SPEECH RHYTHM IN ENGLISH AND ITALIAN: AN EXPERIMENTAL STUDY ON EARLY SEQUENTIAL BILINGUALISM

VINCENZO VERBENI
University of Padova, Padova, Italy
vincenzoverbeni@gmail.com

Abstract
The study investigates the dynamics of speech rhythm in early sequential bilingual children who have access to Italian-English immersion programs. The research focused on the Italian and English semi-spontaneous narrative productions of 9 students, aged between 6;7 and 10;11 and distributed across three different classes (Year 1, Year 3, Year 5). Their speech was recorded and subject to an interval-based analysis via computation of %V/ΔC, PVI and Varco metrics. The retrieved metrics underwent within-group and between-group one-way ANOVAs in order to identify valuable cross-linguistic variations among children of the same age and statistically significant differences between different age groups (Y1, Y3, Y5). The results appear to support a stress-centered interpretation of speech rhythm: according to this view, all languages could be arranged on a stress-timed continuum in which "syllable-timing" is marked by sparser occurrences of (regular) prominence due to the relative absence of vocalic elision and consonantal complexity. Indeed, the comparative analysis drawn between the normalized vocalic indexes of Y1, Y3 and Y5 students revealed a statistically relevant increase in vocalic variation phenomena both in Italian and in English. Moreover, Y1 and Y3 consonantal scores were comparatively higher in the Italian sample: it will be discussed how unpredictable stress-timed patterns can arise as a function of proficiency, speech-rate and age-related disfluencies.

Keywords: speech rhythm, language acquisition, sequential bilingualism, immersion education; interval-based metrics, speech segmentation

1. Introduction

The highly debated *isochrony* hypothesis – postulated by Pike (1945) and partially reformulated by Abercrombie (1967) – frames language rhythm as the output of the distribution and temporal duration of speech segments. Notably, Mehler and colleagues (1988) envisioned the phenomenon as a *class*-specific pattern derived from the alternation of vocalic and consonantal intervals. According to their view, a distinction could be drawn between *syllable*-timed languages (e.g. Italian, Spanish and French) – which exhibit syllables of equal duration – and *stress*-timed languages (e.g. English, Dutch and German) – which are characterized by regular inter-stress intervals. Since vowels represent the nuclei of syllables, languages like Italian exhibit low variability in the durational length of vocalic intervals. Conversely, stress-timing is associated to vowel reduction phenomena, which translate into highly variable vocalic units. Most importantly, the authors claimed that children could only discriminate between languages that belong to different rhythmic classes due to their innate sensitivity to such distributional patterns.

From an acquisitional perspective, the concept of “rhythmic class” has been challenged by various researchers. For instance, White *et al.* (2016) proved that 5-month-old English infants could actually discriminate between French and Spanish (two syllable-timed languages) but not between French and Finnish: in other words, their *perception* of language rhythm did not appear to be categorical. Moreover, Konopczynski (1995) performed a cross-linguistic study on syllabic durations and established that children’s *productions* tend to be syllable-timed in the first stages of acquisition. According to her results, conformity to the native tongue’s pattern is attained depending on the rhythmic complexity of the target language: while French and Hungarian rhythms appeared to be acquired between 2;0 and 2;5 years of age, adult-like English patterns started to arise only after 3 years of age.

Vihman *et al.* (2006) proposed that the production of stress-timing is inherently more complex – relative to syllable-timing – due to the occurrence of phonotactically challenging clusters. Payne *et al.* (2012) validated this hypothesis in a study on English, Spanish and Catalan children of 2, 4 and 6 years of age: while early rhythm displayed a higher percentage of steady vocalic intervals, consonantal outputs appeared to be particularly inconsistent – even at age 6. While the modulation of vocalic intervals could result from the progressive growth of phonological knowledge, the durational variability of consonantal units was attributed to an insufficient development of the articulatory control system.

Finally, Grabe *et al.* (2001) highlighted that stress-timed languages can display different developmental trends: while English children did not seem to master the suprasegmental features of their language at age 4, the speech rhythm of 4-year-old Germans displayed the same structure of adult speech. Therefore, it was proposed that the acquisition rate of rhythmicity mainly depends on the language-specific level of phonological complexity. Besides, these results back up the idea that languages are actually distributed on a *continuum* of rhythmic possibilities rather than arranged within distinct classes.

Given such premises, the purpose of the present research is to better define the structural features of rhythmicity in child speech. To do so, the study investigates a specific instance of multilingualism – i.e. early *sequential* bilingualism – and makes use of three different sets of interval-based rhythm metrics: %V/ΔC, Varcos and PVIs.

2. Theoretical Background

2.1. Speech rhythm and multilingualism

Most of the studies which address the interaction between speech rhythm and multilingualism focus on data collected from simultaneous bilinguals and adult second language learners.

For instance, it was shown that English learners whose native tongue is marked by low levels of vocalic variability manage to develop stress-timed speech in a gradual and progressive fashion (Li and Post, 2014).

Research on simultaneous bilingual children shows that they follow a developmental trend which is very similar to the monolinguals' stages of rhythm acquisition: productions are initially marked by low levels of vocalic variability; then, divergent paths emerge for the acquisition of language-specific temporal templates. However, some studies have found that young simultaneous bilinguals exhibit similar rhythmic patterns in both their native tongues. For instance, Whitworth (2000) analyzed the productions of two German-English bilingual children, noticing that in both languages the ratio between tense and lax vowels corresponded to an intermediate value between the standard ratios identified for the two languages. This *compromise* effect was also found by Kehoe *et al.* (2011), who examined 3-year-old Spanish-German bilinguals: when compared to the speech rhythm of monolingual control groups, bilingual speech showed less vocalic variability in German and greater consonantal variability in Spanish. Other studies suggest that differentiation between two language rhythms in a bilingual context actually occurs, but later than the monolinguals' offset of prosody acquisition. For instance, Mok (2013) showed that the rhythmic patterns of monolingual Cantonese and English children are different at age 2;6, while there is less contrast between the language-specific rhythms of Cantonese-English bilinguals of the same age.

Limited research has been conducted on *early sequential* bilinguals. Indeed, most studies have focused on *late* bilingualism – with subjects ranging in age from their early teenage years up to their thirties (e.g. Carter, 2005). Notably, their speech rhythm has been shown to be characterized by the same "compromise values" that mark simultaneous bilingualism in its early stages (Howell and Borsel, 2011). However, given the great sensitivity that young children display with respect to prosodic cues, it is possible that sequential bilinguals manage to develop native-like rhythms when exposed to their L2 in the *early* stages of their life – i.e. right after 3;0, the age limit established by Meisel (2011) to distinguish between simultaneous and sequential bilingualism. Therefore, one of the aims of the present study is to investigate whether "rhythmic compromise" phenomena are present or not in the speech of early sequential bilingual children.

2.2. Rhythm metrics

Speech rhythm is modulated by a great variety of language-specific factors – e.g. the amount of phonological oppositions, complex consonantal clusters and vowel lengthening phenomena. Within the conceptual framework of isochrony-based hypotheses, these factors are believed to drive the perception of stress-timing and syllable-timing by regulating the duration of vocalic (V) and consonantal (C) intervals: V and C represent the units of measurement of the so-called "*rhythm metrics*". Three specific sets of metrics were selected for the purpose of the present study: % V/ΔC (Ramus *et al.*, 1999), Varcos (Dellwo & Wagner, 2003) and PVIs (Grabe & Low, 2002) – i.e. the three most popular tools that have been employed by research on speech rhythm for the last two decades.

According to Ramus, Nespor and Mehler (1999) the percentage of vowels (%V) and the standard deviations of consonantal (ΔC) intervals are the most reliable indexes to predict language-specific rhythmic tendencies, as they mirror the effects of phonotactic complexity and vowel reduction. For instance, given that stress-timed languages display more syllable types than other rhythmic classes, they are marked by high ΔC and low %V values :

> More syllable types mean more variability in the number of consonants, more variability in their overall duration in the syllable, and thus a higher ΔC (Ramus *et al.*, 1999: 8).

Vice versa, ΔC values in syllable timed-languages are expected to be comparatively lower. However, the metrics developed by Ramus and colleagues are partially flawed due to the intrinsic nature of standard deviation. As it rests on the mean value of a dataset, standard deviation computes higher dispersion indexes for datasets which contain statistical outliers: empirical research on language rhythm can be heavily hindered by statistical outliers, since numerous extra-linguistic factors may trigger the variation of both individual and inter-individual speech tempos (Wagner, 2008). For instance, fast speech determines a decrease in the duration of consonantal intervals, a dynamic which ultimately results into low ΔC values.

In order to factor out the effects of speech tempo, Dellwo and Wagner (2003) emphasized the importance of adopting a *variation coefficient*: dividing the ΔC parameter by the mean value of the consonantal durations allows to measure the variability of consonantal intervals on a relative scale. In this way, the new parameter varcoΔC was obtained:

$$\text{varco}\Delta C = \Delta C / \text{mean}C * 100$$

White and Mattys (2007) showed the relevance of variation coefficients with respect to the standard deviation of vocalic intervals (VarcoΔV). Their research focused on native English, Dutch, French and Spanish speakers, as well as non-native English speakers (whose native language was either Spanish or Dutch), non-native Dutch speakers and non-native Spanish speakers (whose native language was English). The variation coefficient appeared to be particularly useful to identify vocalic variations in L2 speakers. Indeed, mainstream syllable-timed languages – such as Italian, French and Spanish – usually display low values of both varcoC and varcoV, while traditional stress-timed languages – such as English and Dutch – show higher values for both parameters. In their work, the VarcoΔV index was higher in native English speakers' utterances than in native Spanish productions. However, L2 English speakers with a Spanish linguistic background displayed lower values of normalized vocalic variation than English natives – i.e. they produced weaker contrasts between stressed and unstressed syllables; still, they displayed more vocalic variation than native Spanish speakers.

Grabe and Low (2002) proposed a different local approach towards the analysis of speech rhythm. They developed a new metric, the PVI (*Pairwise Variability Index*), which computes the mean of all the durational differences between adjacent pairs of vocalic and consonantal intervals. The so-called rPVI ("raw" PVI) is not subject to normalization:

$$rPVI = \sum_{i=1}^{n-1} \frac{|d_i - d_{i+1}|}{n-1}$$

Conversly, in the nPVI ("normalized" PVI) every interval is divided by the mean value of two adjacent intervals. In this way, the effects of the articulation rate are marginalized:

$$nPVI \ = 100 * \sum_{i=1}^{n-1} |\frac{d_i - d_{i+1}}{d_i + d_{i+1}}| \ /(n-1)$$

PVI indexes were used in a number of studies on bilingual acquisition. For instance, Whitworth (2002) studied the spontaneous speech of six English-German simultaneous bilingual children aged between 5;0 and 13;2: younger children displayed low Vocalic nPVIs and high Consonantal rPVI values, while older children exhibited comparatively higher VnPVIs and lower CrPVIs. According to the author, the bidirectional trend drawn by PVIs proves the following:

> Young children initially produce vowels of roughly equal length and then have to acquire the correct amount of vowel reduction. Similarly, consonant clusters are initially produced with a greater amount of durational variability, which is then gradually reduced (Whitworth, 2002: 201).

3. The study

The present research was conducted on a sample of early sequential bilingual children from H-International School (Roncade, Treviso), where students follow an Italian-English *immersion program* that fosters sequential bilingualism. from kindergarten to 5th grade – the PYP (Primary Years Program). The PYP includes the Early Years Unit (EYU), a nursery school where children are totally immersed in an English-speaking environment; for the following four years of schooling the L2 immersion ratio is reduced to 80%, while it becomes perfectly balanced to L1 exposure (50%) during the last year of primary education.

In an attempt to test the goodness-of-fit of C-V temporal distributions as the building blocks of speech rhythm, the present research looks into (semi-spontaneous) samples produced by early sequential bilingual subjects – i.e. children who have been exposed to the L2 since age 3;0 (Meisel, 2011). Two research questions were established:

1. Do sequential bilinguals display different language-specific rhythms during L1 and L2 speech?
2. How does the rhythmic profile of their speech change with age?

3.1. Participants

The study focused on 9 students (aged between 6;7 and 10;11), equally distributed among Y1 (Year 1), Y3 and Y5 classes. The following Table reports on their age and sex.

Table 1: List of the Subjects.

	Subject (code)	Age	Sex
Y1	HF_3_19	6;7	M
	HF_5_19	6;7	F
	HF_6_19	6;9	F
Y3	HF_8_19	9;0	F
	HF_10_19	8;6	M
	HF_11_19	9;3	F
Y5	HF_14_19	10;7	M
	HF_16_19	10;9	F
	HF_18_19	10;11	F

The linguistic background of each participant was framed by means of an online parental questionnaire, which was arranged and submitted via email to the parents. The questions were designed after the form developed by Roch, Florit and Levorato (Department of Developmental Psychology – University of Padua, 2015).

The aim of the survey was to investigate a number of phenomena and contexts that have an impact on bilingual acquisition, namely: the nationality and geographical origin of the family; the amount of time spent by the children in a foreign country; the Italian varieties and dialects they had been exposed to over the years; the children's active and passive language abilities; any potential influence from other foreign languages.

All participants were born and raised in Veneto region, where they had been exposed to standard Italian and local dialects, with the exception of subject HF_6_19 – who was also exposed to an Umbrian dialect – and subject HF_18_19 – who was raised in Milan and exposed to Milanese dialect. Finally, all participants were exposed to English within the educational context of PYP.

3.2. Method

3.2.1. Data collection

Speech data were obtained by recording the 9 selected Subjects during a narrative test. The test involved the audio-visual presentation of two short stories – one in Italian and one in English – on a computer screen. Each story was narrated by the voice of a mother-tongue speaker on a PPSX file, with three couples of stimulus pictures synchronized with the audio and presented in sequence to reinforce the perception of the events. After listening to each story, all children were engaged in a retelling task. Their productions were recorded by means of a smartphone's internal microphone and audio recorder – 44.1 kHz sampling rate and 16-bit resolution. The device was placed on the table in front of the participant, a few centimeters away from the center of their chest and at an approximate distance of 15-20 cm below their mouth.

The texts employed for the assessment of narrative skills are entitled '*The Cat'* (English text) and '*Il cane*' (Italian text). Both scripts – see APPENDIX A and B – are part of the "*Multilingual Assessment Instrument for Narratives*" (MAIN), a tool specifically designed to evaluate the storytelling abilities of 3 to 10 year-old bilingual children and possibly identify the onset of specific language impairments (Gagarina *et al.*, 2012). The texts portray different characters and events, but they are cross-linguistically comparable for *macrostructure* (i.e. high-order episodic categories such as Internal State, Attempt, Goal and Outcome) and *microstructure* information (e.g. lexical diversity, connectives, referential devices, etc.). Two sentence examples from the scripts are provided hereafter:

(1) *"One day there was a playful cat who saw a yellow butterfly sitting on a bush".* (Extract from *'The Cat'*)

(2) *"Un giorno c'era un cane giocherellone che vide un topo seduto vicino ad un albero".* (Extract from *'Il Cane'*)

The retelling mode represents a rich source of information about the pragmatic use of language and provides a realistic portrayal of the interaction between early grammar and children's cognitive processing of external events (Gillam and Carlisle, 1997). Besides, the elicited nature of the sampling procedure facilitates the arrangement of consistent research protocols by minimizing the variation degree of the outcomes' content, structural complexity and length.

Furthermore, retelling tasks provide interesting material for the study of the physical properties of language. While spontaneous speech is highly unstable – it features disfluencies such as false starts, filled pauses and word repetitions – , the outputs of text reading activities tend not to match the natural prosodic layout of unprepared speech: the administration of retelling tasks may represent a valuable accommodation between the two sampling modes.

3.2.2. Speech segmentation and analysis

The 18 speech samples were subject to a preliminary phase of automatic text/audio alignment. The procedure was carried out via WebMAUS (Schiel, 1999; Kisler *et al.*, 2017), an online tool developed by the Bavarian Archive for Speech Signals (University of Munich). First, the speech content of the recordings was manually transcribed on different .txt files; then, each text file was uploaded on the web interface together with its corresponding audio file.

WebMAUS uses a grapheme-to-phoneme algorithm to convert the tokenized words of the .txt file into SAMPA phonetic forms. The tool produces a statistically-weighted graph of pronunciation variants based on a probabilistic set of language-specific phonological rules. Finally, it uses a set of HMMs (Hidden Markov Models) to time-align the speech spectrogram with the most probable phonetic outcome. The TextGrids obtained from the preliminary phase were uploaded on Praat (Boersma and Weenink, ver. 6.0.45, 2019) and subject to manual realignment within the spectrogram-textgrid interface. Given the purpose of the metric analysis, specific segmentation and labeling criteria were established for a consistent categorization of the following phenomena: sentence-initial voiceless stops and affricates; vocalic epenthesis; vowel lengthening; long transitions between adjacent segments.

It is also important to emphasize that children's speech rate varied significantly both across the two retelling tasks and within their own productions. Local fluctuations in the speech rates of younger students during the Italian retelling task often determined the elision of unstressed vowels and the creation of complex consonantal clusters across word-boundaries (Figure 1).

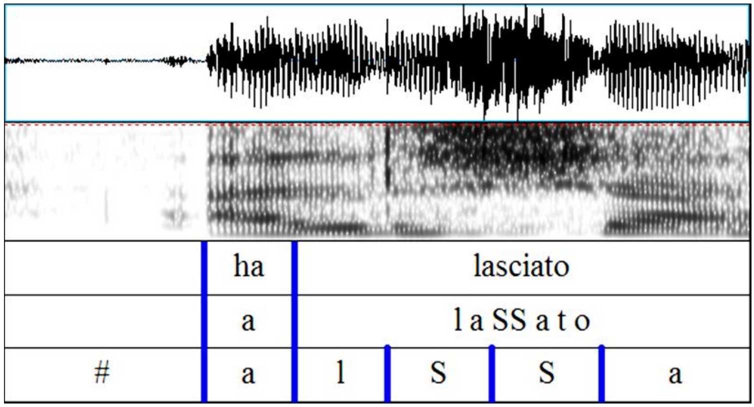

Figure 1: Elision of the unstressed vowel [a] in "lasciato". It can also be noticed that phonologically long phonemes (such as geminate consonants in Italian) were annotated with two labels: the boundary between the two is fictitious, but the procedure was necessary to prepare the transcripts for the metric analysis via Correlatore.

However, high degrees of consonantal complexity might be envisioned as good indicators of language fluency: according to Housen and Kuiken (2009), the speed and ease with which utterances are delivered reflect the speakers' control over linguistic knowledge. Similarly, speech disfluencies can be interpreted as representative indexes of language competence: while hesitation markers were labeled as pauses because they would alter the ratio between consonantal and vocalic intervals, speech disfluencies such as word-repetitions, false starts and corrections were subject to segmentation.

Finally, the metric analysis of the speech samples was computed via *Correlatore* (Mairano, ver. 2.3.4, 2014), a Tcl/Tk program specifically engineered to calculate rhythm metrics. The third tier (MAU) of each TextGrid was uploaded on the program, which transformed the segmented SAMPA transcriptions into CV sequences. Then, vocalic and consonantal intervals were determined as the sum of adjacent vowels and consonants. In this way, the program could compute the rhythmic profile of each audio sample (18 samples in total – i.e. 2 per subject). Three computation series were performed for the analysis: first, the rhythmic profiles of the samples were obtained by employing the %V and ΔC metrics; then, by implementing the VarcoΔC and VarcoΔV metrics; finally, by measuring the PVIs.

4. Results

For each set of metrics, the values generated by *Correlatore* were subject to both within-group *and* between-group one-way ANOVAs to identify valuable cross-linguistic variations among children of the same age and statistically significant differences between different age groups (Y1, Y3, Y5).

Three different predictions could be made:

- children progressively develop divergent rhythmic patterns for the two languages;
- they stabilize on intermediate rhythmic templates, as claimed by the *rhythmic compromise* hypothesis (Howell and Borsel, 2011);
- they exhibit either stress-timed or syllable-timed tendencies in both languages.

4.1. The %V/ ΔC metrics

4.1.1. Vocalic intervals

The percentage of vocalic intervals is not significantly different between Italian and English productions. For instance, Y1 students display %V scores which vary from 51.98% to 63.20% in Italian (M = 57.65) and from 46.69% to 59.37% in English (M = 52.65). Notably, %V values were expected to decrease in the English productions of Y3 and Y5 children, since they have been exposed to the L2 for a longer period of time: indeed, according to Ramus *et al.* (1999), stress-timed patterns are marked by a low percentage of vowels. However, the amount of vocalic intervals drawn from the L2 speech of the Subjects does not appear to change with age.

Standard deviation on %V samples shows that within-group variability (i.e. variability in vocalic productions across children of the same class) decreases with age in relation to the L1 only, a pattern which may reflect the gradual development of native fluency. Italian %V ranges from 54.11% to 60.26% (M = 57.98) in Y3 students, while it only varies from 57.74% to 59.94% (M = 58.84) in Y5 children. On the other hand, the percentage of vocalic intervals in English speech spans from 51.74% to 56.89% (M = 54.26) in Y3 students, while it ranges from 48.92% to over 59.24% (M = 55.53) in Y5 Subjects – i.e. high levels of variability are present across older students, too.

It is interesting to notice that the English %V values retrieved from the analysis are sensibly higher than the average established for the language by other studies. For instance, Ramus et al. (1999) reported that the percentage of vocalic intervals in native English speakers is around 40%. Nonetheless, given the experimental context in which the present research was conducted and the phonetic features of the Subjects' speech, the retrieval of divergent values is not surprising: indeed, the participants' English productions were generally characterized by word-final vocoid insertions (*paragoge*), which probably determined an increase in %V.

4.1.2. Consonantal intervals

The standard deviation of consonantal intervals (ΔC) ranges from 55.86 ms to 72.64 ms (M = 63.72) in Y1 Italian productions, while it varies from 63.41 ms to 82.05 ms (M = 68.73) in Y1 English productions. For Y3 students, the index spans from 47.53 ms to 92.54 ms (M = 69.23) in Italian, while it ranges from 62.61 ms to 74.49 ms (M = 66.12) in English. Finally, Y5 students display ΔC scores which are comprised between 45.80 ms and 73.34 ms (M = 57.08) in Italian, while the index varies from 58.91 ms to 107.70 ms (M = 77.52) in English. From the above-mentioned results it appears that English ΔC values are significantly higher than Italian ΔCs only in Y5 students. However, the difference is not statistically relevant (p >.05) due to great within-group variability. Indeed, it is important to stress how data from the analysis reflect the extreme sensitivity of the ΔC metric towards statistical outliers. For instance, Subject HF_16_19 (Y5) exhibited hesitation in the English retelling task through the lengthening of liquid, nasal and fricative consonants: since standard deviation does not take into account the relative distribution of the intervals, it is plausible that consonantal lengthening caused by thoughtful absorption determined an important leap in the subject's ΔC score (107.70 ms). Such hesitations may have been caused by insecurity in second language use, since the Italian consonantal outputs of the same Subject do not show any lengthening phenomena: indeed, consonantal variation in HF_16_19 appears to be on-target for Italian (ΔC = 45.80 ms). Conversely, during the Italian retelling task both Y1 and Y3 students tended to create complex consonantal clusters across word-boundaries. For instance, Subject HF_11_19 exhibited abrupt accelerations in the speech rate during the articulation of specific phrases (e.g. *“ altra parte*”), a tendency which often resulted in the elision of final-word vowels, while Subject HF_8_19 was prone to erase the vocalic nuclei followed by nasal consonants (e.g. the indefinite determiner “*un*” [u ŋ] was systematically pronounced as [ŋ]) : both phenomena led to the creation of complex consonantal clusters between the offset and the onset of adjacent words.

4.2. The Varcos

4.2.1. Vocalic intervals

Significant trends were identified across different age groups by implementing a variation coefficient to the standard deviation of vocalic intervals (VarcoΔV).

Table 2: VarcoΔV values in Y1, Y3 and Y5 English samples (unit of measurement: ms). Between-group differences are statistically relevant.

Y1		Y3		Y5		p-value
HF_3_19	51.44	HF_8_19	62.15	HF_14_19	72.83	0.0121
HF_5_19	40.07	HF_10_19	58.77	HF_16_19	74.67	
HF_6_19	44.74	HF_11_19	58.34	HF_18_19	57.93	

Y1 English narrative productions display VarcoΔVs which are comparable to Mairano and Romano's (2009) figures on standard Italian (around 43 ms). However, the one-way ANOVA computed between the VarcoΔVs of Y1, Y3 and

Y5 children showed a gradual and statistically significant growth in the values of the metric ($F(2,6) = 10.05566$; $p = 0.0121$). The VarcoΔV scores extracted from Y3 and Y5 productions resemble the average attested by Mairano and Romano for the received pronunciation of English (around 63 ms). Therefore, age and time of exposure appear to correlate positively with the development of vocalic variation in English: this pattern could mirror the (relatively) late acquisition of reduction properties which are typical of stress-timed languages.

Interestingly, a similar proclivity towards the production of uneven vocalic intervals was found in the Italian speech samples of Y3 and Y5 students. More specifically, Italian VarcoΔV ranges from 46.51 ms to 59.50 ms ($M = 52.58$) in Y1 students, while it is comprised between 57.63 ms and 76.66 ms ($M = 68.49$) in Y3 students, and between 62.65 ms and 80.62 ms ($M = 71.92$) in Y5 students. Between-group differences are marginally relevant ($F(2,6) = 4{,}36615$, $p = 0.06755$). Besides, it can be noticed that the VarcoΔV values obtained from the Italian narrative samples are very similar to the values derived from the analysis of English productions. Such results stand out when compared to those obtained by White and Mattys (2007) with respect to English-Spanish speakers, whose L2 "intermediate" VarcoΔV values reflected the impact of native language on the development of L2 rhythm. It could be inferred that the parallel shift of early Italian and English rhythms towards stress-timed templates is determined by the dominant presence of English in PYP educational settings. However, the hypothesis collides with the results obtained from the analysis of the %V metric: is it plausible that children whose L2 is marked by pervasive occurrences of vocalic epenthesis manage to acquire native-like vowel reduction? Further research on a broader sample of subjects is needed in order to test this possibility.

4.2.2. Consonantal intervals

The VarcoΔC values of Italian productions do not vary significantly across children of different classes: the metric ranges from 50.23 ms to 60.13 ms ($M = 55.88$) in Y1 students, from 47.80 ms to 85.43 ms ($M = 65.79$) in Y3 students, from 51.30 ms to 75.54 ms ($M = 60.23$) in Y5 students. The higher mean of the metric in Y3 productions is the result of great within-group variability. The English samples, on the other hand, display an increase in VarcoΔC values between Y3 and Y5 productions: however, the difference is not statistically relevant ($F(2,6) = 2.62814$, $p = 0.15145$) due to the great variability of the metric within the Y5 group, where VarcoΔC is comprised between 50.26 ms and 77.74 ms.

Italian VarcoΔC values also appear to be higher than the average determined by Mairano and Romano (2009) for Italian (around 47 ms). High values of VarcoΔC in the Italian productions of Y3 Subjects HF_8_19 (85.43 ms) and HF_11_19 (64.14 ms) reflect their tendency to form complex consonantal clusters. Vice versa, low within-group variability with respect to the English counterpart of the metric could be associated to the relative absence of fluctuations in the speech rate. Therefore, the approximate articulation rate of the two abovementioned Subjects was calculated as the ratio between the number of vocalic intervals and the speech duration (pauses excluded): in both cases, it was found that the amount of syllables per second was significantly lower in English productions (Table 3).

Table 3: Approximate articulation rates (nV-Int./articulation time) of Subjects HF_8_19 and HF_11_19.

Participant	Syll./s - ITA	Syll./s - ENG
HF_8_19	3.76	2.57
HF_11_19	4.47	2.98

Such results are consistent with findings by Dellwo (2006), who claimed that syllable-timed languages are marked by a positive correlation between speech rate and VarcoΔCs: in other terms, local fluctuations in the speech rate during the Italian narrative test could be envisioned as the source of consonantal complexity and high Italian VarcoΔC values.

As already mentioned, VarcoΔC values appear to be relatively higher in Y5 English narrative productions. However, such values do not diverge significantly from the VarcoΔCs extracted from Y5 Italian productions. In order to outline a more precise rhythmic profile of Y5 productions, the consonantal Varcos of Y5 samples were analyzed individually. In Subject HF_14_19 the values of the metric are 77.55 ms and 61.86 ms for Italian and English respectively, while HF_16_19 shows the opposite ratio: VarcoΔC is comparatively lower in Italian (51.30 ms) than it is in English (77.75 ms). Finally, Subject HF_18_19 displays similar levels of consonantal variation in the two languages: 53.84 ms in Italian and 50.26 ms in English. In an attempt to account for such variability, the above-mentioned values were compared to the individual children’s articulation rates (Table 4).

Table 4: Approximate speech rates (nV-Int./speech time) of Subjects HF_14_19, HF_16_19 and HF_18_19.

Participant	Syll./s - ITA	Syll./s - ENG
HF_14_19	4.34	3.06
HF_16_19	4.60	2.69
HF_18_19	4.24	3.26

The positive correlation between articulation rate and consonantal variation in Subject HF_14_19 supports the hypothesis drawn from the analysis of Y3 students: fluency in the L1 probably triggered the creation of complex consonantal clusters in Italian. Conversely, it is possible that the comparatively higher values of English consonantal variation in HF_16_19 are linked to the invasive occurrence of consonantal lengthening in their L2 speech, as already illustrated during the analysis of ΔC results. Finally, Subject HF_18_19 displays a comparatively higher articulation rate in Italian but similar VarcoΔC values in the two languages. This suggests that other factors can contribute to the durational variability of consonantal intervals: one example is represented by the syllable-initial voiceless stops produced by all nine Subjects, which is marked by Italian-like VOT values (i.e. < 40 ms) that can ultimately alter the durational ratio between vowels and consonants. Specific research on the durational and qualitative features of the phonemes produced by young bilinguals should be conducted in order to test the correlation between early phonetic repertoires and the development of rhythmic patterns.

4.3. The PVIs

4.3.1. Vocalic intervals

The normalized computation of a Vocalic Pairwise Variability Index (VnPVI) on English narrative samples revealed that durational differences between adjacent vocalic intervals tend to increase with age. The value of VnPVIs ranges from 44.50 ms to 52.82 ms (M = 47.38) in Y1 samples, from 53.45 ms to 57.63 ms (M = 54.51) in Y3 productions and from 57.43 ms to 70.89 ms (M = 62.04) in Y5 samples. The one-way ANOVA computed between Y1, Y3 and Y5 VnPVIs disclosed a statistically significant growth in the metric scores (F = 5.46208; p = 0.04456).

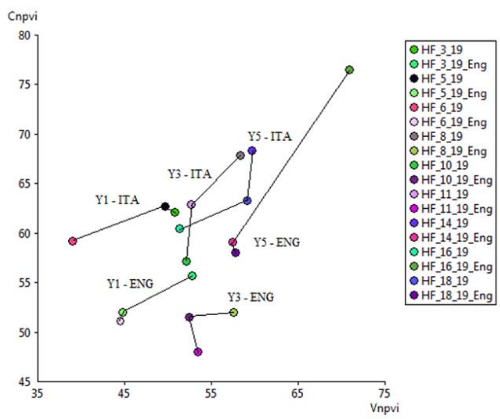

Figure 5: Italian and English narrative productions plotted on a VnPVI/CnPVI diagram. It can be noticed that age correlates with the movement of narrative productions towards the upper-right side of the graph – where stress-timed languages are usually plotted due to their high levels of vocalic variability.

Conversely, the VnPVIs obtained from Italian narrative samples do not increase significantly with age (p >.05): the metric ranges from 38.96 ms to 50.76 ms (M = 46.49) in Y1 speech samples, from 52.08 ms to 58.33 ms (M = 54.35) in Y3 samples and from 51.29 ms to 59.73 ms (M = 56.71) in Y5 productions. It can be noticed that the value of the index for both Italian and English Y1 productions are close to the average established by Mairano and Romano (2009) for standard Italian – around 45-50 ms. However, Y3 and Y5 students exhibit Italian and English VnPVI scores which are closer to the typical values of Queen's English – around 64 ms, according to Mairano and Romano (2009).

4.3.2. Consonantal intervals

Since the articulation of consonants appeared to be particularly affected by individual speech rates, the computation of consonantal PVIs was subject to normalization, too (CnPVIs). Y5 students display the same ratios determined by VarcoΔCs: in Subject HF_14_19 the CnPVIs are 68.29 ms and 59.23 ms for Italian and English respectively; HF_16_19 displays comparatively lower CnPVI scores in Italian (60.54 ms) than in English (76.54 ms); Subject HF_18_19 displays similar levels of consonantal variation in the two languages (63.25 ms in Italian and 58.02 ms in English). However, the one-way ANOVA computed between the values of the metric in Italian and English productions revealed significant patterns within the groups Y1 and Y3. Y1 students exhibit CnPVI values which range from 59.20 ms to 67.72 ms in Italian speech, while the metric varies from 51.08 ms to 55.67 ms in English productions: English CnPVIs are significantly lower than the Italian ones ($F(1,4) = 22.42657$, $p = 0.00907$). Y3 students exhibit CnPVI values which range from 57.23 ms to 67.78 ms (M = 62.65) and from 47.98 ms to 51.98 ms (M = 50.50) in Italian and English productions respectively: the metric scores derived from the Italian narratives are significantly higher ($F(1,4) = 13.52248$, $p = 0.02125$).

5. Conclusion

The study aimed at investigating whether Italian and English speech rhythms undergo any form of differentiation in early sequential bilinguals. The research was conducted on a sample of 9 early bilingual children aged between 6;7 and 10;11 and equally distributed among three different age groups. Their semi-spontaneous narrative productions were recorded in both languages and subject to automatic text/audio alignment and manual re-segmentation. Finally, the segmented transcripts were uploaded on the program *Correlatore*, which computed three sets of interval-based metrics (%V and ΔC; Varcos; PVIs) and traced the rhythmic profile of each Subject. The analysis addressed both within-group rhythmic differences between L1 and L2 productions and any significant change in the values of the metrics across different age groups.

The examined speech samples displayed high %V scores for both Italian and English: indeed, English productions were marked by pervasive occurrences of final-word vocalic epenthesis. However, the comparative analysis performed between the VarcoΔV values of Y1, Y3 and Y5 students revealed a statistically relevant increase in vocalic variation phenomena – which are typical of stress-timed languages – both in Italian and in English productions. The same trend was outlined by the computation of the VnPVI metric.

Conversely, consonantal variation indexes did not appear to change significantly across children of different ages. However, Y1 and Y3 CnPVI scores were significantly higher in Italian productions, a result which mirrored children's propensity to accelerate their speech rate and create complex clusters across word-boundaries through the elision of within-phrase word-final vowels and the nasalization of syllabic nuclei. On the other hand, the analysis of Y5 students' narratives revealed different consonantal variation trends. One Subject exhibited an asymmetric increase in English VarcoΔC and CnPVI values relative to their Italian scores: the result was influenced by the abnormal lengthening of nasal and fricative consonants, probably caused by instances of thoughtful absorption during the English retelling task. A different Subject displayed comparatively lower VarcoΔC and CnPVI scores in both languages, suggesting that other factors – such as the dominance of the L1 phonetic repertoire – might have had an impact on the durational length of consonantal intervals.

In conclusion, the computation of interval-based metrics on early sequential bilingual speech outlined a progressive development of *stress-timed* patterns in both L1 Italian and L2 English. Importantly, Subjects did not display any "intermediate" rhythmic trends. However, data show that interval-based metrics – even when subject to normalization – are extremely sensitive to hesitation and local fluctuations in the speech rate, two aspects of online speech processing that cause frequent disfluencies in young children. Disfluencies are notably problematic for empirical research on prosody, as they can alter the quality of segmental information – and, by extension, the rhythmic profile of a speech stream. In order to factor out the effects that speech tempo has on phonotactic complexity, further research could adopt different sampling modes – such as reading and sentence-repetition tasks – and cross-correlate the metric results that stem from each form of speech elicitation.

References

Abercrombie, David. 1967. *Elements of general phonetics*. Chicago: Aldine Publishing Corporation.

Allen, George D., Hawkins, Sarah. 1980. Phonological Rhythm: Definition and Development. In G. H. Yeni-Komshian, J. F. Kavanagh and Ch. A. Ferguson (eds.), *Child Phonology, Vol. 1: Production*, 227-256. Amsterdam: Elsevier. https://doi.org/10.1016/B978-0-12-770601-6.50017-6

Baker, Colin. 2001. *Foundations of Bilingual Education and Bilingualism* (3rd ed). Clevedon: Multilingual Matters.

Carter, P. M. 2005. Quantifying Rhythmic Differences Between Spanish, English, and Hispanic English. In E. J. Rubin and R. S. Jess (eds), *Theoretical and Experimental Approaches to*

Romance Linguistics: Selected papers from the 34th Linguistic Symposium on Romance Languages (LSRL). Salt Lake City. https://doi.org/10.1075/cilt.272.05car

Dellwo, V. 2006. A variation coefficient for deltaC. In P. Karnowski and I. Szigeti (eds.), *Language and Language Processing*, 231-241. Frankfurt am Main.

Dellwo, Volker, and Wagner, Petra. 2003. Relations between language rhythm and speech rate. *Proceedings of the International Congress of Phonetic Sciences, Barcelona, Spain, 3 August 2003 - 9 August 2003*, 471-474.

Gagarina, Natalia, Klop, Daleen, Kunnari, Sari, Tantele, Koula, Välimaa, Taina, Balčiūnienė, Inrigda, Bohnacker, Ute, and Walters, Joel. 2015. *Assessment of Narrative Abilities in Bilingual Children.* In S. Armon-Lotem J. de Jong, and Natalia Meir (eds.), *Assessing Multilingual Children*, 243-269. https://doi.org/10.21832/9781783093137-011

Grabe, Esther, Gut, U., Post, Brechtje, and Watson, Ian. 2001. The acquisition of rhythm in English, French and German. *Current research in language and communication: Proceedings of the Child Language Seminar*, 156-162.

Grabe, Esther, Low. Ee Ling. 2002. Durational Variability in Speech and the Rhytm Class Hypothesis. *Papers in Laboratory Phonology, 7*, 515-546. https://doi.org/10.1515/9783110197105.2.515

Grabe, E., Watson, Ian, and Post, Brechtje. 1999. The acquisition of rhythmic patterns in English and French. *Proceedings of the 14th International Congress of Phonetic Sciences*, 1201-1204.

Housen, Alex, and Kuiken, Folkert. 2009. Complexity, Accuracy, and Fluency in Second Language Acquisition. *Applied Linguistics, 30(*4), 461-473. https://doi.org/10.1093/applin/amp048

Howell, P., Borsel, J. V. 2011. *Multilingual Aspects of Fluency Disorders*. Multilingual Matters. https://doi.org/10.21832/9781847693570

Johnson, Keith. 2012. *Acoustic and Auditory Phonetics* (3rd ed). Chichester: Wiley-Blackwell.

Kehoe, Margaret, Lleó, Conxita, and Rakow, Martin. 2011. Speech rhythm in the pronunciation of German and Spanish monolingual and German-Spanish bilingual 3-year-olds. *Linguistische Berichte*, *227*, 323-351.

Kisler, Thomas, Reichel, Uwe D., and Schiel, Florian. 2017. Multilingual processing of speech via web services. *Computer Speech & Language, 45,* 326-347. https://doi.org/10.1016/j.csl.2017.01.005

Konopczynski, Gabrielle. 1995. A developmental model of acquisition of rhythmic patterns: results from a cross-linguistic study. *Proceedings of the Intonational Congress of Phonetic Sciences, Stockholm, Sweden, 4*, 22-25.

Li, Aike, and Post, Brechtje. 2014. L2 Acquisition of Prosodic Properties of Speech Rhythm: Evidence from L1 Mandarin and German Learners of English. *Studies in Second Language Acquisition, 36*(2), 223-255. https://doi.org/10.1017/S0272263113000752

Mairano, P., Romano, A. 2009. Un confronto tra diverse metriche ritmiche usando Correlatore 1.0. La dimensione temporale del parlato. *AISV Conference 2009, University of Zurich, Kollegiengebäude, 4-6 february 2009*. Torriana: EDK.

Mehler, Jacques, Jusczyk, Peter, Lambertz, Ghislaine, Halsted, Nilofar, Bertoncini, Josiane, and Amiel-Tison, Claudine. 1988. A precursor of language acquisition in young infants. *Cognition, 29*(2), 144-178. https://doi.org/10.1016/0010-0277(88)90035-2

Meisel, Jürgen M. 2008. Child second language acquisition or successive first language acquisition? In B. Haznedar and E. Gavruseva (eds.), *Current Trends in Child Second Language Acquisition*, 55-80. Amsterdam: John Benjamins. https://doi.org/10.1075/lald.46.04mei

Mok, Peggy. 2013. Speech rhythm of monolingual and bilingual children at 2;06: Cantonese and English. *Bilingualism: Language and Cognition, 16*, 693-703. https://doi.org/10.1017/S1366728912000636

Payne, Elinor, Post, Brechtje, Prieto, Pilar, Vanrell, Maria, and Astruc, Lluisa. 2012. Measuring child rhythm. *Language and Speech, 55*, 202-228. https://doi.org/10.1177/0023830911417687

Pike, Kenneth. 1945. *The Intonation of American English*. University of Michigan Press.

Ramus, F., Nespor, and M. Mehler, J. 1999. Correlates of Linguistic Rhythm in the Speech Signal. *Cognition, 73*(3), 265-292. https://doi.org/10.1016/S0010-0277(99)00058-X
Roch, M., Florit, E., and Levorato, C. 2016. Narrative competence of Italian–English bilingual children between 5 and 7 years. *Applied Psycholinguistics, 37*(1), 9-67. https://doi.org/10.1017/S0142716415000417
Schiel, F. 1999. Automatic Phonetic Transcription of Non-Prompted Speech. *Proc. of the ICPhS*, 607-610.
Vihman, M. M., Nakai, S. and DePaolis, R. A. 2006. Getting the rhythm right: A cross-linguistic study of segmental duration in babbling and first words. In L. Goldstein, D. H. Whalen and C. T. Best. *Laboratory Phonology 8: Phonology and Phonetics*, 341-366. Mouton De Gruyter.
Wagner, P. S. 2008. *The Rhythm of Language and Speech: Constraining Factors, Models, Metrics and Applications*. University of Bonn.
White, Laurence, Floccia, Caroline, and Delle Luche, Claire. 2016. Five-month-old infants discrimination of unfamiliar languages does not accord with "rhythm class". *Conference of Speech Prosody. Boston.* https://doi.org/10.21437/SpeechProsody.2016-116
White, Laurence, and Mattys, Sven L. 2007. Calibrating rhythm: First language and second language studies. *Journal of Phonetics, 35*(4), 501-522. https://doi.org/10.1016/j.wocn.2007.02.003
Whitworth, Nicole. 2000. Acquisition of VOT, and vowel length by English-German bilinguals: A pilot study. *Leeds Working Papers in Linguistics and Phonetics, 8,* 1–16.
Whitworth, Nicole. 2002. Speech rhythm production in three German-English bilingual families. *Leeds Working Papers in Linguistics and Phonetics, 9*, 175–205.