

Calibrating rhythm: First language and second language studies

Laurence White*, Sven L. Mattys

Department of Experimental Psychology, University of Bristol, 12a Priori Road, Bristol BS8 1TU, UK

Received 8 September 2005; received in revised form 5 February 2007; accepted 12 February 2007

Abstract

This paper presents a comparative evaluation of metrics for the quantification of speech rhythm, comparing pairwise variability indices (nPVI-V and rPVI-C) and interval measures (ΔV , ΔC , %V), together with rate-normalised interval measures (VarcoV and VarcoC). First, we examined how well these metrics discriminated “stress-timed” English and Dutch and “syllable-timed” Spanish and French. Metrics of interval standard deviation such as ΔV and ΔC were strongly influenced by speech rate, but rate-normalised metrics of vocalic interval variation, VarcoV and nPVI-V, were shown to discriminate between hypothesised “rhythm classes”, as did %V, an index of the relative duration of vocalic and consonantal intervals. Second, we applied these metrics to quantifying the influence of first language on second language rhythm, with the expectation that speakers switching “rhythm classes” should show rhythm scores different from both their native and target languages. VarcoV offered the most discriminative analysis in this part of the study, with %V also suggesting insights into the process of accommodation to second language rhythm.

© 2007 Elsevier Ltd. All rights reserved.

1. Introduction

Rhythm derives from the repetition of elements perceived as similar. In speech, these elements are syllables, or stressed syllables in particular. Metrics for comparing the rhythm of different languages—interval measures (Ramus, Nespore, & Mehler, 1999) and pairwise variability indices (Grabe & Low, 2002; Low, Grabe, & Nolan, 2000)—have been relatively successful in distinguishing languages which fall into the conventional rhythmic categories, “stress-timed” and “syllable-timed” (e.g. Pike, 1945). This categorisation is based on broad perceptual judgements, however, with no existing quantitative yardstick, so assessment of the relative adequacy of the metrics in these different studies is not straightforward, particularly given the use of different speakers, materials and recording conditions, together with variability in speech rate and in techniques for segmentation of the signal into vowels and consonants.

We have taken two approaches to the comparison of rhythm metrics. First, we assessed the rhythm metrics in the “traditional” task of distinguishing between languages from different rhythm classes. Materials, speakers and methods were kept constant between the different metrics, and we examined both raw and rate-normalised metrics. Second, we applied the metrics to the assessment of the influence of first language (L1) on the rhythm of second language (L2) speakers. We chose L2 speakers who were competent but had non-native

*Corresponding author. Tel.: +44 117 9288450; fax: +44 117 9288588.

E-mail addresses: laurence.white@bristol.ac.uk (L. White), sven.mattys@bristol.ac.uk (S.L. Mattys).

accents, with the expectation that L2 rhythm scores of such speakers should differ both from those of the L2 speakers' native languages and from those of the target languages.

1.1. *Speech rhythm and the rhythm class hypothesis*

Within the languages of Western Europe, Romance languages, such as Spanish and French, have been described as “syllable-timed” and Germanic languages, such as English and Dutch, have been described as “stress-timed” (Abercrombie, 1967; Pike, 1945).¹ This typological dichotomy was originally related to isochronous speech intervals, with the hypothesis that syllables tend to be of equal duration in syllable-timed languages and that stress-delimited feet tend to be of equal duration in stress-timed languages. Instrumental studies have shown, however, that stress-based or syllable-based isochrony is not systematic in either rhythm class (see Ramus et al., 1999, for a review).

Despite the lack of isochronous units of speech timing, there have been empirical demonstrations of perceptual consequences arising from the stress-timed vs syllable-timed distinction. Using low-pass filtered speech to reduce segmental information, Nazzi, Bertoni, and Mehler (1998) showed that French neonates discriminate a combination of stress-timed Dutch and English sentences from syllable-timed Spanish and Italian, while not discriminating cross-rhythm class groupings such as Dutch and Italian compared with Spanish and English. Using speech resynthesised to eliminate segmental and intonational variability, Ramus, Dupoux and Mehler (2003) showed that adult speakers could discriminate between rhythm classes, e.g. English vs. Spanish, but not within rhythm classes, e.g. English vs Dutch.

What within the speech signal underpins these rhythmic distinctions? Dasher and Bolinger (1982) and Roach (1982) suggest that the degree of vowel reduction in unstressed syllables is important in making stressed syllables relatively salient in stress-timed languages. This is just one of the factors which make the durational difference between stressed and unstressed syllables greater than in syllable-timed languages: Delattre (1966) found that non-final open syllables are 50% longer when stressed than unstressed in English and 60% longer in German, but only 10% longer in Spanish. Furthermore, Dauer (1983, 1987) observed that a stress-timed language like English has a wide variety of syllable structures available, with greater complexity allowable in onsets and codas than in syllable-timed languages. It is the heavier syllables that attract stress in stress-timed languages, a trend which is absent or much less marked in syllable-timed languages. Open syllables are more widespread in syllable-timed languages: over half of Spanish or French syllables have a CV structure (Dauer, 1983).

1.2. *Rhythm metrics*

The pairwise variability index (PVI) and interval measures (IM) exploit syllable complexity, vowel reduction and stress-based lengthening to provide metrics of rhythm. They are acoustic measures, deriving from a simple segmentation of the speech string into vocalic and consonantal² intervals, measuring variability in these intervals on the basis that: (a) stress-timed languages tend to have a greater contrast in vowel duration between stressed and unstressed syllables; (b) stress-timed languages tend to have greater variation in the complexity of consonant clusters and hence in the duration of consonantal intervals. In all cases, no reference is made to the phonological constituency of intervals: both Ramus et al. (1999) and Grabe and Low (2002) included onset consonants in the same interval as preceding coda consonants and adjacent heterosyllabic vowels in the same vocalic interval.³

¹Languages like Japanese, which have been held to belong to a third, “mora-timed”, rhythmic class (e.g. Port, Dalby, & O’Dell, 1987), are not investigated in the present study.

²Some studies refer to “intervocalic” intervals rather than “consonantal” intervals. For consistency with the names of the metrics— ΔC , VarcoC, rPVI-C—we use the term “consonantal” throughout.

³Some other rhythm metrics have been proposed recently that look at variation in the duration of phonological constituents, such as syllables (e.g. the Variability Index of Deterding, 2001). There are theoretical and practical reasons for not testing such metrics here. First, by simply appealing to acoustic concepts such as the duration of vocalic and consonantal intervals, Ramus et al. (1999) attempted to provide a developmentally plausible account of rhythm, arguing that even very young infants can distinguish vowels and consonants, but initially lack a concept of syllabification. Second, distinguishing vocalic and consonantal intervals allows the possibility that languages

1.2.1. Interval measures

Ramus et al. (1999) proposed three rhythm metrics: ΔV , the standard deviation of vocalic intervals; ΔC , the standard deviation of consonantal intervals; %V, the proportion of total utterance duration which comprises vocalic intervals. They applied these metrics to four speakers reading five sentences each for: English, Dutch, Polish, categorised as stress-timed; French, Spanish, Italian, Catalan, categorised as syllable-timed; Japanese, categorised as mora-timed. The three rhythm classes had significantly different scores for ΔC and %V, but not for ΔV , primarily because ΔV for Polish was even lower than ΔV for syllable-timed languages.

A subsequent study suggested that the distinction by ΔV of Polish from English and Dutch might have perceptual validity. Ramus et al. (2003) showed that adults listening to segmentally and intonationally degraded speech cannot discriminate English and Dutch, but can discriminate English and Polish. Thus, a combination of ΔV and either ΔC or %V appear to offer useful measures of rhythmic relatedness between languages.

1.2.2. Pairwise variability indices

Taking a parallel approach, but attempting to capture the sequential nature of rhythmic contrasts, Low et al. (2000) proposed the pairwise variability index to exploit the durational difference between successive syllables—specifically stressed and unstressed vowels—which tends to be much greater in stress-timed languages. The normalised Pairwise Variability Index (nPVI) is the mean of the differences between successive intervals divided by the sum of the same intervals, the latter step included to control for speech rate variation:

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

where m is the number of intervals and d is the duration of the k th interval.

Low et al. compared the nPVI scores for vocalic intervals (nPVI-V) for British English and Singapore English. The latter variety has been held to be more syllable-timed than the former and its lower nPVI-V score reflected the perceived rhythmic distinction. Low et al. also applied Ramus et al.'s IM, finding that ΔV showed a similar pattern to the nPVI-V. They questioned the utility of %V, as it did not reflect the hypothesised differences between the accents. Grabe and Low (2002) showed, however, that Singapore English is only slightly more syllable-timed than British English, when placed in the context of a range of languages.

By analogy with ΔC , Low et al. suggested a consonantal PVI, particularly for cross-linguistic comparisons, for example, where languages may have properties of both stress-timing and syllable-timing. Grabe and Low (2002) proposed that speech rate normalisation is not appropriate for consonantal intervals, in part because overall differences between languages in consonantal interval duration are a function of their phonotactics and should be captured by rhythm metrics rather than normalised away. The raw Pairwise Variability Index (rPVI) is simply the mean of the differences between successive intervals:

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

Grabe and Low obtained normalised vocalic (nPVI-V) and raw consonantal (rPVI-C) scores for: stress-timed English, German and Dutch; syllable-timed French and Spanish; mora-timed Japanese; rhythmically indeterminate Polish and Catalan; rhythmically unclassified languages including Greek, Romanian and Estonian. Using one speaker per language reading a passage of text, they found higher nPVI-V scores for English, German and Dutch than for Spanish and French; rPVI-C scores did not clearly discriminate between the two hypothesised classes, however. Many unclassified or indeterminate languages had nPVI-V scores intermediate between stress-timed and syllable-timed, which Grabe and Low suggested indicates a weak

(footnote continued)

may show independent variation on these parameters. Grabe and Low (2002), for example, found that Polish and Estonian show similar patterns of vocalic interval variation but very different consonantal variation. Third, Ong, Deterding, and Low (2005) described some difficulties in applying the Variability Index (VI), such as identification of syllable boundaries and the treatment of insertions and deletions, which led to significant differences between measurers in VI scores for the same speech. Thus, for theoretical consistency and practical simplicity, we examined only acoustic measures in this study.

categorical distinction. Mora-timed Japanese did not occupy a distinct area of PVI space, in contrast with Ramus et al.'s clear separation of Japanese from stress-timed and syllable-timed languages on any pairwise IM mapping. Some languages were clearly distinguished by rPVI-C, with Polish and Estonian having extreme high and low values, respectively.

Grabe and Low compared PVI and IM scores for these speech samples, finding some commonalities and some clear differences, but conceded that conclusions were difficult to draw from results based on one speaker per language.

1.3. *Speech rate and rhythm class metrics*

One of the key differences between the cross-linguistic studies of Ramus et al. and Grabe and Low is in the treatment of speech rate. Ramus et al. attempted to control speech rate between languages by selecting utterances of similar duration and syllable number, whereas Grabe and Low normalised vocalic interval variation (nPVI). Ramus (2002) compared nPVI-V and rPVI-C with ΔV and ΔC on the Ramus et al. rate-controlled corpus and found that the PVI and IM plots produced similar pictures of rhythmic differences and commonalities. Observing that the application of IM to a corpus not controlled for speech rate was likely to produce results that reflect individual idiosyncrasy as much as language typology, he suggested that ΔV and ΔC could be normalised for speech rate by dividing by mean interval durations.

Support for the normalisation of these metrics came from Barry, Andreeva, Russo, Dimitrova and Kostadinova (2003), who found that both ΔC and ΔV were inversely related to speech rate (likewise Dellwo & Wagner, 2003, for ΔC). Dellwo (2006) thus utilised a rate-normalised metric, VarcoC, the standard deviation of consonantal interval duration divided by the mean consonantal duration (and multiplied by 100). Dellwo found that VarcoC produced clearer discrimination than ΔC at all rates between stress-timed English and German and syllable-timed French. VarcoC did vary according to intended rate, but the correlation did not appear systematic across languages. Interestingly, Dellwo and Wagner (2003) found little consistent correlation between %V and speech rate, suggesting that consonantal and vocalic intervals expand to comparable degrees at lower rates. Thus, rate normalisation may be unnecessary for %V.

1.4. *Utilisation of rhythm metrics: first and second language studies*

Since these rhythm metrics have been proposed, a number of studies have used them in various combinations and variations to address questions about rhythmic differences between or within languages. Comparative studies of rhythm between languages include: Gibbon and Gut (2001)—British English and Nigerian rhythm (modified nPVI-V, standard deviation of syllable duration); Dellwo and Wagner (2003)—English, French and German rhythm (ΔC , %V); Asu and Nolan (2005)—Estonian rhythm (PVI variants); Lin and Wang (2005)—Chinese and English rhythm (ΔC , %V). Some studies have considered rhythm between different varieties within a language: Barry et al. (2003) for Italian (ΔV , ΔC , %V, PVI variants); Ferragne and Pellegrino (2004) for British English (ΔV , ΔC , %V, VarcoV, VarcoC, PVI variants). Two studies have examined the relationship between speech rhythm and musical rhythm in English and French, as measured by nPVI-V (Patel & Daniele, 2003; Patel, Iversen, & Rosenberg, 2006). The range of proposed rhythm metrics and the methodological differences in studies that utilise them indicate the strong need for a direct and systematic comparison, as described below.

A number of studies have examined the influence of a speaker's first language on their production of second language rhythm, including: Low et al. (2000)—Singapore English L2 influenced by Chinese L1 (nPVI-V); Gut (2003)—German L2 influenced by Chinese, English, French, Italian and Romanian L1s (ΔC , %V, modified nPVI-V); Lin and Wang (2005)—Canadian English L2 influenced by Chinese L1 (ΔC , %V); Carter (2005)—American English L2 influenced by Mexican Spanish L1 (nPVI-V). Whitworth (2002) considered the rhythm of English–German bilingual children in both languages, as influenced by their parental languages, using nPVI-V and rPVI-C.

Studies of the influence of L1 on L2 production are intrinsically difficult to interpret. Lin and Wang (2005) found, for example, that advanced Chinese speakers of L2 English have ΔC scores that are not different from those of native English speakers, which suggests that these speakers, of an apparently syllable-timed language,

have adopted a native-like stress-timed rhythm for the L2. As discussed above, however, ΔC shows an inverse correlation with speech rate. Given that L2 speakers tend to speak more slowly than L1 speakers, the similarity in rhythm scores could be a reflection of relative rate rather than native-like competence. Unless rhythm metrics are appropriately calibrated, interpretation of such studies will remain problematic.

Relative rhythm scores for L1 and L2 will clearly be affected by a number of considerations, including the rhythmic properties of the native and target language and the degree of non-native accent of the L2 speaker. Carter (2005) tested English Spanish bilingual speakers who had moved from Mexico to North Carolina during childhood. He found that nPVI-V scores for L2 English were intermediate between the low nPVI-V scores for L1 Spanish and the high scores for L1 English, relating this result to the much lower incidence of vowel reduction in the English of native Spanish speakers, who lack significant vowel reduction in their L1. This clearly suggests that nPVI-V scores can be informative about the necessary process of accommodation between rhythmically distinct first and second languages.

1.5. Purpose of experiment

The purpose of this experiment was to evaluate the full range of acoustically-based rhythm metrics described above. The first part of the study evaluated the metrics by applying them all to a comparison of well-researched languages that exhibit distinct rhythmic characteristics: syllable-timed French and Spanish and stress-timed Dutch and English. We used new materials, uncontrolled for speech rate, and larger groups of speakers (six per group) than have hitherto been used in comparative studies of rhythm. As the results reported below show, the number of sentences and speakers were sufficient to provide statistically reliable differences between groups for the best-performing metrics. We examined the effect of speech rate on the original metrics (ΔV , ΔC , %V, nPVI-V, rPVI-C). We also utilised rate-normalised versions of IM (VarcoV and VarcoC).

In the second part of the study, we subjected the rhythm metrics found to be most discriminative in the first study to another test, by examining second language rhythm. It is possible that L2 speakers could achieve native-like rhythm and thus manifest rhythm scores like those of native speakers; equally, L2 speakers could fail to accommodate at all to the target language rhythm, thereby showing rhythm scores comparable to those of their L1. Because we chose speakers who were competent but with a perceptibly non-native accent, however, we adopted the working hypothesis that, where L1 and L2 are rhythmically different, rhythm metrics should show these speakers to have an L2 rhythm distinct both from their L1 and from the target language as spoken by native speakers. We also test the transfer of rhythm between similar languages, English and Dutch, with the prediction that L2 speakers should show rhythmic patterns similar to those of their native language.

2. Method

For the first language analyses, we recorded native English speakers (Eng_{Eng}) and Dutch speakers (Dut_{Dut}) as representative of stress-timed languages, and native Spanish (Sp_{Sp}) and French speakers (Fr_{Fr}) as representative of syllable-timed languages. For the first language vs second language analyses (L1/L2), we looked both within and between rhythm classes. Between classes, we recorded native Spanish speakers (Sp_{Sp}) and English speakers proficient in Spanish (Sp_{Eng}), all reading the same Spanish materials, and native English speakers (Eng_{Eng}) and Spanish speakers proficient in English (Eng_{Sp}), all reading the same English materials. Within classes, we recorded native Dutch speakers (Dut_{Dut}) and native English speakers proficient in Dutch (Dut_{Eng}), all reading the same Dutch materials, and native English speakers (Eng_{Eng}) and Dutch speakers proficient in English (Eng_{Dut}), all reading the same English materials.

2.1. Participants

Six speakers were recorded for each language condition (three female and three male speakers in each condition, except for Fr_{Fr} speakers—four female and two male—and Dut_{Eng} speakers—four male and two female). Both L1 and L2 speakers had accents of their native language that were not markedly different from the commonly accepted standard (i.e. Algemeen Nederlands, standard southern British English, français neutre, castellano). For the L1 recordings, all speakers were currently resident in the country of their native language.

For L2 speakers, our main hypothesis is that speakers with non-native accents should show intermediate rhythm scores between first and second language, so the primary selection criterion was for speakers to have some fluency but a discernible non-native accent. L2 speakers reported no extended exposure to their L2 in early childhood. Although all Dutch speakers reported familiarity with English through the media from an early age, none of the L2 speakers reported any formal training in their L2 before the age of 11 or later. To ensure a minimum level of non-native competence, we recorded only L2 speakers who had lived for an extended period in the country of their L2 (minimum five months). The recordings of L2 speakers who lacked an adequate degree of fluency were discarded, fluency being assessed, in part, by speakers' ability to describe a route around a map with minimum preparation.

2.2. Materials

Five sentences were recorded for each language condition (Appendix A). We attempted to achieve comparability of sentence length between languages, which represented a trade-off between equivalent numbers of syllables and equivalent total duration (given the differences in syllable complexity between languages). The mean number of syllables per sentence was: Spanish 19.4; French 19.6; English 16.2; Dutch 17.2. For the recordings of native speakers, the mean duration of the sentences in seconds was: French 3.6; Spanish 2.5; English 3.1; Dutch 2.9.

The distribution of stressed and unstressed syllables in the materials was not controlled. To construct sentences that were perfectly representative of stress distribution within a language would require data not uniformly available across languages and would seem a tendentious endeavour in any case. We opted rather for selection and construction of sentences that were intended as a semi-random sample, taking no explicit account of the rhythmic purpose of the experiment. English sentences were modified from a larger set reported in [Nazzi et al. \(1998\)](#); sentences for other languages were constructed by analogy with the English sentences, by speakers of these languages who were suggested a range of sentence lengths, but given no instructions regarding rhythm.

As the boundary between an approximant and a preceding or following vowel is difficult to determine reliably, the English sentences were constructed to exclude the approximants /l/, /r/, /w/ and /j/. The sentences for the other languages were then constructed along the same lines, although other allophonic approximants were not systematically excluded. Also, excluding all glide-vowel sequences proved intractable for French, so in the words *biscuits*, *mois* and *nuit*, the glide was taken as part of the vocalic interval, this being the only procedure that would facilitate consistency of segmentation.

2.3. Recordings

For most speakers, the sentence recordings were made at the end of a longer recording session, which included reading a short story, describing directions on a map and reading five other sentences. These other tasks were recorded for analysis in a separate study and their content and design had no connection with the experimental sentences. For the native French speakers and for a small number of others, speakers just read 10 sentences (i.e. the experimental set described above preceded by five others).

In all cases, speakers were given the sentences to read silently before reading them aloud. They were instructed to speak in their normal conversational voice at a rate that felt natural and comfortable. They were asked not to pause during sentences, but were requested to pause between successive sentences and to repeat a sentence if they made a mistake. The experimenter also occasionally requested repetition of sentences, in the interest of fluency, but speakers were not guided further about how to say the sentences. Recordings were made directly to disk, at a sampling rate of 16 kHz or higher.

2.4. Measurements

The first author identified and labelled the location of vowel–consonant and consonant–vowel boundaries, primarily by visual inspection of speech waveforms and wideband spectrograms in Praat ([Boersma & Weenink, 2006](#)). This procedure was carried out with reference to standard criteria (e.g. [Peterson & Lehiste,](#)

1960); where labels were associated with the start or end of pitch periods, they were placed at the point of zero crossing on the waveform.

The primary determiner of the placement of a vowel–consonant boundary was the end of the pitch period preceding a break in formant structure associated with a significant drop in waveform amplitude. Additional criteria which facilitated the location of the boundary in certain contexts included:

- Where the vowel offset was glottalised, a change in the shape of successive pitch periods, for example, lengthening or doubling.
- Before fricatives, the onset of visible frication.
- Before nasals, the appearance of nasal formant structure and a waveform amplitude minimum.

The consonant–vowel boundary was the beginning of the pitch period at the onset of vocalic formant structure, where this was associated with the appearance of pitch periods consistent with the body of the vowel (e.g. unfricated and of comparable amplitude). Aspiration following stop release was therefore included within the consonantal interval.

The durations of the vocalic and consonantal intervals were extracted using a Praat script. Following the approach taken by Grabe and Low (2002), silent pauses within sentences were excluded from the measured intervals and interval durations were summed across the pause. For consistency, we therefore excluded the silent interval in the few cases where a perceptible pause preceded a stop consonant release, summing the durations of the consonantal interval preceding the onset of silence and the interval from post-silence stop release to vowel onset. Where the silence was preceded by a vowel, the interval from the end of formant structure to the onset of complete silence was taken as consonantal—consistent with how it would be treated preceding a stop release—and summed with the post-silence consonantal interval. The duration of the silent interval itself was excluded. This was determined to be the most reliable and replicable method of dealing with the small number of pre-stop release pauses: stop consonant durations thus estimated were of the same order of magnitude as phrase-internal stops; including the silent interval would have greatly overestimated the true stop duration. Following Grabe and Low (2002), pre-pausal and utterance-final intervals were not excluded, despite the likelihood of lengthening effects (e.g. Klatt, 1976). First, phrase-final lengthening may occur in the absence of pausing and some pauses—for example, disfluent or hesitation pauses—may not be preceded by lengthening. Second, the locus and extent of final lengthening and other prosodic lengthening processes, such as accentual lengthening (e.g. Turk & White, 1999), may be language specific and may contribute to the overall perception of cross-linguistic differences in rhythmicity.

Intervals of glottalisation between successive vowels were treated as in the same manner as silent pauses. Glottalised intervals were identified by changes in the shape of pitch periods such as attenuation and lengthening/doubling (see Dilley, Shattuck-Hufnagel, & Ostendorf, 1996, for discussion of criteria for identification of glottalised intervals); these intervals were omitted and the intervals before and after the pitch period irregularity were summed. In order to achieve regularity between sentences, utterance-initial consonants, where present, were excluded from all analyses. As these metrics are acoustically based, vocalic intervals were only identified where there was evidence of a voiced vowel: syllabic consonants and devoiced or wholly elided vowels were treated as part of the adjacent consonantal interval.

2.5. Calculation of rhythm metrics

Rhythm metrics were calculated for each of the five sentences spoken by each of the six speakers for each language condition (L1 or L2). Thus, there were a total of 30 data points per language condition for each rhythm metric. The IM, all based on interval durations in milliseconds, were:

- ΔV , the standard deviation of vocalic interval duration.
- ΔC , the standard deviation of consonantal interval duration.
- %V, the sum of vocalic interval duration divided by the total duration of vocalic and consonantal intervals and multiplied by 100.

- VarcoV, the standard deviation of vocalic interval duration divided by the mean vocalic interval duration and multiplied by 100.
- VarcoC, the standard deviation of consonantal interval duration divided by the mean consonantal interval duration and multiplied by 100.

As outlined above, the PVI utilises the difference in duration of successive intervals, either vocalic or consonantal. The raw pairwise variability index (rPVI) is simply the mean of the differences between successive intervals; the normalised pairwise variability index (nPVI) is the mean of the differences between successive intervals divided by the sum of the same intervals. The PVI calculated here are:

- nPVI-V, the normalised Pairwise Variability Index for vocalic intervals.
- rPVI-C, the raw Pairwise Variability Index for consonantal intervals.

2.6. Statistical analysis

Clark (1973) advocated the use of the $\min F'$ statistic, derived from by-Subjects (F_1) and by-Items (F_2) analyses of variance, for experimental designs where items (i.e. individual sentences) were a nested factor within each condition (e.g. between languages), as is the case here. We therefore utilised $\min F'$ as the index of statistical reliability of differences. All post hoc tests are two-tailed Tukey HSD; p -level is only reported for significant (i.e. $p < 0.05$) or near-significant differences.

3. Analysis of rhythm metrics for first languages

Table 1 shows the mean scores and standard errors for all the rhythm metrics for the first language speakers: Sp_{Sp}, Fr_{Fr}, Eng_{Eng} and Dut_{Dut}. The different metrics are evaluated below in terms of how successfully they discriminate languages hypothesised to be rhythmically distinct and in terms of their consistency with previous studies, in particular, Ramus et al. (1999) for IM and Grabe and Low (2002) for PVI.

3.1. Interval measures

3.1.1. ΔV , ΔC and %V

For ΔV , there was a main effect of Language [$\min F'$ (3,32) = 5.08, $p < 0.01$]. Post hoc comparisons showed that the Sp_{Sp} score was significantly lower than for all the other languages [vs Dut_{Dut}, $p < 0.001$; vs Eng_{Eng}, $p < 0.001$; vs Fr_{Fr}, $p < 0.005$]. There were no significant differences between any other pairs of languages, even

Table 1
Means (standard errors) of rhythm metrics for Spanish, French, Dutch and English as first languages

	Spanish	French	English	Dutch
	Sp _{Sp}	Fr _{Fr}	Eng _{Eng}	Dut _{Dut}
<i>Interval measures</i>				
ΔV	32 (1.9)	44 (2.2)	49 (2.2)	49 (2.6)
ΔC	40 (2.3)	51 (3.6)	59 (2.4)	49 (4.1)
%V	48 (0.8)	45 (0.5)	38 (0.5)	41 (1.2)
VarcoV	41 (2.0)	50 (0.9)	64 (1.7)	65 (1.5)
VarcoC	46 (2.0)	44 (0.8)	47 (1.0)	44 (1.8)
<i>Pairwise variability indices</i>				
nPVI-V	36 (1.6)	50 (1.8)	73 (1.2)	82 (2.4)
rPVI-C	43 (2.1)	56 (4.3)	70 (2.8)	52 (4.2)
<i>Speech rate</i>				
Syllables/second	8.0 (0.3)	5.6 (0.3)	5.2 (0.2)	6.0 (0.3)

though the mean score for Fr_{Fr} was lower than for Dut_{Dut} or Eng_{Eng}, as shown in Table 1. Ramus et al. also found that French ΔV was higher than Spanish; however, as they only carried statistical tests of differences between presumed rhythm classes rather than between individual languages, direct comparisons of differences cannot be made. Despite our larger number of speakers per language, six compared to four in the Ramus et al. study, the differences between stress-timed English and Dutch and syllable-timed French were not significant.

The effect of Language approached significance for ΔC [$\min F'$ (3,35) = 2.75, $p = 0.057$]. The only significant post hoc comparison showed that ΔC was significantly lower for Sp_{Sp} than for Eng_{Eng} [$p < 0.005$]. The mean scores for Fr_{Fr} and Dut_{Dut} occupied an intermediate position between Sp_{Sp} and Eng_{Eng}, not differing statistically from either. The pattern for ΔC differs somewhat from Ramus et al., who found that ΔC was the measure with the smallest contrasts within rhythm classes and a clear distinction between rhythm classes. Speech rate was controlled in that study, however, by selecting rate-matched sentences. The effects of rate on IM are examined in the next section.

There was a main effect of Language for %V [$\min F'$ (3,27) = 6.64, $p < 0.05$]. The differences in scores were significantly different or approached significance for all pairwise language comparisons in post hoc tests. Sp_{Sp} had the highest %V score, significantly greater than Dut_{Dut} and Eng_{Eng} [$p < 0.001$ in both cases]. Fr_{Fr} was slightly lower than Sp_{Sp}, the difference approaching significance [$p = 0.085$]; Fr_{Fr} was significantly higher on %V than Dut_{Dut} or Eng_{Eng} [$p < 0.05$; $p < 0.001$, respectively]. Eng_{Eng} had the lowest score and the difference with Dut_{Dut} was almost significant [$p = 0.062$].

In terms of both replication of the results of Ramus et al. (1999) and discrimination between languages, %V appears the most consistent IM across studies. For both studies, English had the lowest %V score, followed by Dutch, and these were significantly lower than French/Spanish. In the present study, there was a trend towards Spanish having a significantly higher score than French (there was no evidence of this in Ramus et al.).

3.1.2. Effect of speech rate on ΔV , ΔC and %V

The absence of the expected discrimination, particularly between syllable-timed Fr_{Fr} and stress-timed Eng_{Eng} and Dut_{Dut}, in ΔV and ΔC , may partly be attributable to differences in speech rate. Mean speech rates for all first language conditions, measured in syllables per second, are shown in Table 1. The influence of speech rate on these measures is confirmed by the correlations between speech rate and the variance-based measures (Table 2).

All L1 conditions (Sp_{Sp}, Fr_{Fr}, Eng_{Eng}, Dut_{Dut}) showed significant, or near significant, inverse correlations between ΔV and speech rate, and significant inverse correlations between ΔC and speech rate. These correlations strongly encourage the use of the normalised variance measures, VarcoV and VarcoC, to compensate for speech rate variations. There were, however, no significant correlations between %V and speech rate for any L1 condition.

Table 2
Correlations between rhythm metrics and speech rate (syllables/second) for all first and second language conditions

	Spanish		French	English			Dutch	
	Sp _{Sp}	Sp _{Eng}	Fr _{Fr}	Eng _{Eng}	Eng _{Sp}	Eng _{Dut}	Dut _{Dut}	Dut _{Eng}
<i>Interval measures</i>								
ΔV	-0.466**	-0.647**	-0.425*	-0.394*	-0.710**	-0.374*	-0.324***	-0.137
ΔC	-0.557**	-0.579**	-0.430*	-0.497**	-0.666**	-0.488**	-0.583**	-0.639**
%V	0.241	-0.127	0.295	0.303	0.239	0.140	0.251	0.229
VarcoV	-0.307***	-0.204	-0.008	0.188	-0.393*	0.215	0.228	0.127
VarcoC	-0.085	-0.223	0.111	0.155	-0.183	0.274	-0.224	-0.235
<i>Pairwise variability indices</i>								
nPVI-V	-0.002	-0.187	-0.234	0.304	-0.413*	0.300	0.318***	0.267
rPVI-C	-0.383*	-0.558**	-0.352***	-0.540**	-0.605**	-0.554**	-0.636**	-0.618**

(* indicates $p < 0.05$; ** indicates $p < 0.01$; *** indicates $0.05 < p < 0.10$)

3.1.3. VarcoV and VarcoC

For VarcoV, there was a main effect of Language [$\min F' (3,23) = 9.89, p < 0.001$]. Post hoc tests showed significant differences in VarcoV scores between all languages except Eng_{Eng} and Dut_{Dut} (Table 1). Thus, both Eng_{Eng} and Dut_{Dut} had significantly higher VarcoV than both Fr_{Fr} [$p < 0.001$ in both cases] and Sp_{Sp} [$p < 0.001$ in both cases]. In addition, Fr_{Fr} had significantly higher VarcoV than Sp_{Sp} [$p < 0.005$].

We are not aware of any previous studies which utilise VarcoV to examine rhythmic differences between languages but, interestingly, the pattern that emerged was similar to that for ΔV , for both Ramus et al. and this study: Dutch and English had the highest scores, almost identical to each other, Spanish had the lowest score and French was intermediate. For VarcoV, however, French was closer to Spanish than to English/Dutch and the difference between the stress-timed languages and French was highly significant: thus VarcoV seems to capture the hypothesised typological difference better than ΔV , as well as indicating the possibility of differences within syllable-timed languages. The adequacy of the normalisation procedure for VarcoV is shown by the lack of significant correlations between VarcoV and speech rate for any of the first language conditions (Table 2), although there was a trend towards an inverse correlation for Spanish.

An ANOVA showed no effect of Language on VarcoC [$\min F' (3,26) = 0.19, n.s.$]. As shown in Table 1, there was little difference between mean scores for the different languages, and no suggestion of a systematic pattern reflecting rhythm classes. This is in contrast to the results obtained by Dellwo (2006), where at all speech rates, English and German had higher VarcoC values than French. At “normal” rate, however, which one must assume most comparable to the present case, the VarcoC scores were (to the nearest digit): German 62; English 53; French 46. Thus, for the two common languages, the results were not dramatically dissimilar (in the present study: English 47; French 44). In addition, Dellwo (2006) did not report statistical tests, so we cannot know if the difference between French and English at normal rates was reliable. There were no significant correlations between VarcoC and speech rate for any of the L1 conditions.

3.2. Pairwise variability indices

For nPVI-V, an ANOVA showed a main effect of Language [$\min F' (3,23) = 23.96, p < 0.001$]. Post hoc tests revealed that all languages significantly differed from each other: Dut_{Dut} had a higher score than Sp_{Sp}, Fr_{Fr} and Eng_{Eng} [$p < 0.001$; $p < 0.001$; $p < 0.01$, respectively], Eng_{Eng} had a higher score than Sp_{Sp} and Fr_{Fr} [$p < 0.001$ in both cases] and Fr_{Fr} had a higher score than Sp_{Sp} [$p < 0.001$]. The utility of the normalisation procedure in the nPVI metric was demonstrated by the lack of significant correlations between nPVI-V scores and speech rate for any language category (Table 2).

The comparison between the present study and Grabe and Low (2002) is instructive. For nPVI-V, the order of scores for the four languages was identical in the two studies, although the range was somewhat greater in the present study (36–82, vs about 30–65 for Grabe and Low). In both cases, there were large differences in scores within, as well as between, rhythm classes; in the present study, with six speakers per language rather than one, the differences between classes were clearly greater than within. Interestingly, the pattern for the two normalised vocalic measures, VarcoV and nPVI-V, was similar, the one difference being the discrimination between Dutch and English by nPVI-V, but not by VarcoV.

For rPVI-C, there was a main effect of Language [$\min F' (3,33) = 4.29, p < 0.05$]. Post hoc comparisons showed that the score for Eng_{Eng} was significantly higher than for all other languages [vs Dut_{Dut}, $p < 0.01$; vs Fr_{Fr}, $p < 0.05$; vs Sp_{Sp}, $p < 0.001$] and that the difference between Fr_{Fr} and Sp_{Sp} approached significance [$p = 0.078$]. There were no other significant differences in rPVI scores between languages. The lack of speech rate normalisation for rPVI scores was evidenced by the inverse correlations, significant or approaching significance, for all first languages between rPVI-C and speech rate (Table 2).

The parallels between PVI and IM are further evidenced by the pattern of scores for rPVI-C. The obvious comparison for this non-normalised metric is with ΔC : for both metrics, English had the highest score (concurring with expectations given its greater complexity in syllable onsets and codas than the Romance languages), and Spanish had the lowest score, with French and Dutch not discriminated in between. This pattern differs from that found by Grabe and Low for rPVI-C where there was little spread across languages, and Spanish and Dutch had the intermediate scores.

3.3. Discussion: rhythm metrics applied to first languages

In general, the pattern of results for the original IM replicated that of Ramus et al. (1999), certainly as regards the broad differences between rhythm classes. The pattern of significant differences between specific languages did not, however, concur with the hypothesised rhythm classes for ΔV and ΔC and the correlations between these metrics and speech rate strongly suggest the need for a normalisation procedure. In contrast, %V did not show a correlation with speech rate and clearly discriminates between rhythm classes. The normalised metrics, VarcoV and VarcoC, contrasted in their utility for first language discrimination: VarcoV discriminated between rhythm classes, and also within syllable-timed languages; VarcoC showed no linguistic discrimination at all, as though the normalisation removed all variability, echoing the concern of Grabe and Low for rPVI-C regarding the normalisation of the PVI measure for consonantal intervals.

Fig. 1 shows the scores for all first languages for %V plotted against VarcoV, indicating that both metrics provide support for a primary distinction between stress-timed Dutch and English and syllable-timed French and Spanish. In addition, both metrics, and %V in particular, also provide some support for the idea of gradient distinctions in rhythmicity. The difference between English and Dutch is consistent with the phonology of the two languages, as Dutch has been reported to have less widespread vowel reduction than English (Swan & Smith, 1987).

First language nPVI-V and rPVI-C scores are plotted in Fig. 2. The scores for rPVI-C do not clearly reflect hypothesised rhythm classes and are somewhat inconsistent with the scores of Grabe and Low. There are similarities between rPVI-C and ΔC scores, both showing French and Dutch as intermediate. It may be that French and Dutch are similar on consonantal interval variability, at least for these materials, but this does not concur with what has been proposed about the rhythm class differences in phonotactically allowable onsets and codas. The pattern may also be affected by differences in speech rate: as seen for VarcoC, however, proposed speech rate-normalisation procedures for these consonantal measures seem to eliminate much of the critical variation.

As mentioned above, the pattern of scores for nPVI-V mirrors to a large extent that found for VarcoV, with both metrics discriminating between and within rhythm classes. The following L2 studies assess further the relative utility of these measures.

4. Analysis of rhythm metrics for stress-timed vs syllable-timed L1 and L2 (English/Spanish)

Table 3 shows the means and standard errors for all the rhythm metrics in the English/Spanish L1/L2 comparison. Correlations between rhythm metrics and speech rate are shown in Table 2.

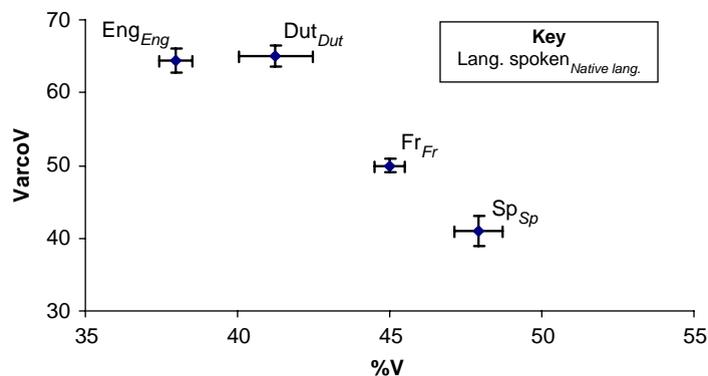


Fig. 1. Distribution of Spanish, French, English and Dutch as first languages over the %V, VarcoV plane. Bars represent one standard error around the mean.

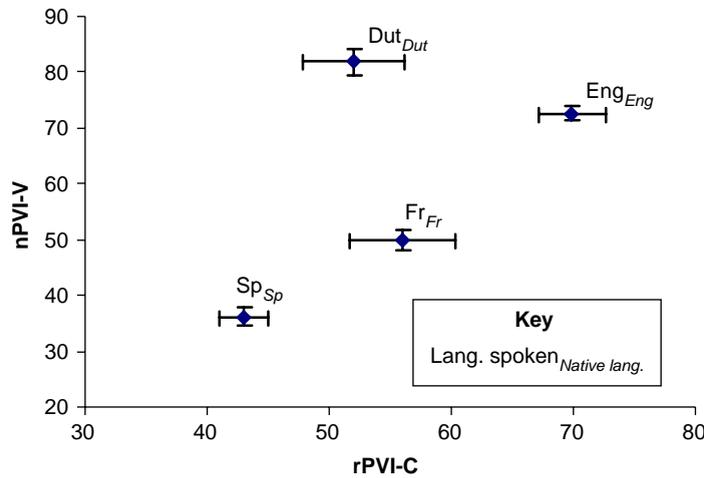


Fig. 2. Distribution of Spanish, French, English and Dutch as first languages over the rPVI-C, nPVI-V plane. Bars represent one standard error around the mean.

Table 3

Means (standard errors) of rhythm metrics for Spanish and English as first and second languages

	Language spoken: Spanish		Language spoken: English	
	Sp _{Sp}	Sp _{Eng}	Eng _{Eng}	Eng _{Sp}
<i>Interval measures</i>				
ΔV	32 (1.9)	51 (3.9)	49 (2.2)	47 (3.5)
ΔC	40 (2.3)	43 (2.2)	59 (2.4)	57 (3.6)
%V	48 (0.8)	52 (0.8)	38 (0.5)	41 (0.9)
VarcoV	41 (2.0)	52 (1.3)	64 (1.7)	54 (3.2)
Varco C	46 (2.0)	45 (1.7)	47 (1.0)	45 (1.9)
<i>Pairwise variability indices</i>				
nPVI-V	36 (1.6)	51 (2.4)	73 (1.2)	66 (4.3)
rPVI-C	43 (2.1)	46 (2.3)	70 (2.8)	65 (4.1)
<i>Speech rate</i>				
Syllables/second	8.0 (0.3)	6.6 (0.3)	5.2 (0.2)	4.8 (0.2)

4.1. Interval measures

4.1.1. ΔV , ΔC and %V

For ΔV , there was no effect of Language Spoken [$\min F'$ (1,14) = 1.42, n.s.], but there was an effect of Native Language [$\min F'$ (1,28) = 9.14, $p < 0.01$]. There was also a significant interaction between the two factors: [$\min F'$ (1,28) = 6.02, $p < 0.05$]. Post hoc comparisons showed that, for ΔV , Sp_{Sp} is significantly lower than Sp_{Eng}, Eng_{Eng} and Eng_{Sp} [$p = 0.001$; $p < 0.005$; $p < 0.01$, respectively]; none of the other groups differed significantly from each other.

For ΔC , there was a main effect of Language Spoken [$\min F'$ (1,13) = 8.60, $p < 0.05$], but no effect of Native Language [$\min F'$ (1,28) = 0.53, n.s.] and no interaction [$\min F'$ (1,16) = 0.25, n.s.]. Post hoc tests for ΔC confirmed no significant difference in ΔC between first and second language speakers for either Spanish or English, but did show differences between the two languages: both Eng_{Eng} and Eng_{Sp} had higher scores than both Sp_{Sp} and Sp_{Eng} [$p < 0.01$ for all comparisons]. As shown in Table 2, there were significant correlations between speech rate and ΔV , and speech rate and ΔC in all the Sp/Eng L1/L2 groups.

For %V, there was a main effect of Language Spoken [$\min F' (1, 12) = 38.20, p < 0.05$], but no effect of Native Language [$\min F' (1, 26) = 0.28, \text{n.s.}$]. There was a significant interaction between Language Spoken and Native Language [$\min F' (1, 26) = 10.69, p < 0.003$]. Post hoc tests showed that %V scores were significantly lower for Eng_{Eng} than Sp_{Sp} and Sp_{Eng} [both $p < 0.001$]. Eng_{Eng} %V scores were also lower than for Eng_{Sp}, the difference approaching significance [$p = 0.083$]. The %V score for Sp_{Sp} was significantly greater than for Eng_{Sp} and significantly lower than Sp_{Eng} [$p < 0.001$; $p < 0.05$, respectively]. There were no significant correlations between %V and speech rate (Table 2).

4.1.2. VarcoV and VarcoC

For VarcoV, there was a main effect of Language Spoken [$\min F' (1, 12) = 6.87, p < 0.05$] and a main effect of Native Language [$\min F' (1, 26) = 13.03, p < 0.005$], but no significant interaction between Language Spoken and Native Language [$\min F' (1, 23) < 0.001, \text{n.s.}$]. Post hoc comparisons showed that Eng_{Eng} had significantly higher VarcoV scores than all other groups [vs Sp_{Sp}, $p < 0.001$; vs Sp_{Eng}, $p < 0.005$; vs Eng_{Sp}, $p < 0.05$]. Sp_{Sp} had significantly lower scores than Sp_{Eng} and Eng_{Sp} [$p < 0.05$; $p < 0.005$, respectively]. Sp_{Eng} and Eng_{Sp} were not significantly different. There was slight evidence of a negative correlation between VarcoV and speech rate: as shown in Table 2, there was a significant negative correlation for Eng_{Sp} and a correlation approaching significance for Sp_{Sp}, but no correlations for the other two groups.

For VarcoC, there was no effect of Language Spoken [$\min F' (1, 11) = 0.01, \text{n.s.}$], no effect of Native Language [$\min F' (1, 24) = 0.04, \text{n.s.}$] and no interaction [$\min F' (1, 25) = 0.36, \text{n.s.}$]. There were no significant correlations between VarcoC and speech rate (Table 2).

4.2. Pairwise variability indices

For nPVI-V, there was a main effect of Language Spoken [$\min F' (1, 13) = 20.70, p < 0.001$] and a main effect of Native Language [$\min F' (1, 23) = 7.97, p < 0.01$]. There was no significant interaction between the two factors [$\min F' (1, 23) = 1.09, \text{n.s.}$]. Post hoc comparisons showed no significant difference between Eng_{Eng} and Eng_{Sp} scores; Sp_{Sp} scores were significantly lower than those for Sp_{Eng} [$p = 0.005$] and both groups' scores were significantly lower than Eng_{Eng} and Eng_{Sp} [$p < 0.005$ in all cases]. The results for nPVI-V were similar to those for VarcoV, showing influences on vocalic interval variability of the language spoken and the native language. Unlike on VarcoV, however, native and non-native speakers of English were not distinguished on nPVI-V scores.

For rPVI-C, there was a main effect of Language Spoken [$\min F' (1, 15) = 18.33, p < 0.001$] but no effect of Native Language [$\min F' (1, 28) = 1.14, \text{n.s.}$] and no significant interaction [$\min F' (1, 28) = 0.05, \text{n.s.}$]. This exactly mirrored the pattern found for ΔC , as did post hoc tests showing no differences between L1 and L2 speakers of either Spanish or English, but significant differences between languages: thus scores for both Sp_{Sp} and Sp_{Eng} were significantly lower than for both Eng_{Eng} and Eng_{Sp} [$p \leq 0.001$ in all cases].

4.3. Discussion: rhythm metrics applied to stress-timed vs syllable-timed L1 and L2

The results for ΔV did not accord with expectations. Both Spanish speakers of English and English speakers of Spanish appeared to realise vocalic interval variation like native English speakers. This does not concur with the subjective perception of these speakers' abilities in the second language: the speakers for both L2 groups were competent but with obvious non-native accents. L1 speakers tended to have faster speech rate than L2 speakers in each case: as seen for the analysis of first languages, there were significant negative correlations between speech rate and ΔV which encourage the use of the rate-normalised VarcoV.

For ΔC , there was no effect of native language, with scores for L1 and L2 speakers not differing for either English or Spanish. This would suggest that L1 and L2 speakers were producing the consonants in the texts similarly, supporting the claim that the L2 speakers in this study were reasonably competent and suggesting that consonantal metrics may be relatively uninformative about the processes of accommodation from L1 to L2. This conclusion is tentative, however, given the established correlations between ΔC and speech rate.

As already seen in Fig. 1, the combination of %V and VarcoV metrics appeared particularly useful for discriminating first languages, both between and within rhythm classes. Fig. 3 shows VarcoV plotted against

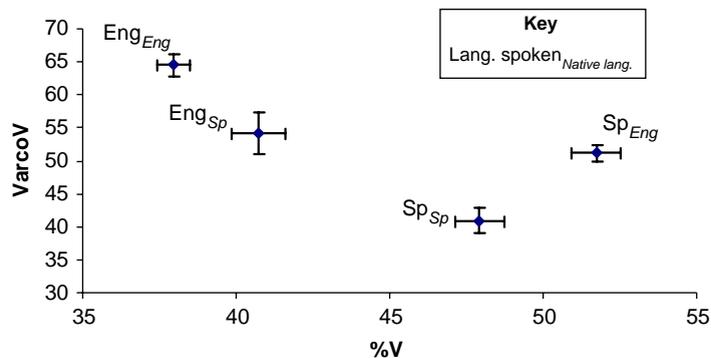


Fig. 3. Distribution of Spanish and English as first and second languages over the VarcoV, %V plane. Bars represent one standard error around the mean.

%V for the English/Spanish L1/L2 analysis. VarcoV was higher for Eng_{Eng} than for Sp_{Sp}, and within each language spoken, higher for native English speakers than for native Spanish speakers. VarcoV thus showed both sets of second language speakers occupying an intermediate position between their first language and the target language in terms of the realisation of vocalic intervals. This suggests that Eng_{Sp} speakers made unstressed vowels shorter than Sp_{Sp} speakers, but did not make the contrast between stressed and unstressed syllables as great as Eng_{Eng} speakers, a result echoing Carter's (2005) study of the English of American Hispanic bilinguals. Sp_{Eng} speakers appeared to be making less of a durational contrast between stressed and unstressed vowels than Eng_{Eng} speakers, but not attenuating this contrast as strongly as Sp_{Sp} speakers. These findings agree with the subjective perception of both sets of L2 speakers as being competent, but with a non-native accent.

Native Spanish speakers of English, Eng_{Sp}, were intermediate between their native language and their second language on %V, reinforcing the impression given by VarcoV that they were making unstressed vowels shorter than in Sp_{Sp} but not as short as in Eng_{Eng}. Native English speakers of Spanish, Sp_{Eng}, appear to have overshoot the linguistic target, however, with a higher %V even than Sp_{Sp} speakers.

This result appears surprising, as a working hypothesis is that second language speakers should be intermediate between L1 and L2. Given that %V reliably showed the expected discrimination in the other comparisons reported here, and that speech rate cannot account for the overshoot (there being no correlation between %V and speech rate), explanations may be sought in the segmental and suprasegmental timing differences between English and Spanish.

Specific processes which English native speakers may bring to Spanish as a second language include: at the segmental level, diphthongal realisation of Spanish monophthongs; at the suprasegmental level, greater lengthening of vowels in accented syllables and in phrase-final syllables. An example phrase, *su coche*, from one of the Spanish sentences spoken in this experiment: *A mí no me gustaba su coche pequeño y viejo*, may serve as an illustration. The VarcoV scores suggest that Sp_{Eng} speakers have, in general, greater durational contrast than Sp_{Sp} speakers between unstressed vowels like the initial and final vowels in *su coche* and stressed vowels like the initial vowel of *coche*, but there are a number of additional processes which may affect the durational balance. First, the word-final vowel in *coche*, the monophthong [e] in Spanish, may be realised with some quality of the diphthong [ei] by native English speakers, which would tend to make it longer than in Spanish. Second, the word *coche* may carry a pitch accent in this sentence. One of consequences of pitch accent is lengthening of segments within the accented word (e.g. Turk & White, 1999). Evidence suggests that there is less lengthening associated with pitch accent in Spanish than in English (Ortega-Llebaria & Prieto, 2007), but English speakers of Spanish may well retain their native pattern of lengthening. Likewise, some speakers may realise a phrase break following *su coche*, another prosodic feature associated with greater lengthening (phrase-finally) in English than in Spanish (Delattre, 1966). Typically, phrase-final lengthening begins with the final stressed vowel and continues to the phrase break. As a high proportion of syllables are CV in Spanish, such prosodic lengthening is likely to disproportionately affect vowels, thus increasing the overall %V score. Both unstressed and stressed vowels are likely to be affected by accentual lengthening and final lengthening

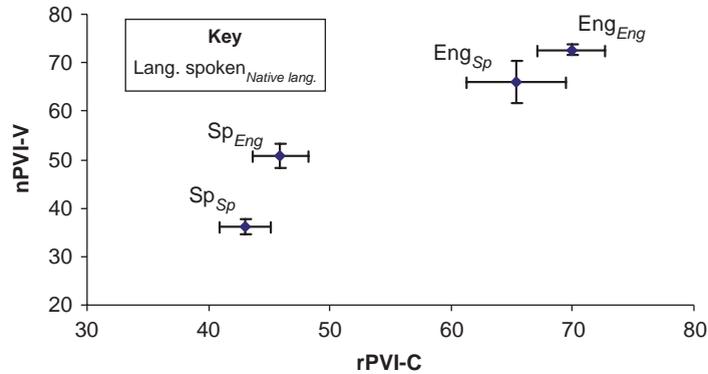


Fig. 4. Distribution of Spanish and English as first and second languages over the rPVI-C, nPVI-V plane. Bars represent one standard error around the mean.

(except in monophthongs or, for final lengthening, in cases of word-final stress), so the overall durational balance between syllables may be less affected, and VarcoV would be not raised as greatly as if only stressed syllables were subject to prosodic lengthening effects.

Fig. 4 shows nPVI-V plotted against rPVI-C for English/Spanish L1/L2 speakers. Once again, the pattern for rPVI-C mirrored that for ΔC , with the same caveats about its interpretation regarding the dependence of the metrics on speech rate. The pattern for nPVI-V was likewise similar to that for VarcoV; post hoc tests do not discriminate the nPVI-V scores for Eng_{Eng} and Eng_{Sp}, however. Given the clear non-native accents of the Eng_{Sp} speakers, the VarcoV metric which does distinguish Eng_{Eng} and Eng_{Sp}, appears to be somewhat more effective in this case. The third set of analyses—between first and second languages speakers of two rhythmically-similar languages—address this issue further.

5. Analysis of rhythm metrics for stress-timed vs stress-timed L1 and L2 (English/Dutch)

Table 4 shows the means and standard errors for all the rhythm metrics in the Dutch vs English L1 vs L2 comparison. Correlations between IM and speech rate are shown in Table 2.

5.1. Interval measures

5.1.1. ΔV , ΔC and %V

The ΔV scores showed no effect of Language Spoken [$\text{min}F'$ (1,14) = 0.10, n.s.], no effect of Native Language [$\text{min}F'$ (1,28) = 0.36, n.s.] and no interaction [$\text{min}F'$ (1,28) = 0.10, n.s.]. As before, there was evidence of a correlation between speech rate and ΔV for most groups (excluding Dut_{Eng}, see Table 2).

For ΔC , there was no effect of Language Spoken [$\text{min}F'$ (1,13) = 0.05, n.s.], but there was an effect of Native Language [$\text{min}F'$ (1,27) = 5.68, $p < 0.05$]. There was no interaction between Language Spoken and Native Language [$\text{min}F'$ (1,27) = 0.49, n.s.]. Thus, native English speakers showed greater variability in consonantal intervals than did native Dutch speakers. Post hoc tests did not approach significance for any of the comparisons between groups, however. As before, all groups showed strong inverse correlations between speech rate and ΔC (Table 2), with all groups apart from Dut_{Dut} having very similar speech rates.

For %V, there was no effect of Language Spoken [$\text{min}F'$ (1,12) = 0.05, n.s.], but there was a significant main effect of Native Language [$\text{min}F'$ (1,28) = 5.69, $p < 0.02$]. There was no significant interaction between Language Spoken and Native Language [$\text{min}F'$ (1,28) = 0.52, n.s.]. The %V difference between Dut_{Dut} and Dut_{Eng} was the only one that approached significance in post hoc tests [$p = 0.093$]. Once again, %V showed no effect of speech rate (Table 2).

Table 4
Means (standard errors) of rhythm metrics for Dutch and English as first and second languages

	Language spoken: Dutch		Language spoken: English	
	Dut _{Dut}	Dut _{Eng}	Eng _{Eng}	Eng _{Dut}
<i>Interval measures</i>				
ΔV	49 (2.6)	52 (3.4)	49 (2.2)	48 (2.2)
ΔC	49 (4.1)	60 (3.1)	59 (2.4)	53 (2.6)
%V	41 (1.2)	38 (1.6)	38 (0.5)	40 (0.4)
VarcoV	65 (1.5)	65 (1.7)	64 (1.7)	61 (2.7)
Varco C	44 (1.8)	45 (0.8)	47 (1.0)	45 (2.1)
<i>Pairwise variability indices</i>				
nPVI-V	82 (2.4)	75 (1.6)	73 (1.2)	70 (1.6)
rPVI-C	52 (4.2)	61 (3.9)	70 (2.8)	62 (3.5)
<i>Speech rate</i>				
Syllables/second	6.0 (0.3)	5.1 (0.2)	5.2 (0.2)	5.2 (0.3)

5.1.2. VarcoV and VarcoC

For VarcoV, there was no effect of Language Spoken [$\min F'$ (1,11) = 0.18, n.s.], no effect of Native Language [$\min F'$ (1,25) = 0.40, n.s.] and no significant interaction [$\min F'$ (1,25) = 0.28, n.s.]. For VarcoC, there was likewise no effect of Language Spoken [$\min F'$ (1,11) = 0.09, n.s.], no effect of Native Language [$\min F'$ (1,25) = 0.53, n.s.] and no significant interaction [$\min F'$ (1,25) = 0.17, n.s.]. There was no evidence of any correlations between VarcoV and speech rate, or VarcoC and speech rate (Table 2).

5.2. Pairwise variability indices

For nPVI-V, there was no effect of Language Spoken [$\min F'$ (1,11) = 2.14, n.s.], no effect of Native Language [$\min F'$ (1,16) = 0.55, n.s.] and no significant interaction [$\min F'$ (1,16) = 2.39, n.s.]. These results mirror those for VarcoV, indicating the rhythmic similarity of Dutch and English, when spoken either by natives or by native speakers of the other language. Once again, there was little evidence of correlations between nPVI-V and speech rate (Table 2).

For rPVI-C, there was no effect of Language Spoken [$\min F'$ (1,13) = 1.59, n.s.], but the effect of Native Language approached significance [$\min F'$ (1,28) = 4.12, $p = 0.052$]. There was no significant interaction between the two factors [$\min F'$ (1,28) = 0.01, n.s.]. While these results apparently mirror those for ΔC , post hoc tests suggested that this Native Language tendency arose largely from the greater rPVI-C scores for Eng_{Eng} than for Dut_{Dut} [$p < 0.05$]. No other differences between groups were significant in post hoc tests. There were strong inverse correlations between rPVI-C and speech rate (Table 2), which could, at least in part, underpin the differences in nPVI-C scores between Eng_{Eng} and Dut_{Dut} (the latter having higher speech rate and lower rPVI-C).

5.3. Discussion

As in the previous analyses, the metrics for consonantal interval variability presented an equivocal picture. The rate-normalised metric, VarcoC seemed once again to eliminate not only the variability due to rate variation but also that due to cross-linguistic differences in onset or coda phonotactics. The non-normalised measures were slightly more suggestive of the nature of the accommodation process to the second language, with both ΔC and rPVI-C showing evidence of an influence of first language only. This could be taken as evidence that speakers of rhythmically similar languages do not accommodate between the L1 and L2, preferring to realise the latter like the former, at least in terms of consonantal interval variability. This conclusion is relatively well supported by the pattern of mean scores for ΔC , but not for rPVI-C; the influence of speech rate on both of these metrics confounds further interpretation.

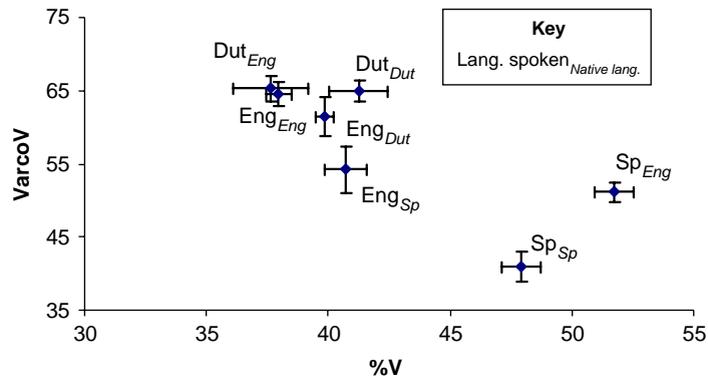


Fig. 5. Distribution of Dutch and English as first and second languages over the VarcoV, %V plane. The data for Spanish and English L1/L2 speakers are shown for reference. Bars represent one standard error around the mean.

None of the measures of vocalic interval variability, whether normalised or not, showed any effect of language spoken or native language (although the pattern of individual scores was much less consistent for nPVI-V than for either ΔV or VarcoV), reinforcing the idea that Dutch and English are rhythmically similar. Further, this picture serves to aid interpretation of the previous second language results. The lack of VarcoV differences between Eng_{Dut} and Eng_{Eng}, for example, supports the view that the greater syllable-timing of Eng_{Sp} than Eng_{Eng} speakers does not emerge simply from the fact of speaking an L2 rather than an L1, but is, at least in part, a function of L1 rhythm.

Fig. 5 shows VarcoV plotted against %V for the English/Dutch L1/L2 analysis (the English/Spanish data are shown for comparison). The effect of native language on %V can be seen here; as reported in the L1 analysis, Dut_{Dut} speakers have slightly higher %V than do Eng_{Eng} speakers. This trend appears also to influence second language production, so that Eng_{Dut} had a %V near to that of Dut_{Dut}, and Dut_{Eng} had a %V near to that of Eng_{Eng}. These effects were small, however, suggesting that the subtle accommodations required between an L1 and a rhythmically similar L2 may not be undertaken by most speakers. It may be communicatively sufficient to realise the timing of vocalic and consonantal intervals just as in the L1, a strategy that would be sub-optimal when moving between rhythmically distinct languages, such as Spanish and English.

6. General discussion

6.1. Evaluation of rhythm metrics

The first test for the rhythm metrics in this study was discrimination between stress-timed languages such as English and Dutch and syllable-timed languages such as French and Spanish. The metrics %V, VarcoV and nPVI-V all showed scores that supported this distinction, whilst also showing evidence of differences within rhythm classes almost as large as those between classes.

Non-rate-normalised measures of interval variation did not clearly discriminate between rhythm classes: French was not distinguished from Dutch and English by ΔV scores; French and Dutch appeared intermediate between English and Spanish in terms of ΔC and rPVI-C, unadjusted for speech rate. One reason for the lack of expected discrimination is likely to be the inverse correlation between speech rate and scores for ΔV , ΔC and rPVI-C.

The normalisation procedure of dividing the standard deviation by the mean worked well for the vocalic IM, VarcoV, as described above, producing a similar pattern of results to that of the normalised nPVI-V. The rate-normalised VarcoC showed almost no language-related variation at all, however.

We then tested whether the rhythm metrics that proved most discriminant in distinguishing between rhythm classes were also successful in distinguishing first and second language rhythm. For the English/Spanish comparison, VarcoV scores for L2 speakers—Eng_{Sp} and Sp_{Eng}—showed the expected pattern of

discrimination from both the speaker's native language and the target language. This suggests one reason for slightly favouring VarcoV over nPVI-V, which did not distinguish Eng_{Sp} from Eng_{Eng} speakers, despite the subjective perception of non-native accent for the latter.

The %V metric passed the primary test of distinguishing L1 and L2 rhythm: Eng_{Sp} speakers had intermediate %V scores; somewhat unexpectedly, Sp_{Eng} speakers had %V scores even higher than those for native Spanish speakers. As %V clearly showed the expected discrimination between categories in the rest of study, and is not subject to variation with speech rate, this result is unlikely to be anomalous. The rhythmic overshoot therefore demands an assessment of the relative timing properties of Spanish and English and the nature of the required linguistic accommodation for Sp_{Eng}, as discussed above.

As the first part of the study suggests, scores for the non-rate-normalised metrics should be interpreted with caution. ΔV scores suggest that Eng_{Sp} achieved native-like variability in vocalic intervals, whereas Sp_{Eng} did not accommodate at all to L2 rhythm, neither result according with the subjective perception of these speakers' abilities. The consonantal metrics, ΔC and rPVI-C, showed a strong language dependence for the English/Spanish L1/L2 comparison, regardless of speakers' linguistic origins.

For the English/Dutch comparison, VarcoV pointed, as expected, to the essential rhythmic similarity between the two languages. Although neither VarcoV nor nPVI-V scores showed an effect of either target language or native language, the scores for the four groups—Dut_{Dut}, Dut_{Eng}, Eng_{Eng}, Eng_{Dut}—were much more internally homogenous for VarcoV than for nPVI-V, suggesting another slight reason for favouring VarcoV over nPVI-V.

The %V scores for L1 Dutch and English did show a difference, with Dutch having a higher %V, and this difference was reflected somewhat in the L2 scores, suggesting that most L2 speakers make little rhythmic accommodation if their L1 and L2 are rhythmically similar. The results for consonantal intervals reinforced this idea of rhythmic inertia in the face of subtle distinctions. As before, however, speech rate caveats apply to the interpretation of measures such as ΔC and rPVI-C.

Previous studies have met with mixed results for consonantal metrics: Ramus et al. (1999) found ΔC to distinguish rhythm classes in rate-controlled materials, but the expected distinctions were not observed by Grabe and Low (2002). Both these metrics are highly correlated with speech rate, but the straightforward rate normalisation in VarcoC appears to eliminate almost all the critical linguistic variability. In addition to the obvious need for an effective but not over-effective speech rate control, more investigation is required regarding the dependence of consonantal interval metrics on the specific materials involved.

Patterns of L1 scores found here were broadly similar to those found in other studies, particularly for vocalic measures. Absolute scores differ substantially between studies, however, in part due to variable methodologies used in dividing the speech signal into vocalic and consonantal intervals. Some of these methodological issues are tractable, such as whether to include phrase-final intervals or not, and whether to treat successive vowels as a single interval or as discrete intervals, but caution should be exercised in making comparisons of rhythm scores between studies.

6.2. Interpretation of rhythm metrics

As discussed above, differences in rhythm scores between languages arise from language-specific variations in syllable construction and distribution, as well as from segmental timing processes. The rhythmic distinctions suggested by these metrics are therefore emergent phenomena; an alternative view of rhythm, as a product of top-down control of the timing of successive syllables or successive stresses, is discussed below. Language-specific prosodic timing processes, such as accentual lengthening, word-initial lengthening and phrase-final lengthening, should clearly also be considered in a full model of influences on vocalic and consonantal interval durations. How far rhythm class distinctions correlate with differences in prosodic timing processes is an open question. As discussed above, Spanish is marked by lesser degrees of accentual lengthening and final lengthening than English, suggesting that strong durational marking of both stressed syllables and prosodic structure may tend to co-occur in languages.

The aggregation of these timing factors contributes to the emergence of rhythm classes based upon the relative strength of successive syllables. In stress-timed languages, stressed syllables have much greater strength and it is their repetition that primarily conditions the perception of rhythm—"Morse code" rhythm

in the terms of Lloyd James (1940). In syllable-timed languages, stressed syllables are not greatly stronger in acoustic terms than unstressed syllables and rhythm arises from the repetition of syllables *per se*—“machine gun” rhythm. As syllabic strength is clearly gradient and a product of a variety of more or less independent factors, distinctions between rhythm classes should not be expected to be categorical: PVI data for a range of languages show clear evidence of middle ground between canonical stress-timed and syllable-timed languages (Grabe & Low, 2002).

It is not necessarily possible to extrapolate from relative acoustic strength of stressed syllables to their relative salience to native speakers. Arvaniti (1994) argued against an account of rhythm classes in terms of the relative salience of stressed and unstressed syllables, observing that in a language like Greek, with predominantly CV syllables and no phonemic vowel length distinction, stressed syllables are highly salient to native speakers, who are consequently sensitive to misaccentuations. This is clearly true of some other syllable-timed languages: although French speakers, for instance, may exhibit “stress deafness” when perceiving other languages (Peperkamp & Dupoux, 2002), the difference between stressed and unstressed syllables is crucial in the perception of Spanish or Italian (e.g. the Spanish minimal pair *hablo*—*I speak*—vs *hab lo*—*she spoke*).

Given that rhythm metrics such as VarcoV clearly show that Spanish has less variation in vocalic interval duration than English, what account can we give of the fact that Spanish stressed syllables are perceptible as such to Spanish listeners? First, non-durational cues such as intonation and amplitude variation also contribute to stress perception. Second, categorical perception can, of course, apply equally to acoustically subtle and gross distinctions where they are meaningful within languages.

There are consequences in terms of rhythmic organisation for the greater relative strength of stressed syllables in stress-timed languages. As Dauer (1983) noted, English, with its high contrast between stressed and unstressed syllables, has secondary lexical stresses, stress shift and the possibility of stress insertion, which all conspire to prevent long sequences of unstressed syllables. Spanish, with low contrast between stressed and unstressed syllables, may, however, have relatively long sequences of unstressed syllables. Thus, there is a correlation between syllable strength and stress distribution, and rhythm metrics scores—which provide direct information about the former—should allow predictions to be made about the latter.

The quality of direct information that rhythm metrics provide about stress distribution is minimal, however. Although the PVI was designed to quantify sequential variation, there is a limit to how much information a single parameter can convey about the organisation of a complex heterogeneous sequence. In addition, there are some, perhaps linguistically implausible, cases where PVI could actually fail to discriminate very different series. As Gibbon (2003) noted, the PVI scores derived from alternating patterns and monotonic geometric series may be the same, so that, for example, PVI(2,4,2,4,2,4) and PVI(2,4,8,16,32,64) are equal. The latter sequence is highly unlikely in natural language, of course, but is suggestive of a formal problem with the PVI that the simpler interval measures do not present.

Clearly, a full quantification of rhythm should take account of both relative syllable strength and syllable distribution. Insofar as rhythm metrics only provide information about the former, they are a necessarily incomplete, though informative, account of speech rhythm.

Finally, it is useful to consider how the results reported in this study bear upon our understanding of the concept of speech rhythm. After all, stress-timing and syllable-timing were originally conjectured as syntagmatic descriptions, relating to the temporal arrangement of successive syllables. The empirically discredited notion of the strict isochronous occurrence of stressed syllables or syllables *per se* necessarily required planning of speech timing throughout the utterance, implying some form of top-down control of speech segment duration to regularise the language-specific rhythmic intervals. Acknowledging the lack of strict isochrony in English, but looking nonetheless for evidence of top-down control in rhythmic timing—i.e. some planning of the temporal arrangement of stressed and unstressed syllables—Cummins and Port (1998) found that English speakers could produce a sequence of stressed syllables entrained to a metronome-governed cycle. The fact that isochronous performance can be induced experimentally need not imply that this reflects an underlying principle of timing, of course. Cummins (2002) further showed that speakers of Spanish and Italian did not readily exhibit such entrainment of stressed syllables and suggested that it was the lack of the appropriate level—the foot—in a rhythmic timing hierarchy that prevented such speakers from completing the task. Appeal to a concept of globally-planned rhythmic timing may not be

necessary to account for such results, however: the strength of successive Spanish syllables is much less variable than in English, so it may simply be that there is no strong attraction of stressed syllables towards the regular beat.

7. Summary

We have evaluated various timing metrics that have been suggested for quantifying rhythmic differences between and within languages. The comparison with previous findings showed that results for measures of vocalic interval variance are relatively reproducible, while measures of consonantal variance showed variability in scores which suggested they may be influenced by the nature of the linguistic materials used. Both types of metrics showed an inverse correlation with speech rate: for vocalic intervals, rate-normalised metrics such as VarcoV and nPVI-V appeared more reliable and discriminative than raw metrics; for consonantal intervals, rate normalisation appeared to eliminate linguistically-interesting variation. A simple measure of the relative proportion of vocalic and consonantal intervals, %V, was robust to changes in speech rate and was discriminative between hypothesised rhythm classes.

The combination of %V and VarcoV appeared particularly complementary, reinforcing the concept of a rhythmic typology of languages, while challenging the notion of a strictly categorical distinction by showing that variation within hypothesised rhythmic classes is sometimes as great as between. This combination also offers the possibility of providing insights into less-studied aspects of speech rhythm, such as the influence of first language on second language rhythm.

Rhythm is primarily a perceptual property, however, and more work is required to ascertain how these metrics relate to the subjective experience of linguistic rhythm. Some studies have begun this process: Ramus et al. (2003) showed, for example, that rhythmic distinctions suggested by measures of vocalic interval variability were supported by perceptual discrimination tasks; White and Mattys (2007) showed VarcoV to be a strong predictor of the rating of a Spanish speaker's English accent as native or non-native. The possible link between rhythmic classification and the realisation of prosodic timing processes is also worthy of further cross-linguistic investigation. The most effective of the rhythm metrics investigated here allow the possibility of tackling questions, such as these, that might formerly have seemed intractable.

Acknowledgements

This research was supported by a grant from the Biotechnology and Biological Sciences Research Council (BBSRC) to Sven Mattys (7/S18783). We thank Sarah Davies, Juan Manuel Toro, Casimier Ludwig and Ineke Mennen for help with stimulus preparation and Elizabeth Johnson, Reinier Salverda, Astrid Schepman, Mike Sharwood Smith, Juan Manuel Toro, Isabelle Viaud-Delmon, Atie Vogelenzang de Jong and Eric-Jan Wagenmakers for help with recordings. We are grateful to three anonymous reviewers for their very helpful comments on an earlier draft of this paper and to Amalia Arvaniti and Laura Dilley for interesting discussions of some of the issues raised here.

Appendix A. Sentence materials

Spanish.

A mí no me gustaba su coche pequeño y viejo.
Vicente y Susana van de vacaciones este mes a Escocia.
A pocos pasos de mi casa está una tienda bonita.
Un chico me dijo hace poco que no había pasado nada.
Pienso que todo va bien con mis tíos estas Navidades.

French.

Des fonds d'assistance ont été donnés à deux banques au sud des zones dominantes.
Cette fameuse échoppe vend de bons petits pains et d'appétissants biscuits secs.
Neuf mois ont passé depuis cet événement choquant qui fit tant de dégâts.

Enquêtes et sondages ne sont que des outils de vague estimation.
 Une nuit d'été au Mexique est assez douce mais peu de gens aiment son niveau d'humidité.

English.

The supermarket chain shut down because of poor management.
 Much more money must be donated to make this department succeed.
 In this famous coffee shop they serve the best doughnuts in town
 The chairman decided to pave over the shopping centre garden.
 The standards committee met this afternoon in an open meeting.

Dutch.

Tegen het einde van mei gaan ze met hun oom en tante op vakantie.
 In deze zaak verkopen ze de beste biefstuk van de stad.
 De ondeugd had met een stift gekke gezichten op het behang getekend.
 Hij had de kans om aan een tocht van zes maanden mee te doen.
 Ze kwam pas tegen de ochtend thuis en was zó moe dat ze in bed dook.

References

- Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh: Edinburgh University Press.
- Arvaniti, A. (1994). Acoustic features of Greek rhythmic structure. *Journal of Phonetics*, 22, 239–268.
- Asu, E. L. & Nolan, F. (2005). Estonian rhythm and the Pairwise Variability Index. In *Proceedings of Fonetik 2005* (pp. 29–32). Gothenburg.
- Barry, W. J., Andreeva, B., Russo, M., Dimitrova, S., & Kostadinova, T. (2003). Do rhythm measures tell us anything about language type? In *Proceedings of the 15th international congress of phonetics sciences* (pp. 2693–2696). Barcelona.
- Boersma, P., & Weenink, D. (2006). *Praat: Doing phonetics by computer* (version 4.3.04) [computer program]. Retrieved 21st March 2006, from <<http://www.praat.org/>>.
- Carter, P. M. (2005). Quantifying rhythmic differences between Spanish, English, and Hispanic English. In R. S. Gess, & E. J. Rubin (Eds.), *Theoretical and experimental approaches to romance linguistics: Selected papers from the 34th linguistic symposium on romance languages (Current issues in linguistic theory 272)* (pp. 63–75). Amsterdam, Philadelphia: John Benjamins.
- Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behaviour*, 12, 335–359.
- Cummins, F. (2002). Speech rhythm and rhythmic taxonomy. *Proceedings of prosody 2002* (pp. 121–126). Aix en Provence.
- Cummings, F., & Port, R. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics*, 26, 145–171.
- Dasher, R., & Bolinger, D. (1982). On pre-accentual lengthening. *Journal of the International Phonetic Association*, 12, 58–69.
- Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, 11, 51–62.
- Dauer, R. M. (1987). Phonetic and phonological components of language rhythm. In *Proceedings of the 11th international congress of phonetic sciences* (pp. 447–450). Tallinn.
- Delattre, P. (1966). A comparison of syllable length conditioning among languages. *International Review of Applied Linguistics in Language Teaching*, 4, 183–198.
- Dellwo, V. (2006). Rhythm and speech rate: A variation coefficient for deltaC. In P. Karnowski, & I. Szigeti (Eds.), *Language and language processing: Proceedings of the 38th linguistic colloquium* (pp. 231–241). Piliscsaba 2003. Frankfurt: Peter Lang.
- Dellwo, V., & Wagner, P. (2003). Relations between language rhythm and speech rate. In *Proceedings of the 15th international congress of phonetics sciences* (pp. 471–474). Barcelona.
- Deterding, D. (2001). The measurement of rhythm: A comparison of Singapore and British English. *Journal of Phonetics*, 29, 217–230.
- Dilley, L., Shattuck-Hufnagel, S., & Ostendorf, M. (1996). Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics*, 24, 423–444.
- Ferragne, E. & Pellegrino, F. (2004). A comparative account of the suprasegmental and rhythmic features of British English dialects. In *Proceedings of "Modelisations pour l'Identification des Langues"*. Paris.
- Gibbon, D. (2003). Computational modelling of rhythm as alternation, iteration and hierarchy. In *Proceedings of the 15th international congress of phonetic sciences* (pp. 2489–2492). Barcelona.
- Gibbon, D., & Gut, U. (2001). Measuring speech rhythm. In *Proceedings of Eurospeech* (pp. 91–94). Aalborg.
- Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In N. Warner, & C. Gussenhoven (Eds.), *Papers in laboratory phonology 7* (pp. 515–546). Berlin: Mouton de Gruyter.
- Gut, U. (2003). Prosody in second language speech production: The role of the native language. *Fremdsprachen Lehren und Lernen*, 32, 133–152.
- Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, 59, 1208–1220.

- Lin, H., & Wang, Q. (2005) Vowel quantity and consonant variance: A comparison between Chinese and English. In *Proceedings of between stress and tone*. Leiden, June 2005.
- Lloyd James, A. (1940). *Speech signals in telephony*. London: Pitman & Sons.
- Low, E. L., Grabe, E., & Nolan, F. (2000). Quantitative characterisations of speech rhythm: ‘Syllable-timing’ in Singapore English. *Language and Speech*, 43, 377–401.
- 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.
- Ong, F., Deterding, D., & Low, E. L. (2005). Rhythm in Singapore and British English: A comparative study of indexes. In D. Deterding, A. Brown, & E. L. Low (Eds.), *English in Singapore: Phonetic research on a corpus* (pp. 74–85). Singapore: McGraw-Hill Education (Asia).
- Ortega-Llebaria, M., & Prieto, P. (2007). Stress and focus in Spanish and Catalan: Patterns of duration and vowel quality. In P. Prieto, J. Mascaró, & M.-J. Solé (Eds.), *Segmental and prosodic issues in romance phonology (Current issues in linguistic theory series)*. Amsterdam, Philadelphia: John Benjamins.
- Patel, A. D., & Daniele, J. R. (2003). An empirical comparison of rhythm in language and music. *Cognition*, 87, B35–B45.
- Patel, A. D., Iversen, J. R., & Rosenberg, J. C. (2006). Comparing the rhythm and melody of speech and music: The case of British English and French. *Journal of the Acoustical Society of America*, 119, 3034–3047.
- Peperkamp, S., & Dupoux, E. (2002). A typological study of stress ‘deafness’. In C. Gussenhoven & N. Warner (Eds.), *Papers in laboratory phonology 7* (pp. 203–240). Berlin: Mouton de Gruyter.
- Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America*, 32, 693–703.
- Pike, K. (1945). *The intonation of American English*. Ann Arbor: University of Michigan Press.
- Port, R., Dalby, J., & O’Dell, M. (1987). Evidence for mora timing in Japanese. *Journal of the Acoustical Society of America*, 81, 1574–1585.
- Ramus, F. (2002). Acoustic correlates of linguistic rhythm: Perspectives. In *Proceedings of speech prosody 2002* (pp. 115–120). Aix-en-Provence.
- Ramus, F., Dupoux, E., & Mehler, J. (2003). The psychological reality of rhythm classes: Perceptual studies. In *Proceedings of the 15th international congress of phonetic sciences* (pp. 337–342). Barcelona.
- Ramus, F., Nespors, 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” and “syllable-timed” languages. In D. Crystal (Ed.), *Linguistic controversies*. London: Edward Arnold.
- Swan, M., & Smith, B. (1987). *Learner English*. Cambridge: Cambridge University Press.
- Turk, A. E., & White, L. S. (1999). Structural influences on accentual lengthening in English. *Journal of Phonetics*, 27, 171–206.
- White, L., & Mattys, S. L. (2007). Rhythmic typology and variation in first and second languages. In P. Prieto, J. Mascaró, & M.-J. Solé (Eds.), *Segmental and prosodic issues in romance phonology (Current issues in linguistic theory series)*. Amsterdam, Philadelphia: John Benjamins.
- Whitworth, N. (2002). Speech rhythm production in three German–English bilingual families. In D. Nelson (Ed.), *Leeds working papers in linguistics and phonetics* (Vol. 9, pp. 175–205).