RHYTHMIC TYPOLOGY AND VARIATION IN FIRST AND SECOND LANGUAGES*

LAURENCE WHITE & SVEN L. MATTYS
Department of Experimental Psychology, University of Bristol

Abstract

This paper explores the concept of linguistic rhythm classes through a series of studies exploiting metrics designed to quantify speech rhythm. We compared the rhythm of 'syllable-timed' French and Spanish with that of 'stress-timed' Dutch and English, finding that rate-normalised metrics of vocalic interval variability (VarcoV and nPVI-V), together with a measure of the balance of vocalic and intervocalic intervals (%V), were the most discriminant between the two rhythm groups. The same metrics were also informative about the adaptation of speakers to rhythmically-similar (Dutch and English) or rhythmically-distinct (Spanish and English) second languages, and showed evidence of rhythmic gradience within accents of British English. Patterns of scores in all studies support the notion that rhythmic typology is not strictly categorical. A perceptual study found VarcoV to be the strongest predictor of the rating of a second language speaker's accent as native or non-native.

1. *Introduction*

1.1 *Speech rhythm and rhythm classes*

It has long been asserted that languages fall into distinct rhythm classes (e.g. Pike 1945). Within the languages of Western Europe, Romance languages, such as Spanish, are described as 'syllable-timed' and Germanic languages, such as English, are described as 'stress-timed'. Initial attempts to quantify this distinction appealed to the notion of isochronous units in speech timing: the syllable in syllable-timed languages and the stress-delimited foot in stress-timed languages (e.g. Abercrombie 1967:96-98). Dauer (1983) and

* 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, Casimier Ludwig and Ineke Mennen for help with stimulus preparation and Elizabeth Johnson, Klaske van Leyden, Reinier Salverda, Astrid Schepman, Mike Sharwood Smith, Juan Manuel Toro, Isabelle Viaud-Delmon, Atie Vogelenzang de Jong, Eric-Jan Wagenmakers and Rod Walters for help with recordings. Thanks are also due to James Melhorn for assisting in stimulus preparation for the perceptual study reported here, and for running that experiment. The cross-linguistic production study reported here was first presented at the *Phonetics and Phonology in Iberia* conference 2005. We thank the organisers of this conference, the conference delegates for some very interesting discussions, and Klaske van Leyden, Rod Walters and three anonymous reviewers for very useful comments on an earlier draft of this paper.

others showed, however, that syllable duration varies substantially in syllable-timed languages and, conversely, that inter-stress intervals are highly variable in stress-timed languages (see Ramus, Nespor & Mehler 1999, for a review).

Despite the lack of evidence for isochrony-based rhythm classes, there remains the perception that the contrast between stressed and unstressed syllables is greater in, for example, English than Spanish, at least to native English listeners. Lloyd James (1940) expressed this as the distinction between 'Morse code' English rhythm and 'machine gun' Spanish rhythm. In line with this distinction, Dauer and others (Roach 1982; Dasher & Bolinger 1982) suggested that differences between rhythm classes emerge from contrasts in syllable structure. The phonotactic rules of stress-timed languages typically allow greater complexity in syllable onsets and codas, so that a syllable like *strands*, with three onset and three coda segments, is permissible in English but phonotactically illegal in Spanish or French, the latter two having a much higher preponderance of open syllables such as simple CV syllables. In addition, stress-timed languages have stressed vowels that are substantially longer than (typically reduced) unstressed vowels, whereas syllable-timed languages have much contrast in vowel duration between stressed and unstressed syllables.

1.2 *Rhythm metrics*

A number of rhythm metrics have recently been proposed to exploit these differences in syllable structure and vowel duration. Ramus et al. (1999) suggested indices of rhythm based on a division of the speech signal into vocalic and consonantal intervals, specifically:

- ΔV Standard deviation of vocalic interval duration
- ΔC Standard deviation of consonantal interval duration
- %V Percentage of utterance duration that is made up of vocalic rather than consonantal intervals

They showed that combinations of these 'interval measures' successfully captured the stress-timed vs syllable-timed distinction for a range of languages in a speech-rate controlled corpus.

Ramus (2002) suggested that speech rate normalisation might be necessary when applying interval measures to corpora with variable speech rate. Barry, Adreeva, Russo, Dimitrova and Kostadinova (2003) provided evidence supporting normalisation, showing that both ΔV and ΔC are inversely related to speech rate. In contrast, Dellwo and Wagner (2003) found little evidence of a consistent relationship between %V and speech rate, suggesting normalisation may not be necessary for this metric. Dellwo (2006) exploited a rate-controlled metric, VarcoC—the standard deviation of consonantal interval duration divided by the mean consonantal interval duration—and found that it was better than ΔC at all speech rates for discriminating stress-timed English and German from syllable-timed French.

Taking a parallel approach, again based on a division of speech into vocalic and consonantal intervals, Low, Grabe and Nolan (2000) argued that it is the sequential nature of rhythm that is critical. They proposed a rate-normalised pairwise variability index (PVI) to exploit specifically the durational contrast between successive vocalic intervals, derived by dividing the difference between pairs of vocalic intervals by the sum of the intervals. The normalised PVI for vocalic intervals, nPVI-V, is calculated thus (where m is the number of intervals and d is the duration of the kth interval):

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

Grabe and Low (2002) further proposed a non-rate normalised PVI measure for consonantal intervals, suggesting that normalisation could mask rhythmically-relevant variation in onset and coda structure. The raw PVI for consonantal intervals, rPVI-C, is calculated thus:

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

Utilising both of these metrics, Grabe and Low examined a range of languages and found evidence for stress-timed and syllable-timed groups, as well as rhythmically-intermediate languages, such as Polish and Catalan.

Few direct comparisons have been made between interval measures and pairwise variability indices. We report here on studies in which we attempted to evaluate these metrics. Firstly, we looked for evidence of the traditional distinction between stress-timed and syllable-timed languages, and also we examined the influence of first language (L1) on second language (L2) rhythm. Secondly, we sought evidence for rhythmic distinctions between accents of British English. Finally, we considered how well rhythm metrics predict the ratings of speakers of English as native or non-native.

2. *Production studies of speech rhythm*

Here we present results from two studies designed to evaluate the power of the various rhythm metrics. The first study, reported in detail in White and Mattys (in press), examined speech rhythm in first and second languages. The study was designed to test how well different metrics supported the distinction between stress-timed (English and Dutch) and syllable-timed (Spanish and French) languages. In addition, the effect of L1 on L2 rhythm was considered, by analysing the L2 rhythm of speakers with first and second languages in different rhythm classes (English and Spanish). We hypothesised that, where L2 speakers have a clearly non-native accent, rhythm metrics scores in L2s should reflect the rhythmic properties of both the L1 and the L2. There has been a limited amount of previous research on the influence of L1 on L2 rhythm metric scores, such as a study by Carter (2005) of American Hispanic bilinguals. He found that these speakers had nPVI-V scores which were intermediate between the higher scores for L1 English and the lower scores for L1 Spanish. We predicted that the most useful rhythm metrics should show this pattern in our study, and similar intermediacy of scores for native English speakers of L2 Spanish. We also looked at the L2 rhythm of speakers with first and second languages in the same rhythm class (English and Dutch), with the expectation that there should be little difference in rhythm scores between L1 and L2 speakers in this case.

The second study, reported in detail in White and Mattys (in preparation), examined evidence for rhythmic contrasts between different accents of British English. Accents such as Bristolian English have been held to manifest less widespread vowel reduction and less contrast between the length of tense and lax vowels (Hughes & Trudgill 1996). Accents with pitch-peak delay, such as Welsh Valleys English, may show levelling of the duration contrast between stressed and post-stress syllables. This may arise either from relative shortening of the stressed syllable, or at least the vocalic part of the stressed syllable, or from relative lengthening of the post-stress syllable (see Walters 2003, for Welsh Valleys English). This durational levelling may underpin the perception of Welsh Valleys English as being more syllable-timed than Standard Southern British English (Mees & Collins 1999).

The native dialects of speakers from the Orkney Islands also show pitch peak delay (van Leyden 2004), and this may also be manifest in Orcadians' production of Standard English. For comparison with Orkney English, we also analysed the rhythm of English as spoken by natives of Shetland, which, despite being geographically proximate to Orkney and sharing a Scandinavian substrate for its indigenous dialect, lacks the distinctive pitch-peak delay of Orcadian (van Leyden 2004).

Orkney and Shetland speech have some phonological features in common that distinguish them from SSBE and the other accents in this study. In particular, both show the operation of the Scottish Vowel Length Rule (SVLR), whereby most vowels are short, but are lengthened in certain contexts, such as before voiced fricatives in stressed syllables (e.g. Aitken 1981). The consequences for rhythm metric scores of the SVLR have not been examined: given the existence of phonological contexts in which certain vowels may be substantially lengthened, it seems likely that SVLR speakers will—other things being equal—show more durational variability between vowels than, for example, Spanish speakers.

2.1 *Method*

2.1.1 *Participants: Cross-linguistic study.* There were eight groups of speakers and six speakers in each group. Four groups were composed of L1 speakers, all speaking near-standard European varieties of their native languages: Standard Southern British English – $\mathrm{Eng}_{\mathrm{Eng}}$; Dutch (*Algemeen Nederlands*) – $\mathrm{Dut}_{\mathrm{Dut}}$; Spanish (*castellano*) – $\mathrm{Sp}_{\mathrm{Sp}}$; French (*français neutre*) – $\mathrm{Fr}_{\mathrm{Fr}}$.

The other four groups were composed of L2 speakers: English speakers of Dutch – Dut_{Eng}; Dutch speakers of English – Eng_{Dut}; English speakers of Spanish – Sp_{Eng}; Spanish speakers of English – Eng_{Sp}. Because our hypotheses regarding expected rhythm scores rely on native speakers manifesting some degree of non-native accent, we only used L2 speakers who sounded non-native and had not learnt their L2 in early childhood. L2 speakers had a minimum residency of five months in the country of their second language and had to have at least reasonable competence in the L2, as evidenced by their ability to describe a route around a map with minimal preparation, as well as to read sentences and a short story intelligibly and without widespread hesitations or restarts.

2.1.2 *Participants: British accents study.* There were five groups of speakers and six speakers in each group. The Standard Southern British English (SSBE) speakers were the same as those in the cross-linguistic study, with the same utterance tokens analysed for both studies. In addition, there were four groups of speakers of regional accents of English, from Bristol, the Welsh Valleys, Orkney and Shetland. It should be noted that both Orkney and Shetland dialects are significantly different in syntax and lexis from Standard English or Standard Scots. Speakers reading Standard English sentences, as in this study, are therefore speaking what could be regarded as a second language, albeit an early acquired one, and may not necessarily manifest all the prosodic features of their native dialect.

2.1.3 *Materials.* All sentence materials for the analyses reported here are listed in full in White and Mattys (in press). The five English sentences were adapted from a larger set created by Nazzi, Bertoncini and Mehler (1998) and used in Ramus et al.'s (1999) investigation of rhythm metrics. The adaptations were designed to exclude the approximants /j/, /w/, /r/ and /l/, to facilitate measurement of speech interval duration, given that boundaries between vowels and approximants can be difficult to identify reliably from visual analysis of waveforms and spectrograms. Speakers from Bristol, Orkney and Shetland may, however, realise orthographic post-vocalic "r" as an approximant. Where this occurred, the approximant was included within the vocalic interval: the impact of this procedure on rhythm scores is discussed below.

The sentences for the other languages were constructed along similar lines, with the same set of approximants excluded, although other allophonic approximants were not systematically excluded. Excluding the approximant [ɥ] from all the French sentences proved problematic, so it was taken as part of the vocalic interval in the words *biscuits* "biscuits", *mois* "month" and *nuit* "night", as it could not be separated from the vowel with adequate consistency. Given the cross-linguistic variation in syllable complexity, we attempted to match sentences for overall duration by constructing sentences with slightly more syllables for Spanish and French than for English and Dutch.

2.1.4 *Procedure.* Each recorded sentence was labelled into vocalic intervals and consonantal intervals through visual analysis of the waveform and spectrogram in Praat (www.praat.org), based on standard criteria (e.g. Peterson & Lehiste 1960; see White & Mattys in press, for a description of the specific measurement criteria applied). The duration of these intervals was derived from the labelled speech files using a Praat script. Mid-utterance pauses were excluded from the analysis, as were utterance-initial consonants, and glottalised sections between vowels.

The rhythm metrics calculated for each utterance from the vocalic and consonantal interval durations were as follows.

ΔV	Standard deviation of vocalic interval duration
ΔC	Standard deviation of consonantal interval duration
%V	Sum of vocalic interval duration divided by the total duration of vocalic and consonantal intervals
VarcoV	Standard deviation of vocalic interval duration divided by mean vocalic interval duration, multiplied by 100
VarcoC	Standard deviation of consonantal interval duration divided by mean consonantal interval duration, multiplied by 100
nPVI-V	Normalised Pairwise Variability Index for vocalic intervals (see formula 1a)
rPVI-C	Raw Pairwise Variability Index for consonantal intervals (see formula 1b)

All pairwise comparisons reported below between rhythm metric scores are two-tailed Tukey HSD.

2.2 *Results and discussion*

2.2.1 *First and second language rhythm.* First-language scores for all the rhythm metrics are shown in Table 1. As outlined below, the only metrics that consistently discriminated stress-timed English and Dutch from syllable-timed Spanish and French were %V, VarcoV and nPVI-V.

	Sp_{Sp}	Fr_{Fr}	Eng_{Eng}	Dut_{Dut}
Interval measures				
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)
Pairwise variability indices				
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)

Table 1: *Rhythm metric scores (standard errors) for first language speakers.*

Spanish had significantly lower ΔV scores than all other groups [vs Eng_{Eng}: $p<0.001$; vs Dut_{Dut}: $p<0.001$; vs Fr_{Fr}: $p<0.005$]. Spanish ΔC scores were significantly lower than those for English [$p<0.005$]. There were no other significant differences between languages for either ΔV or ΔC scores. Thus, both metrics suggested that Spanish belongs to a distinct rhythm class, but French appeared to be rhythmically more similar to Dutch and English. One likely reason for this is the greater speech rate for Spanish than French (8.0 syls/s vs 5.6 syls/s): as discussed above, both ΔV and ΔC scores are inversely correlated with speech rate. Indeed, as seen below, the rate-normalised vocalic interval metric, VarcoV, was more successful in distinguishing between traditional rhythm classes.

Both Spanish and French %V scores were higher than those for English and Dutch [Sp_{Sp} vs Eng_{Eng}: $p<0.001$; Sp_{Sp} vs Dut_{Dut}: $p<.001$; Fr_{Fr} vs Eng_{Eng}: $p<0.001$; Fr_{Fr} vs Dut_{Dut}: $p<0.05$]. Stress-timed languages, given their widespread vowel reduction in unstressed syllables and their higher occurrence of onset and coda consonant clusters, had lower %V than syllable-timed languages. However, there was also suggestive evidence of differences within rhythm classes [Eng_{Eng} vs Dut_{Dut}: $p = 0.062$; Sp_{Sp} vs Fr_{Fr}: $p = 0.085$]. If the difference between Dutch and English is reliable, this could relate to the reported less widespread occurrence of vowel reduction in Dutch (Swan & Smith 1987).

English and Dutch VarcoV scores were significantly higher than those for Spanish and French [$p<0.001$ for all four comparisons]. French also had significantly higher VarcoV than Spanish [$p<0.005$]. For VarcoC, however, rate normalisation appeared to eliminate all distinctions between languages, with no significant differences in scores.

Both English and Dutch had higher nPVI-V scores than Spanish and French [$p<0.001$ for all four comparisons]. In addition, Dutch had a higher nPVI-V score than English [$p<0.01$] and French had a higher score than Spanish [$p<0.001$].

The patterns for VarcoV and nPVI-V were similar to each other and in line with expectations based on the greater durational difference between stressed and unstressed vowels in stress-timed languages. As with %V, differences within rhythm classes were also evident, suggesting that this classification is not straightforwardly categorical.

English had significantly higher rPVI-C scores than all other languages [vs Dut_{Dut}: $p<0.01$; vs Fr_{Fr}: $p<0.05$; vs Sp_{Sp}: $p<0.001$] and the difference between French and Spanish approached significance [$p = 0.078$]. As with ΔC, scores did not strongly support the traditional rhythmic distinctions between these languages, perhaps due, once again, to the lack of rate normalisation for these metrics of consonantal interval variation and the consequent influence of speech rate on scores.

Given their power to discriminate expected rhythm classes, reporting of subsequent results will focus on the metrics %V, VarcoV and nPVI-V. (Scores for all metrics in this study are given in White & Mattys in press.) Results for comparisons of L1 and L2 rhythm are shown in Table 2 (there was no L2 analysis for French).

	English			Dutch		Spanish	
	Eng_{Eng}	Eng_{Dut}	Eng_{Sp}	Dut_{Dut}	Dut_{Eng}	Sp_{Sp}	Sp_{Eng}
Interval measures							
%V	38 (0.5)	40 (0.4)	41 (0.9)	41 (1.2)	38 (1.6)	48 (0.8)	52 (0.8)
VarcoV	64 (1.7)	61 (2.7)	54 (3.2)	65 (1.5)	65 (1.7)	41 (2.0)	52 (1.3)
Pairwise variability indices							
nPVI Voc	73 (1.2)	70 (1.6)	66 (4.3)	82 (2.4)	75 (1.6)	36 (1.6)	51 (2.4)

Table 2: *Rhythm metric scores (standard errors) for L1 and L2 speakers.*

We consider first the case of languages from the same rhythm class. There was no significant difference in VarcoV between Eng_{Eng} and Eng_{Dut} and no difference between Dut_{Dut} and Dut_{Eng}. The same was true for nPVI-V. The lack of differentiation between first and second languages by the measures of vocalic interval variability reinforces the idea that Dutch and English are rhythmically similar.

There was no significant difference in %V scores between Eng_{Eng} and Eng_{Dut}, but %V for Dut_{Dut} was higher than for Dut_{Eng}, the difference approaching significance [$p = 0.093$], and the %V scores for Dut_{Eng} being the same as for Eng_{Eng}. This trend was the only distinction between L1 and L2 within the same rhythm class. The fact the English speakers of L2 Dutch have %V scores that are so similar to those of L1 English—likewise Dutch speakers of L2 English and L1 Dutch speakers—suggests that L2 speakers either do not perceive subtle rhythmic distinctions, such as between Dutch and English, or not do not realise them because communication is not compromised by ignoring them.

All three metrics showed discrimination between L1 and L2 when the two languages belonged to different rhythm classes. Eng_{Eng} had lower %V scores than Eng_{Sp} [approaching significance: $p = 0.083$]. Sp_{Sp} had significantly lower %V scores than Sp_{Eng} [$p<0.05$], a surprising result given that L1 Spanish had higher %V than L1 English. Eng_{Eng} had significantly higher VarcoV scores than Eng_{Sp} [$p<0.05$] and Sp_{Sp} had significantly lower scores than Sp_{Eng} [$p<0.05$]. For nPVI-V, there was no significant difference between Eng_{Eng} and Eng_{Sp}, but scores were significantly lower for Sp_{Sp} than for Sp_{Eng} [$p<0.005$].

Comparing metrics of vocalic interval variability, VarcoV appeared slightly more successful in capturing the differences between L1 and L2 rhythm than nPVI-V. The latter metric showed no difference between English L1 and Spanish L2 speakers of English; in contrast, both L1 vs L2 English and L1 vs L2 Spanish were distinguished by VarcoV. As discussed further below, there are several segmental and suprasegmental processes which contribute to patterns of vowel and consonant duration. Thus, the working assumption of this study is that L2 speakers, where they have clear non-native accents, should be distinguished from L1 speakers in terms of rhythm scores, which exploit these patterns of vowel and consonant duration. The marginal preference for VarcoV over nPVI-V stems from this working assumption. Figure 1 shows the VarcoV scores for all L1 and L2 speakers plotted against the %V scores.

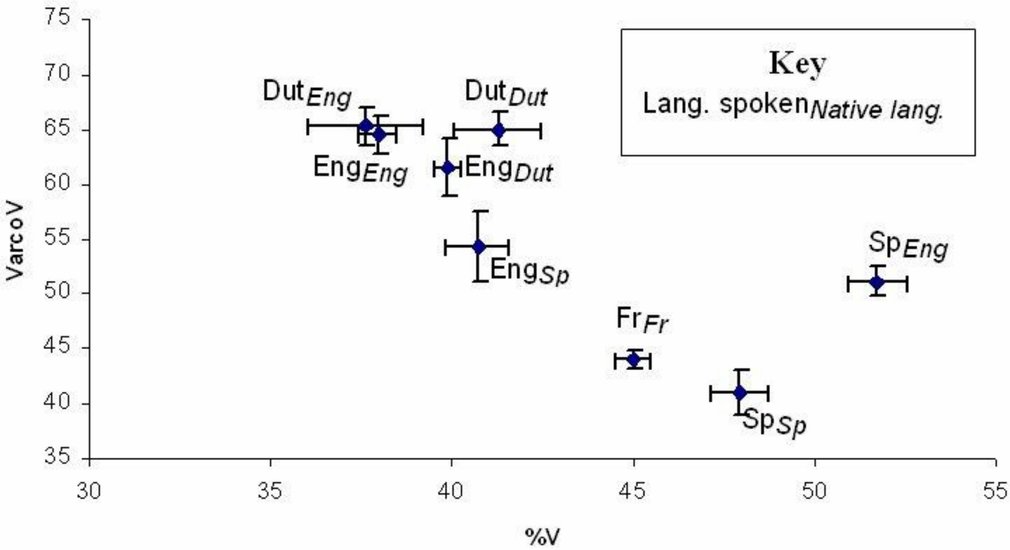

Figure 1: *VarcoV and %V scores and standard error bars for all first and second language groups. Eng – English; Dut – Dutch; Sp – Spanish; Fr – French.*

For VarcoV, L2 speakers had rhythm scores intermediate between scores for their L1 and those for native speakers of the L2. VarcoV scores suggest that Spanish speakers of English accommodate towards the shorter unstressed vowels of their L2, and may also produce longer stressed vowels than in their L1, but do not make the distinction between stressed and unstressed vowels as great as do native English speakers. English speakers of Spanish appear to produce unstressed vowels that are longer than those of English but not as long as those of native Spanish speakers.

Given this general intermediacy of L2 rhythm, the pattern for %V in L1 and L2 Spanish is rather surprising: as shown in Figure 1, English speakers of L2 Spanish actually had %V scores higher than those of L1 Spanish speakers. There are a number of segmental and suprasegmental differences between Spanish and English that could account for this pattern. At the segmental level, English speakers of Spanish may produce vowels with generally greater duration than Spanish speakers, particularly where the closest English vowel is a diphthong: for example, English speakers may realise Spanish [e] as the diphthong [eɪ], or at least retain its greater duration. At the suprasegmental level, English may have more marked prosodic lengthening. For example, Ortega-Llebaria and Prieto (this volume) reported little evidence of accentual lengthening in Castilian Spanish, in contrast with American or British English (Turk & Sawusch 1997; Turk & White 1999). Likewise, perceived phrase-final lengthening is less widespread in Castilian Spanish than some other Romance languages (Frota, D'Imperio, Elordieta, Prieto & Vigário, this volume), whereas phrase-final lengthening is very well attested in varieties of English (e.g. Wightman, Shattuck-Hufnagel, Ostendorf & Price 1992). If English speakers retain their native prosodic lengthening patterns in L2 Spanish, these processes may conspire to increase the vocalic proportion of the total utterance: the preponderance of open syllables in Spanish means, for example, that final lengthening is likely to affect vowels more than consonants.

The overall pattern that emerges from this study of first and second language rhythm is that certain rhythm metrics, particularly %V and VarcoV, and to a slightly lesser extent nPVI-V, capture cross-linguistic differences in vowel duration and syllable onset and coda phonotactics. It is these durational and phonotactic contrasts that conspire to make stressed syllables relatively strong in English and Dutch, and make the relative strengths of stressed and unstressed syllables less different in Spanish and French. Clearly, given the range of contributing factors, the relative strength of stressed and unstressed syllables is gradiently variable, so, although there is evidence here for stress-timed vs syllable-timed language grouping, this typology is highly unlikely to be categorical.

2.2.2 *Accents of British English.* We report the results for the three rhythm metrics that were found to be most useful for discriminating stress-timed from syllable-timed languages, namely, %V, VarcoV and nPVI-V. (Scores for accents of British English for the other metrics discussed above are reported in White & Mattys in preparation.) Table 3 shows the scores for these metrics for the five accents of British English studied.

	SSBE (= Eng_{Eng})	Shetland	Orkney	Welsh Valleys	Bristol
Interval measures					
%V	38 (0.5)	39 (0.6)	40 (1.0)	42 (0.7)	41 (0.5)
VarcoV	64 (1.7)	59 (1.2)	53 (2.0)	53 (1.8)	57 (2.2)
Pairwise variability indices					
nPVI-V	73 (1.2)	77 (2.6)	70 (1.9)	66 (2.9)	70 (2.1)

Table 3: *Means (standard errors) of rhythm metrics for SSBE (Standard Southern British English), Shetland, Orkney, Welsh Valleys and Bristol English.*

SSBE had significantly lower %V scores than Bristol and Welsh Valleys accents [$p<0.005$ for both comparisons], and the difference between SSBE and Orkney approached significance [$p = 0.059$]. Shetland also had significantly lower %V scores than Bristol and Orkney accents [$p<0.05$ for both comparisons]. No other differences in %V scores between accents were significant.

A similar pattern of discrimination was observed for VarcoV: scores for SSBE were significantly greater than those for all other accents [vs Bristol: $p<0.05$; vs Welsh Valleys: $p = 0.001$; vs Orkney: $p = 0.001$; vs Shetland: $p<0.05$]. In addition, VarcoV scores for Shetland were significantly higher than those for Welsh Valleys and Orcadian [$p<0.05$ for both comparisons]. No other differences in VarcoV scores between accents were significant.

Thus, both %V and VarcoV suggest a rhythmic accent grouping of Bristolian, Welsh Valleys and Orcadian, with a lesser degree of stress-timing and more evidence of syllable-timing than SSBE. This grouping appears to accord well with the segmental and/or suprasegmental properties of these accents which reduce the vowel duration contrast between stressed and unstressed syllables. The %V scores of Bristolian, Orcadian and Shetland may have been somewhat elevated due to the inclusion of post-vocalic /r/ in vocalic intervals, so some caution must be exercised in interpreting these results. The fact that Orkney had higher %V than Shetland, however, indicates that the treatment of rhoticity is not the sole cause of the apparent greater syllable timing in the former, given the both Orkney and Shetland speakers manifest rhoticity.

In contrast, the scores for nPVI-V were less supportive of this distinction. Although the mean scores were lower for Bristol, Welsh Valleys and Orkney than for SSBE, only the comparison between SSBE and Welsh Valleys approached significance [$p = 0.077$]. Shetland had the highest nPVI-V scores, significantly greater than those for Welsh Valleys [$p<0.05$], with the difference approaching significance for the comparison with Orkney English [$p = 0.064$].

2.2.3 *Discussion: Evaluation of rhythm metrics.* Thus, the results of the cross-linguistic study and the study of British accents converge on the conclusion that %V and VarcoV offer the best discrimination between accent and language groups that are usually held to differ rhythmically. These distinctions are illustrated for British accents in Figure 2, where it can be seen that Bristol, Orkney and Welsh Valleys English have %V and VarcoV scores that are intermediate between those of SSBE and Castilian Spanish.

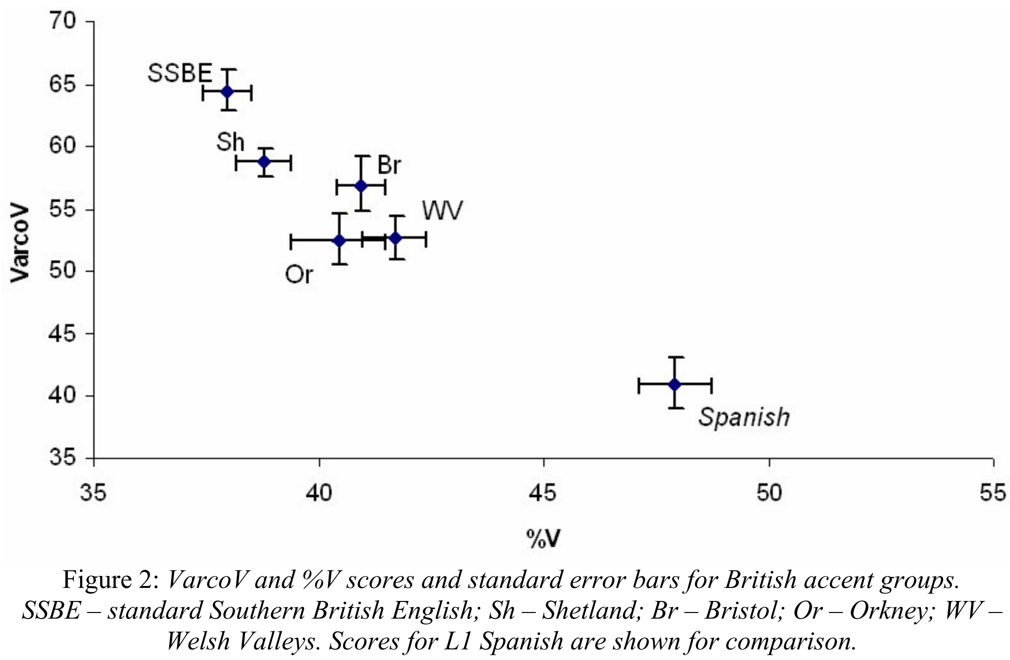

Figure 2: *VarcoV and %V scores and standard error bars for British accent groups. SSBE – standard Southern British English; Sh – Shetland; Br – Bristol; Or – Orkney; WV – Welsh Valleys. Scores for L1 Spanish are shown for comparison.*

The scores for nPVI-V also offer some support for these distinctions, but in both studies expected differences were not found and some differences emerged which had not been predicted. In particular, nPVI-V did not discriminate between native English and L2 English spoken by Spanish speakers. It is also worth noting that Dutch and English had rather different, though not significantly different, nPVI-V scores, with Dutch appearing more stress-timed, a distinction not apparently motivated by what is known about the languages and not reflected in the rhythmic differences between L1 and L2 speakers in the English/Dutch comparison. Finally, nPVI-V did not discriminate between the rhythm of SSBE and that of Bristol and Orkney accents of English, a distinction consistent with the processes affecting stressed vs unstressed vowel duration and which is supported by the scores for %V and VarcoV. It may be that nPVI-V, based as it is on syntagmatic comparisons of vowel duration, is actually too sensitive to the characteristics of individual utterances, with the cruder global measures %V and VarcoV better able to capture broad rhythmic trends.

3. *Rhythm metrics and the perception of native and non-native accent*

Of all the rhythm metrics assessed in the production studies described above, %V and VarcoV best capture expected differences in speech rhythm between languages and language varieties. There is a strong case for arguing, however, that linguistic rhythm is primarily a perceptual phenomenon. Research is clearly required on how these rhythm metrics relate to an individual's perception of speech and of the differences between languages, language varieties and individual speakers.

Studying rhythmic perception in isolation is not straightforward. Segmental information can be removed from speech by low-pass filtering and fundamental frequency contours can be flattened to remove distinctive intonation patterns, but such an approach may leave insufficient information for listeners to make perceptual judgments (e.g. van Leyden 2004). Other researchers have used resynthesis to generate segmentally and intonationally monotonous speech whilst preserving the durational characteristics of the original signal (e.g. Ramus, Dupoux & Mehler 2003).

As a first pass at establishing perceptual correlates of rhythm metrics, we here report a simple accent judgment experiment, in which native English participants rated accents of English as sounding more or less native or non-native. Clearly, with unprocessed speech, there will be many segmental and suprasegmental indicators of a speaker's linguistic origin available, some of which will have no bearing on rhythm. Given, however, that a range of linguistic processes influence the duration of segments, in particular, the relative duration of stressed and unstressed vowels, we predict that the most effective rhythm metrics should capture something of the variability that leads to perceptions of speech as native and non-native.

3.1 *Method*

The methodology for the perception experiment was that utilised in a review study of native accent assessment (Piske, MacKay & Flege 2001). Participants were given a nine-point scale of accent nativeness/non-nativeness and told to rate each of a series of auditorily-presented utterances according to this scale. The only difference in our methodology was that we counterbalanced the polarity of the scale: half of participants were told that a rating of 1 should indicate ‘no foreign accent’ and a rating of 9 should indicate ‘strong foreign accent’; for the other participants, the ratings scale was reversed, with 1 indicating ‘strong foreign accent’ and 9 indicating ‘no foreign accent’. This was done to control for potential response bias in the use of the scale. In calculating the mean overall ratings, the ratings were inverted for the second set of participants, so that lower ratings consistently meant a more native-like accent.

3.1.1 *Participants.* Participants were twelve native speakers of English with no self-reported hearing or speaking problems. They were paid a small honorarium or received course credit for their participation.

3.1.2 *Materials.* Three groups of speakers were used from the cross-linguistic study described above: native SSBE speakers; native Dutch speakers and native Spanish speakers, with three female and three male speakers in each group. Each speaker read each of the five English experimental sentences. Thus there were a total of ninety experimental utterances and another ninety similar utterances were also rated.

3.1.3 *Procedure.* Participants were seated in front of a computer monitor and keyboard, and were presented with the utterances over headphones. A nine-point scale was marked on the keyboard and participants were told to rate each utterance according to its degree of foreign accent, as described above. After a short practice block using a sample of the utterances, participants were played all of the 180 utterances three times, in three blocks of separately-randomised order.

3.2 *Results*

Mean accent ratings were: for SSBE speakers, 1.4; for Dutch English speakers, 3.6; for Spanish English speakers, 6.7. A by-Subjects repeated measures ANOVA showed a main effect of accent group on ratings [$F(2,22) = 445.83$, $p<0.001$]. Mean ratings for all accent groups differed from each other at the $p<0.001$ level.

Figure 3 shows the mean ratings for each accent group broken down by speakers. ANOVAs showed main effects of speaker on ratings for all accent groups [SSBE: $F(5,55) = 2.88$, $p<0.05$; Dutch English: $F(5,55) = 49.41$, $p<0.001$; Spanish English: $F(5,55) = 38.66$, $p<0.001$]. Thus, for all groups, accent rating varied between speakers. As Figure 3 indicates, this variation was much less for SSBE speakers than for the other groups.

In line with the primary distinction between stress-timed and syllable-timed languages, Dutch speakers were rated as being more native-like in their production of English than Spanish speakers. As can be seen from Figure 3, two Dutch speakers were rated as much more non-native than the other four, indicating that, of course, factors other than rhythm influence accent perception.

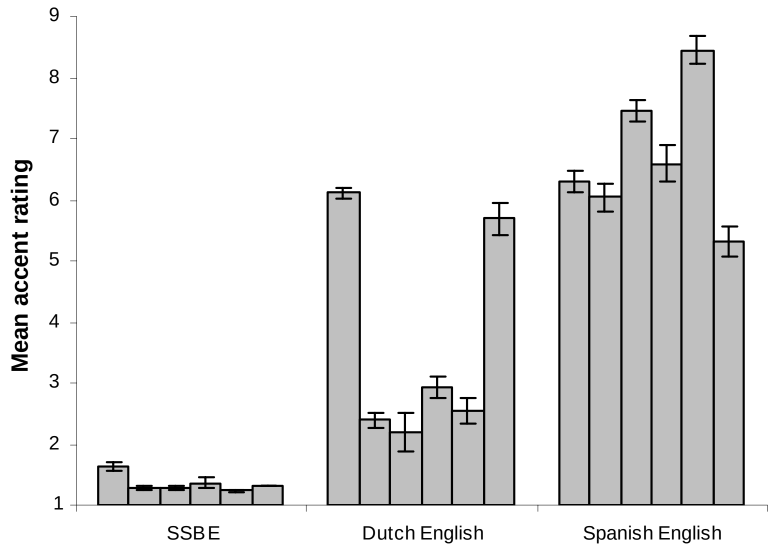

Figure 3: *Mean accent ratings by speakers in each accent group.*

Table 4 shows a correlation matrix between accent ratings, rhythm metrics and speech rate. There are a number of trends to note in the pairwise correlations. Firstly, there are expected correlations between measures of vocalic intervals variability (ΔV, VarcoV, nPVI-V) and between measures of consonantal interval variability (ΔC, rPVI-C). Secondly, the non-rate-normalised measures (ΔV, ΔC, rPVI-C) show evidence of inverse correlations with speech rate, as was discussed in the introduction. Thirdly, only the three metrics that emerged from the production studies as most effectively discriminant between rhythmic groups show strong significant positive (%V) or negative (VarcoV, nPVI-V) correlations with accent ratings. As these measures are also correlated with each other, further analysis is reported below to assess which metrics best predict accent rating. Finally, speech rate and accent rating are also inversely correlated, as would be expected, given that fluent speakers should tend to speak more quickly than less fluent speakers.

The mean scores for the seven rhythm metrics and speech rate were used as predictor variables for accent ratings in a stepwise linear regression, the results of which are shown in Table 5. VarcoV was found to be the best single predictor of accent rating (r^2 = 0.541). A model incorporating both VarcoV and speech rate accounted for a greater proportion of the accent ratings (r^2 = 0.669) than VarcoV alone, but no additional rhythm metrics made significant further contributions to predicting accent ratings.

	ΔV	ΔC	%V	VarcoV	VarcoC	nPVI-V	rPVI-C	Speech rate
Accent rating	−0.127	0.019	0.647**	−0.735**	−0.264	−0.561*	−0.123	−0.486*
ΔV		0.601**	0.000	0.556**	0.170	0.661**	0.495*	−0.685**
ΔC			−0.151	0.183	0.664**	0.472*	0.953**	−0.576*
%V				−0.603**	−0.177	−0.539*	−0.227	−0.338
VarcoV					0.316	0.863**	0.193	0.178
VarcoC						0.463†	0.698**	0.136
nPVI-V							0.451†	−0.095
rPVI-C								−0.437†

Table 4: *Pairwise Pearson correlations between accent ratings, rhythm metrics and speech rate. Two-tailed significance levels: †: 0.10>p≥0.05; *: p<0.05; **: p<0.01.*

	r	r^2	Adjusted r^2	β	t
Model 1	0.735	0.541	0.512		
VarcoV				−0.735	−4.34**
Model 2	0.818	0.669	0.625		
Varco V				−0.671	−4.44**
Speech rate				−0.363	−2.41*

Table 5: *Stepwise linear regression with accent rating as dependent variable and rhythm metrics (ΔV, ΔC, %V, nPVI-V, rPVI-C, VarcoV, VarcoC) and speech rate as potential predictor variables. Significance levels: *: p<0.05; **: p<0.01.*

To look at the contribution of VarcoV further, we examined the partial correlation of VarcoV with accent rating, taking speech rate into account, separately for the three native speaker groups: English, Dutch and Spanish. Given the small number of different speakers, we used the scores for each of the five utterances by each of the six speakers (rather than the speaker means) giving a total of thirty scores for each language group. The partial correlation between VarcoV rating was not significant for English and Dutch native speakers [English: $r = -0.003$, p = 0.998; Dutch: $r = -0.147$, p = 0.446], but it was significant for Spanish speakers [$r = -0.440$, $p<0.05$]. Figure 4, showing a scatter-plot and a partial regression line, gives an indication of this relationship, demonstrating that utterances by Spanish speakers are likely to be rated as more non-native if VarcoV scores are lower.

3.3 *Discussion*

The results of this simple perceptual experiment offer support for the conclusions drawn from both of the production studies discussed above regarding the value of the VarcoV metric as a measure of speech rhythm. VarcoV was found overall to be the best predictor of accent ratings, even when the contribution of speech rate, clearly an important difference between native and non-native speakers, was also considered.

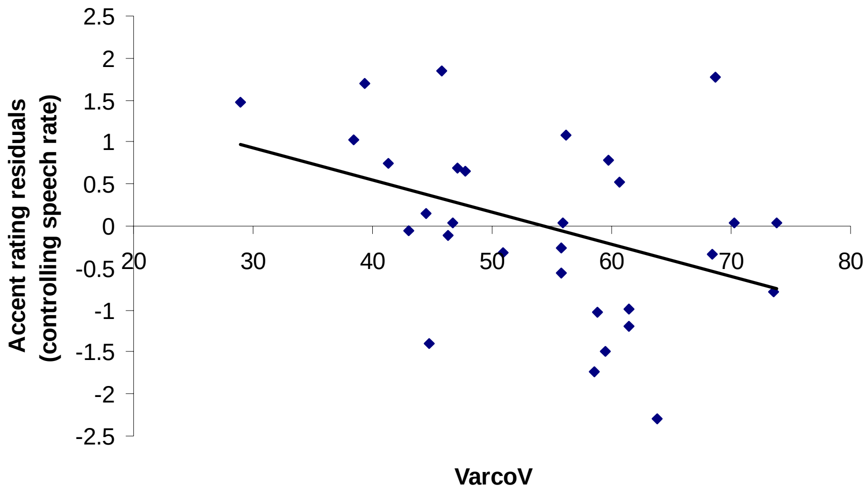

Figure 4: *Relationship between accent ratings and VarcoV, showing partial regression line, for Spanish speakers of L2 English.*

The power of VarcoV as a predictor of accent rating was clearly demonstrated for the utterances produced by native Spanish speakers, though not for the other two groups.

The correlation between accent rating and VarcoV is readily interpretable in terms of the accommodation required for a native Spanish speaker to produce L2 English: speakers who produce short, reduced vowels in unstressed syllables in L2 English speech will, other things being equal, have higher VarcoV scores than speakers who consistently produce long, full unstressed vowels in L2 English. The former group, with higher VarcoV scores, are likely to be rated as more English-like in their speech production. We would also predict that English speakers who have lower VarcoV scores in L2 Spanish should be rated as more native-like, reflecting their production of relatively long, full-vowel unstressed syllables and stressed syllables that are not greatly longer.

The failure to find a similar relationship between VarcoV and accent rating for the other two language groups is not surprising. For native English speakers, although there was evidence of slight variation between speakers, accent ratings were essentially at floor level, with a mean (1.35) only just above the lowest possible rating (1). For Dutch speakers, the fact that rhythm may play a lesser role in accent perception is indicated by the presence of an essentially bimodal distribution of accent ratings (see Figure 3): thus, two speakers had much higher ratings of non-nativeness than the other four, although their scores across the range of rhythm metrics were not markedly different. As indicated in the cross-linguistic production study reported above, the rhythmic differences between Dutch and English are slight and, to the extent that they exist at all, are indexed by %V rather than VarcoV. Thus segmental and prosodic processes not impinging on linguistic rhythm seem likely to account for variations in the perception of the nativeness of Dutch speakers of English.

4. *General discussion*

4.1 *Rhythm metrics*

We have reported on three studies designed to assess the discriminative performance of a range of speech rhythm metrics. The metrics %V, VarcoV and nPVI-V most clearly discriminated stress-timed English and Dutch from syllable-timed Spanish and French. Of these measures, nPVI-V did not discriminate quite as effectively as %V or VarcoV in the comparison of first and second language rhythm, finding no rhythmic difference between native English speakers and Spanish speakers of L2 English, despite the perceptible non-native accent of the latter group.

We also looked for evidence of rhythmic differences between accents of British English. Once again, %V and VarcoV were the most discriminant metrics, suggesting that there may be significant variability in rhythm even within a canonically stress-timed language like English. This variability is likely to arise, at least in part, from segmental and suprasegmental processes that affect the durational balance of stressed and post-stress syllables.

Finally, we reported a perceptual experiment which tested the power of rhythm metrics to predict ratings of English speech as native or non-native. VarcoV proved the most effective metric for this task, indexing variability in vowel duration—a factor that should be a key indicator of native or non-native accent. Scores for %V and nPVI-V also showed correlations with accent ratings, but their correlations with VarcoV (negative and positive respectively) meant that they did not emerge as independently reliable predictors of rating.

4.2 *Rhythmic typology and variation*

The evidence from the production study of first and second languages reported above, and from studies such as Grabe and Low (2002), suggests that stress-timing vs syllable-timing is a gradient distinction between languages. The study of English accents reported here also showed support for postulated rhythmic variation between varieties of British English. Clearly, rhythmic variation may also be found between varieties of other languages. It has been held, for example, that southern varieties of Italian are relatively more stress-timed than the syllable-timed northern varieties (e.g. Grice, D'Imperio, Savino & Avesani 2005), although few quantitative data appear to be available. Also, Latin American Spanish subjectively conveys the impression of conferring greater salience on at least some stressed syllables than Castilian Spanish. Rhythmic metrics suggest a method of quantifying these apparent variations within Romance languages.

Drawing together different strands of work on speech timing provides suggestive evidence that gradient rhythmic distinctions may be paralleled in differences in prosodic timing processes. Stress-timed languages such as English use strong durational cues to indicate syllable stress, and also indicate prosodic structure with localised lengthening effects (e.g. Wightman et al. 1992). As we have seen, the durational difference between stressed and unstressed syllables is less marked in Spanish. Recent studies suggest that timing may also have a lesser role in the indication of prosodic structure in Spanish. Ortega-Llebaria and Prieto (this volume) reported that pitch accent in Castilian Spanish was not consistently marked by lengthening within the accented word, in contrast with English (e.g. Turk & White 1999). Similarly, Frota et al. (this volume) reported that perceived final lengthening was not a reliable feature of Castilian Spanish prosodic phrasing.

Any link between rhythm typology and prosodic timing processes is unlikely to be categorical. Indeed, the evidence from Frota et al.'s study is that phrase-final lengthening is widespread in Italian, but this could reflect the particular variety of Italian analysed. Their study analysed Neapolitan Italian, held to be more stress-timed than northern varieties. The possibility that varieties of a single language could show covariance in their degree of stress-timing and in their durational marking of prosodic boundaries is intriguing. More empirical work is required to settle this question, but it does suggest a promising direction for research utilising rhythm metrics.

4.3 *Rhythm perception*

It is clear that further perceptual studies are necessary to determine what rhythm metrics such as %V and VarcoV tell us about the experience of linguistic rhythm for the listener. Firstly, the role of rhythm in linguistic discrimination should be explored further. Using speech resynthesis techniques designed to eliminate cues other than rhythm, Ramus et al. (2003) showed that language classes suggested by rhythm metric scores corresponded to listeners' perceptual groupings. Similar techniques could be applied to assessing the extent of rhythmic variation between accents of the same language.

The possibility of gradience in the role of rhythm in the perception of linguistic juncture should also be explored. It has been shown that the statistical predominance of word-initial stress in Germanic languages is used by listeners in the identification of word boundaries (e.g. Cutler & Norris 1988, for English; Vroomen & de Gelder 1997, for Dutch), at least when other, more reliable cues are not available (Mattys, White & Melhorn 2005). In contrast, Romance languages tend to have stressed syllables in penultimate or word-final position. The consequence of this latter arrangement for segmentation has barely been tested, however. The results of this and other recent research raise the possibility that rhythmic gradience within languages—specifically, variation in the relative strength of stressed and unstressed syllables—may be paralleled by gradience in listeners' exploitation of stress-based segmentation strategies. The rhythm metrics positively evaluated here, in particular %V and VarcoV, provide tools that will facilitate future research in such directions.

References

Abercrombie, David. 1967. *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.

Aitken, Adam J. 1981. "The Scottish vowel length rule". *So meny People, Longages and Tonges: Philological Essays in Scots and Mediaeval English presented to Angus McIntosh* ed. by Michael Benskin & Michael L. Samuels. 131-157. Edinburgh: The Middle English Dialect Project.

Barry, William J., Bistra Andreeva, Michela Russo, Snezhina Dimitrova & Tania Kostadinova. 2003. "Do rhythm measures tell us anything about language type?" *Proceedings of the 15th International Congress of Phonetics Sciences* ed. by Maria-Josep Solé, Daniel Recasens & Joaquín Romero. 2693-2696. Barcelona: Causal Productions.

Carter, Phillip M. 2005. "Quantifying rhythmic differences between Spanish, English, and Hispanic English". *Theoretical and Experimental Approaches to Romance Linguistics: Selected Papers from the 34th Linguistic Symposium on Romance Languages* ed. by Randall S. Gess & Edward J. Rubin. 63-75. Amsterdam & Philadelphia: John Benjamins.

Cutler, Anne & Dennis G. Norris. 1988. "The role of stressed syllables in segmentation for lexical access". *Journal of Experimental Psychology: Human Perception and Performance* 14. 113-121.

Dasher, Richard & Dwight L. Bolinger. 1982. "On pre-accentual lengthening". *Journal of the International Phonetic Association* 12. 58-69.

Dauer, Rebecca M. 1983. "Stress-timing and syllable-timing reanalyzed". *Journal of Phonetic*s 11. 51-62.

Dellwo, Volker. 2006. "Rhythm and speech rate: A variation coefficient for deltaC". *Language and Language Processing: Proceedings of the 38th Linguistic Colloquium* ed. by Pawel Karnowski & Imre Szigeti. 231-241. Frankfurt: Peter Lang.

---------- & Petra Wagner. 2003. "Relations between language rhythm and speech rate". *Proceedings of the 15th International Congress of Phonetics Sciences* ed. by Maria-Josep Solé, Daniel Recasens & Joaquín Romero. 471-474. Barcelona: Causal Productions.

Grabe, Esther & Ee Ling Low. 2002. "Durational variability in speech and the rhythm class hypothesis". *Papers in Laboratory Phonology 7* ed. by Natasha Warner & Carlos Gussenhoven. 515-546. Berlin: Mouton de Gruyter.

Grice, Martine, Mariapaola D'Imperio, Michelina Savino & Cinzia Avesani. 2005. "A strategy for intonation labelling varieties of Italian". *Prosodic Typology: The Phonology of Intonation and Phrasing* ed. by Sun-Ah Jun. 362-389. Oxford: Oxford University Press.

Hughes, Arthur & Peter Trudgill. 1996. *English Accents and Dialects*. London: Arnold.

Leyden, Klaske van. 2004. Prosodic characteristics of Orkney and Shetland dialects. PhD diss., Leiden University. Utrecht: LOT Dissertation Series 92.

Lloyd James, Arthur. 1940. *Speech Signals in Telephony*. London: Pitman.

Low, Ee Ling, Esther Grabe & Francis Nolan. 2000. "Quantitative characterisations of speech rhythm: 'Syllable-timing' in Singapore English". *Language and Speech* 43. 377-401.

Mattys, Sven L., Laurence White & James F. Melhorn. 2005. "Integration of multiple speech segmentation cues: A hierarchical framework". *Journal of Experimental Psychology: General* 134. 477-500.

Mees, Inger M., & Beverley Collins. 1999. "Cardiff: A real-time study of glottalization". *Urban Voices: Accent Studies in the British Isles* ed. by Paul Foulkes & Gerard Docherty. 185-202. London: Arnold.

Nazzi, Thierry, Josiane Bertoncini & Jacques Mehler. 1998. "Language discrimination by newborns: Towards an understanding of the role of rhythm". *Journal of Experimental Psychology: Human Perception and Performance* 24. 756-766.

Peterson, Gordon E. & Ilse Lehiste. 1960. "Duration of syllable nuclei in English". *Journal of the Acoustical Society of America* 32. 693-703.

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

Piske, Thorsten, Ian R.A. MacKay & James E. Flege. 2001. "Factors affecting degree of foreign accent in an L2: A review". *Journal of Phonetics* 29. 191-215.

Ramus, Franck. 2002. "Acoustic correlates of linguistic rhythm: Perspectives". *Proceedings of Speech Prosody* ed. by Bernard Bel & Isabel Marlien. 115-120. Aix-en-Provence: Laboratoire Parole et Langage.

----------, Marina Nespor & Jacques Mehler. 1999. "Correlates of linguistic rhythm in the speech signal". *Cognition* 73. 265-292.

----------, Emmanuel Dupoux & Jacques Mehler. 2003. "The psychological reality of rhythm classes: Perceptual studies". *Proceedings of the 15th International Congress of Phonetic Sciences* ed. by Maria-Josep Solé, Daniel Recasens and Joaquín Romero. 337-342. Barcelona: Causal Productions.

Roach, Peter. 1982. "On the distinction between 'stress-timed' and 'syllable-timed' languages". *Linguistic controversies* ed. by David Crystal. 73-79. London: Edward Arnold.

Swan, Michael & Bernard Smith. 1987. *Learner English*. Cambridge: Cambridge University Press.

Turk, Alice E. & James R. Sawusch. 1997. "The domain of accentual lengthening in American English". *Journal of Phonetics* 25. 25-41.

---------- & Laurence White. 1999. "Structural influences on accentual lengthening in English". *Journal of Phonetics* 27. 171-206.

Vroomen, Jean & Beatrice de Gelder. 1997. "Activation of embedded words in spoken word recognition". *Journal of Experimental Psychology: Human Perception and Performance* 23. 710-720.

Walters, J. Robert. 2003. "On the intonation of a South Wales 'Valleys accent' of English". *Journal of the International Phonetic Association* 332. 211-238.

Wightman, Colin W., Stefanie Shattuck-Hufnagel, Mari Ostendorf & Patti J. Price. 1992 "Segmental durations in the vicinity of prosodic phrase boundaries". *Journal of the Acoustical Society of America* 91. 1707-1717.

White, Laurence & Sven L. Mattys. In press. "Calibrating rhythm: First language and second language studies". *Journal of Phonetics*.