilingual infants. In J. Cenoz & F. Genesee (eds.), *Trends in bilingual acquisition*, pp. 71–93. Amsterdam: John Benjamins.

Bunta, F., & Ingram, D. (2007). The acquisition of speech rhythm by bilingual Spanish- and English-speaking 4- and 5-year-old children. *Journal of Speech, Language and Hearing Research, 50 (4),* 999–1014.

Chao, Y. R. (1947). *Cantonese primer*. New York: Greenwood Press.

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics, 11,* 51–62.

Dellwo, V. (2006). Rhythm and speech rate: A variation coefficient for ΔC. In P. Karnowski & I. Szigeti (eds.), *Language and language-processing*, pp. 231–241. Frankfurt am Main: Peter Lang.

Dellwo, V., & Wagner, P. (2003). Relations between language rhythm and speech rate. In Solé et al. (eds.), pp. 471–474.

Deterding, D. (2001). The measurement of rhythm: A comparison of Singapore and British English. *Journal of Phonetics, 29,* 217–230.

Fletcher, P., Leung, S. C. S., Stokes, S. F., & Weizman, Z. O. (2000). *Cantonese preschool language development: A guide*. Hong Kong: Department of Speech and Hearing Sciences, University of Hong Kong.

Forrester, M. (2002). Appropriating cultural conceptions of childhood: Participation in conversation. *Childhood, 9,* 255–278.

Gibbon, D., & Gut, U. (2001). Measuring speech rhythm. In P. Dalsgaard, B. Lindberg, H. Benner & Z. Tan (eds.), *Proceedings of Eurospeech 2001 Scandinavia*, pp. 95–98. http://www.isca-speech.org/archive/archive_papers/eurospeech_2001/e01_0095.pdf.

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven & N. Warner (eds.), *Laboratory phonology VII*, pp. 515–546. Berlin: Mouton de Gruyter.

Grabe, E., Post, B., & Watson, I. (1999). The acquisition of rhythmic patterns in English and French. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville & A. C. Bailey (eds.), *Proceedings of the 14th International Congress of Phonetic Sciences (ICPhS),* pp. 1201–1204.

Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language, 36,* 3–15.

Hashimoto, A. (1972). *Cantonese phonology*. Cambridge: Cambridge University Press.

Holm, A., & Dodd, B. (1999). A longitudinal study of phonological development of two Cantonese–English bilingual children. *Applied Psycholinguistics, 20,* 349–376.

Jeon, H. S. (2006). Acoustic measure of speech rhythm: Korean learners of English. MSc thesis, University of Edinburgh.

Jian, H. L. (2004). On the syllable timing in Taiwan English. In *Proceedings of Speech Prosody 2004,* pp. 247–250.

Kao, D. (1971). *Structure of the syllable in Cantonese*. The Hague: Mouton.

Kehoe, M. (2002). Developing vowel systems as a window to bilingual phonology. *The International Journal of Bilingualism, 6,* 315–334.

Kehoe, M., Lleó, C., & Rakow, M. (2004). Voice onset time in bilingual German–Spanish children. *Bilingualism: Language and Cognition, 7,* 71–88.

Ladefoged, P. (1975). *A course in phonetics*. New York: Harcourt Brace Jovanovich.

Liberman, I. Y., Shankweiler, D., Fischer, F. W., & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young children. *Journal of Experimental Child Psychology, 18,* 201–212.

Lleó, C. (2002). The role of markedness in the acquisition of complex prosodic structures by German–Spanish bilinguals. *The International Journal of Bilingualism, 6,* 291–313.

Lleó, C., & Kehoe, M. (2002). On the interaction of phonological systems in child bilingual acquisition. *The International Journal of Bilingualism, 6,* 233–237.

Lleó, C., Rakow, M., & Kehoe, M. (2007). Acquiring rhythmically different languages in a bilingual context. In J. Trouvain & W. J. Barry (eds.), *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS)*, pp. 1545–1548.

Low, E. L., Grabe, E., & Nolan, F. (2000). Quantitative characterisations of speech rhythm: Syllable-timing in Singapore English. *Language and Speech, 43 (4),* 377–401.

Mok, P. P. K. (2009). On the syllable-timing of Cantonese and Beijing Mandarin. *Chinese Journal of Phonetics, 2,* 148–154.

Mok, P. P. K., & Dellwo, V. (2008). Comparing native and non-native speech rhythm using acoustic rhythmic measures: Cantonese, Beijing Mandarin and English. In P. A. Barbosa, S. Madureira & C. Reis (eds.), *Proceedings of Speech Prosody 2008*, pp. 423–426.

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Towards an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 756–766.

Nazzi, T., Jusczyk, P. W., & Johnson, E. K. (2000). Language discrimination by English-learning 5-month-olds: Effects of rhythm and familiarity. *Journal of Memory and Language, 43,* 1–19.

Nolan, F., & Asu, E. L. (2009). The pairwise variability index and coexisting rhythms in language. *Phonetica, 66,* 64–77.

Paradis, J. (2001). Do bilingual two-year-olds have separate phonological systems? *The International Journal of Bilingualism, 5,* 19–38.

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America, 32,* 693–703.

Pike, K. L. (1945). *The intonation of American English*. Ann Arbor, MI: University of Michigan Press.

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition, 73,* 265–292.

Roach, P. (1982). On the distinction between stress-timed languages and syllable-timed languages. In D. Crystal (ed.), *Linguistic controversies: Essays in honour of F. R. Palmer*, pp. 73–79. London: Arnold.

Setter, J. (2006). Speech rhythm in world Englishes: The case of Hong Kong. *TESOL Quarterly, 40 (4),* 763–782.

Solé, M. J., Recasens, D., & Romero, J. (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS)*.

Treiman, R. (1985). Onsets and rimes as units of spoken syllables: Evidence from children. *Journal of Experimental Child Psychology, 39,* 161–181.

Treiman, R., & Zukowski, A. (1996). Children's sensitivity to syllables, onsets, rimes and phonemes. *Journal of Experimental Child Psychology, 61,* 193–215.

Vihman, M. (1996). *Phonological development*. Oxford: Blackwell.

Warner, N., & Arai, T. (2001). Japanese mora-timing: A review. *Phonetica, 58,* 1–25.

White, D. (2009). *Speech rhythm development in a balanced Cantonese–English bilingual child*. MA thesis, Chinese University of Hong Kong.

White, D., & Mok, P. P. K. (2008). On the development of speech rhythm in a Cantonese–English bilingual child: A case study. Presented at the Conference on Bilingual Acquisition in Early Childhood, Hong Kong.

White, L., & Mattys, S. L. (2007). Calibrating rhythm: First language and second language studies. *Journal of Phonetics, 35,* 501–522.

Whitworth, N. (2002). Speech rhythm production in three German–English bilingual families. *Leeds Working Papers in Linguistics and Phonetics, 9,* 175–205.

Wiget, L., White, L., Schuppler, B., Grenon, I., Rauch, O., & Mattys, S. L. (2010). How stable are acoustic metrics of contrastive speech rhythm? *Journal of the Acoustical Society of America, 127,* 1559–1569.

Yip, V., & Matthews, S. (2007). *The bilingual child: Early development and language contact*. Cambridge: Cambridge University Press.

Yuan, J. H. (1960). *Hanyu Fangyan Gaiyao* [A survey of Chinese dialects]. Beijing: Wenzi Gaige Chubanshe.