

Rhythm, Timing and the Timing of Rhythm

Amalia Arvaniti

University of California, San Diego, Calif., USA

Abstract

This article reviews the evidence for rhythmic categorization that has emerged on the basis of rhythm metrics, and argues that the metrics are unreliable predictors of rhythm which provide no more than a crude measure of timing. It is further argued that timing is distinct from rhythm and that equating them has led to circularity and a psychologically questionable conceptualization of rhythm in speech. It is thus proposed that research on rhythm be based on the same principles for all languages, something that does not apply to the widely accepted division of languages into stress- and syllable-timed. The hypothesis is advanced that these universal principles are grouping and prominence and evidence to support it is provided.

Copyright © 2009 S. Karger AG, Basel

1. Introduction

The existence of two rhythmic categories, stress- and syllable-timing, has been the foundation of phonetic research on rhythm [among many, Bolinger, 1965; Abercrombie, 1967; Lehiste, 1977; Nakatani et al., 1981; Bertinetto, 1989] and has also been occasionally employed in phonological research [e.g., Nespor and Vogel, 1989; Coetzee and Wissing, 2007]. However, empirical studies failed for a long time to show evidence for isochrony, the equal duration of feet and syllables in stress- and syllable-timed languages, respectively, that is the cornerstone of the stress-/syllable-timing division. As a result, the notion of rhythmic types began to lose its appeal, despite some evidence that infants can differentiate languages depending on rhythmic type [e.g., Nazzi et al., 1998, 2000; Nazzi and Ramus, 2003] and that speech processing by adults relies on syllables or feet depending on the rhythmic type of the listeners' native language [Cutler et al., 1986, 1992]. (The use of the mora in the processing of languages classified as mora-timed, such as Japanese and Telugu, has been demonstrated by Otake et al. [1993], Cutler and Otake [1994] and Murty et al. [2007].)

The stress-/syllable-timing distinction received renewed interest with the advent of rhythm metrics, formulas that seek to quantify consonantal and vocalic variability and use this quantification to classify languages rhythmically. The first such attempt was made by Ramus et al. [1999]; since then several alternative metrics have been

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

© 2009 S. Karger AG, Basel
0031–8388/09/0662–0046
\$26.00/0
Accessible online at:
www.karger.com/pho

Amalia Arvaniti
Department of Linguistics, UCSD
9500 Gilman Dr. No. 0108
La Jolla, CA 92093-0108 (USA)
Tel. +1 858 534 8409, Fax +1 858 534 4789,
E-Mail amalia@ling.ucsd.edu

proposed [Frota and Vigário, 2001; Grabe and Low, 2002; Wagner and Dellwo, 2004; Dellwo, 2006; White and Mattys, 2007], as well as other methods of quantifying rhythmic distinctions [Lee and McAngus Todd, 2004; Rouas et al., 2005]. Nevertheless, it is fair to say that the most widely used metrics remain (a) %V and ΔC – the percentage of vocalic intervals in speech and the standard deviation of consonantal duration – proposed by Ramus et al. [1999], and (b) the pairwise variability indices nPVI and rPVI (pairwise comparisons of successive vocalic and intervocalic intervals) introduced by Grabe and Low [2002], following Low et al. [2000].

Since the success of metrics is often taken for granted and has been used to validate the notion of rhythmic types, in what follows, I briefly review the results of research using metrics and discuss the reasons why they turned out to be much less reliable than originally envisioned. I further argue that these reasons are directly linked to the confounding of rhythm with timing and the reliance on the notion of rhythmic types and all it entails for our understanding of speech rhythm. In particular, I show that the notion of rhythmic types is problematic and largely incompatible with psychological evidence on rhythm. Finally, I sketch an alternative view of how rhythm may be created in all languages on the basis of grouping and prominence and provide some evidence in its favor.

2. Qualifying the Success of Metrics

In their seminal article, Ramus et al. [1999] tested several ways of quantifying durational profiles and concluded that of these %V and ΔC best reflect rhythmic type. Soon afterwards, however, Grabe and Low [2002] showed that different measures can yield different classifications for the same language: for example, PVIs classify Thai as stress-timed, but %V – ΔC classify it as syllable-timed, while the reverse obtains with Luxembourgish; %V – ΔC set Japanese apart from the other languages, supporting the existence of mora-timing as a distinct rhythmic category, but PVI scores group it with syllable-timed languages.

In addition to such disagreements, the results of Grabe and Low [2002] – the only study in which a large number of languages has been examined – show that metrics are not very successful at classifying nonprototypical languages. Of the 18 languages Grabe and Low [2002] tested, stress-timed British English, Dutch and German, and syllable-timed Spanish and French are prototypical and were classified as expected. Of the rest, 9 languages remained unclassifiable (Greek, Malay, Romanian, Singapore English, Tamil, Welsh) or were said to have mixed rhythm (Catalan, Estonian, Polish). Only 4 languages were unambiguously classified by the PVIs as stress-timed (Thai) or syllable-timed (Japanese, Luxembourgish, and Mandarin). Thus, of the 13 nonprototypical languages in the sample, only 4 can be said to be classified with some success (if we disregard the fact that Japanese, Luxembourgish and Thai are classified differently by %V – ΔC and that Japanese is impressionistically classified as mora-timed). This gives, at best, a classification success rate of 33% among nonprototypical languages.

In an attempt to explain the rather limited success of Grabe and Low [2002], Ramus [2002] argued that their study lacked a controlled corpus of data and a method for cross-linguistic control of speech rate (even though Grabe and Low's [2002] nPVI is a measure that normalizes for speaking rate). In support, Ramus [2002, p. 116] calculated PVI scores for the controlled corpus of Ramus et al. [1999] and showed that if this corpus is used, then the two types of metrics provide figures of 'striking' similitude and thus

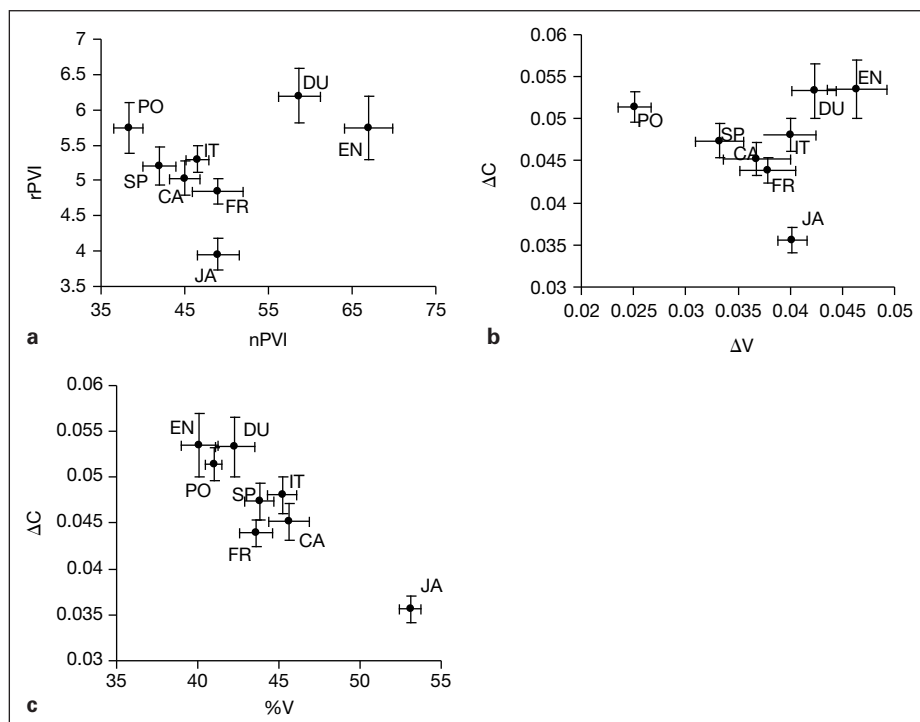


Fig. 1. PVI scores (**a**), $\Delta V - \Delta C$ scores (**b**) for the corpus of Ramus et al. [1999], as presented in Ramus [2002], and $\%V - \Delta C$ scores for the same corpus (**c**), as presented in Ramus et al. [1999]. CA = Catalan, DU = Dutch, EN = English, FR = French, IT = Italian, JA = Japanese, PO = Polish, SP = Spanish. Reproduced from Ramus [2002] and Ramus et al. [1999], with permission.

‘largely equivalent results’. It is worth noting, however, that any similarity obtains only if PVI scores are compared to $\Delta V - \Delta C$, not to $\%V - \Delta C$, the two metrics that according to Ramus et al. [1999] best classify languages rhythmically. The differences in classification that result from the choice of metrics are evident in figure 1; note, e.g., that, according to PVIs, Polish is close to languages classified as syllable-timed (fig. 1a), but, according to $\%V - \Delta C$, it is grouped with stress-timed English and Dutch (fig. 1c).

Ramus [2002] also took exception to Grabe and Low’s [2002] weak support for a categorical distinction. Ramus [2002, p. 116] argued that a continuum can be supported only if languages are ordered on it according to their accepted classifications; otherwise, he rightly pointed out, a random continuum with 18 data points is ‘the simplest thing in the world’ to obtain. Yet, given the expected scores for stress- and syllable-timed languages – high PVIs and ΔC for the former and low for the latter, low $\%V$ for the former and high for the latter – one should expect a correlation between vocalic and consonantal scores (which would arise whether the data formed a true continuum or two distinct clusters). However, as shown clearly in figure 2, neither PVIs nor $\%V - \Delta C$ show such a correlation, rendering moot the issue of language ordering on the continuum (for nPVI and rPVI, $r = 0.019$; for $\%V$ and ΔC , $r = -0.26$) [correlations calculated by the author using the scores of Grabe and Low, 2002, pp. 544–545].

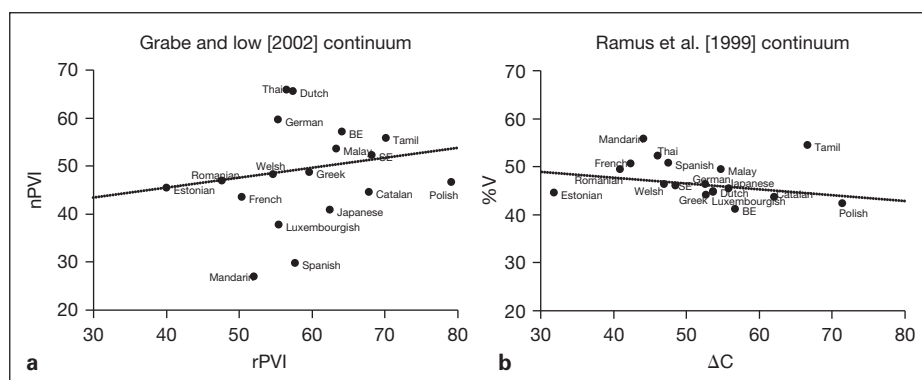


Fig. 2. **a** Scatterplot and regression line for the rPVI and nPVI scores of the 18 languages tested by Grabe and Low [2002]. **b** Scatterplot and regression line for ΔC and %V scores of the same data. Based on the data of Grabe and Low [2002], produced with permission.

Results casting further doubt on the reliability of metrics are currently being analyzed in the author's laboratory (in collaboration with Tristie Ross and Naja Ferjan). Briefly, in this study, three sets of metrics are being tested, %V – ΔC , PVIs, and Varcos (variation coefficients or measures of relative variation, proposed by Dellwo [2006] and also employed by White and Mattys [2007] for rhythmic classification). The languages examined include English and German, which are traditionally classified as stress-timed, Italian and Spanish, which are traditionally classified as syllable-timed, and Greek and Korean, the classification of which is unclear. Greek is unclassifiable, according to Grabe and Low [2002], while other studies suggest it is syllable-timed [Barry and Andreeva, 2001] or has mixed rhythm [Baltazani, 2007; for a review see Arvaniti, 2007]. Similarly uncertain is the classification of Korean, which has been said to be syllable-timed, stress-timed or intermediate between the two [see Lee et al., 1994, for a review and data that show a change in progress]. Data from these six languages were collected using three elicitation methods: read sentences, as in Ramus et al. [1999], read running text (the story of 'The North Wind and the Sun'), as in Grabe and Low [2002], and spontaneous speech. The sentence corpora in particular were divided into three sets of five sentences each: one subset was designed to be as 'syllable-timed' as possible (showing simple consonant-vowel alternations), the second subset was as 'stress-timed' as possible (showing as much segmental variability as was feasible for each language), while the third subset consisted of uncontrolled sentences selected from the writings of well-known authors of the languages investigated (e.g., F. Scott Fitzgerald for English, Gabriel García Márquez for Spanish); English examples are shown in table 1.

It was hypothesized that the elicitation method could have two possible outcomes. If running speech shows similar reduction patterns in languages of different rhythmic types [Barry and Andreeva, 2001], score differences among languages should be minimized in running speech, where languages would exhibit comparable degrees of durational variability. On the other hand, if rhythmic classifications are correct, then score differences should be maximized in spontaneous speech, which should best reflect different rhythmic patterns. Regarding the three sentence subsets, it was expected that their differences in segmental structure could affect metric scores, since metric scores measure durational variability. In particular it was expected that 'stress-timed'

Table 1. Examples of ‘stress-timed’, ‘syllable-timed’ and uncontrolled sentences from the English materials of the author’s study in progress

Type	Example sentences
‘Stress-timed’	<i>The production increased by three fifths in the last quarter of 2007</i>
‘Syllable-timed’	<i>Lara saw Bobby when she was on the way to the photocopy room</i>
Uncontrolled	<i>I called Gatsby’s house a few minutes later, but the line was busy</i>

materials would show more ‘stress-timed’ scores, that ‘syllable-timed’ materials would show more ‘syllable-timed’ scores, and that uncontrolled materials would show scores in between those extremes. Alternatively, if the classification of languages into distinct rhythmic classes is correct, then the scores of uncontrolled materials should pattern with the scores of the materials that are most representative of each language’s rhythm; e.g., uncontrolled English materials should yield scores similar to the English ‘stress-timed’ corpus, while uncontrolled Italian materials should yield scores similar to the Italian ‘syllable-timed’ corpus.

The results from 5 speakers of each language so far agree with the results of previous studies using metrics in that they show similar discrepancies between metric scores and rhythmic classification; further, they highlight some additional problems. First, these results show that many cross-linguistic score differences are not statistically significant; e.g., VarcoV scores pooled over elicitation methods show no significant effect of language. The lack of statistically significant results could be due to the high inter-speaker variability in the data; this is illustrated in figure 3, where it can clearly be seen that the scores of individual speakers of each language do not form distinct groups. In addition, when differences are present, they do not always pattern in the expected direction and the patterns are not the same for all metrics (fig. 4). Thus, for %V, scores are lower in German than in Italian, Korean, Spanish and Greek (pairwise comparisons are based on Tukey HSD post-hoc tests; $p < 0.05$ for all reported differences); this also applies to the comparison of English with Italian, Korean, and Greek, but the English %V score is not statistically distinct from the Spanish score. Regarding nPVI, the English score is higher than that of all the other languages tested, but the German score is not statistically different from that of Korean or Greek. The intervocalic metrics, ΔC , VarcoC and rPVI, all show scores that are *generally* higher for German and English than for the other four languages, but the English VarcoC score is not different from that of Italian, and the German rPVI score is not different from that of either Korean or Greek. As a result of these differences among scores (or the lack thereof), the three sets of metrics yield different rhythmic classifications: %V – ΔC and the PVI’s separate English and German from the other four languages, but %V – ΔC suggests that German is more stress-timed than English while the PVI’s suggest the opposite (fig. 4). Varcos, on the other hand, do not show a clear separation of languages into two groups, while the Varco scores for Greek and Korean are practically identical. Euclidean distances calculated with English as the point of reference (table 2) support this interpretation of the results.

Additionally, the results presented here show that the elicitation method can significantly affect metric scores in such a way that rhythmic classifications can change depending on the type of data on which scores are calculated. This is illustrated in

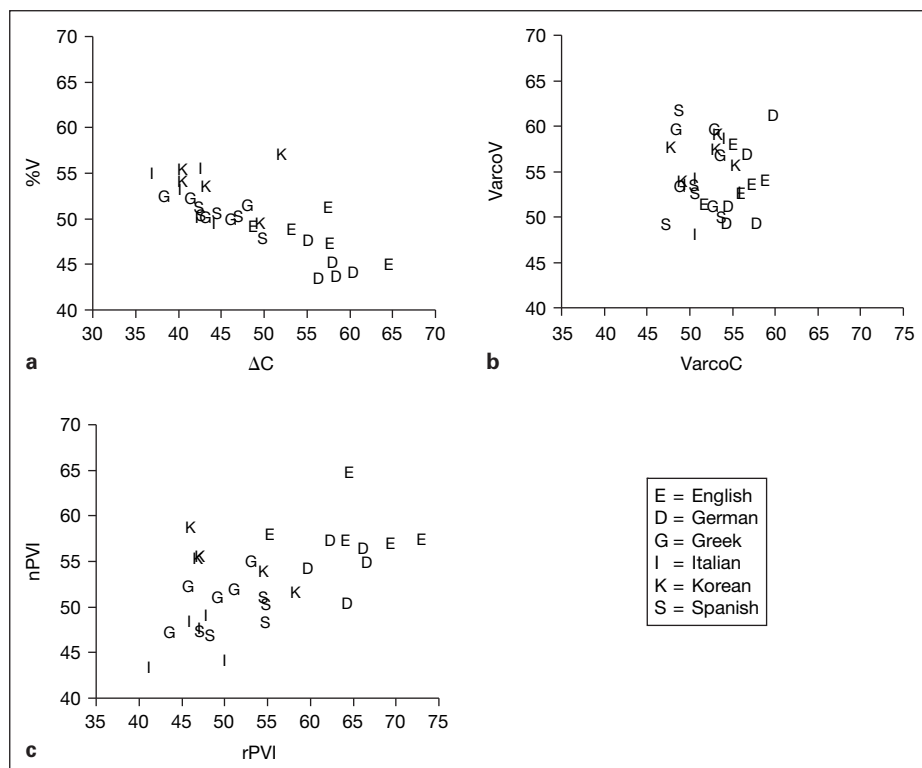


Fig. 3. **a** Scatterplot of %V – ΔC scores separately for each speaker of each language in the author's study ($n = 5$ for each language). **b, c** Scatterplots of Varco and PVI scores for the same data.

figure 5, which shows metric scores calculated on the data from spontaneous speech (for which scores differed substantially from the scores of the two types of read speech); e.g., these data show much greater separation of English and German for PVIs and Varcos than the pooled data presented in figure 4. It is important to note, however, that many of these differences in rhythmic classification are not supported by statistical differences among spontaneous speech scores (possibly due to greater intraspeaker variability than what was found in read speech). Specifically, none of the cross-language comparisons of vocalic scores was statistically significant. Intervocalic scores do not fare substantially better: VarcoC shows no statistically significant differences between languages; for ΔC , only the score of Italian is lower than the scores of English and German; for rPVI, the score of English is higher than those of Italian, Spanish and Greek, and the score of German is higher than that of Italian, but all other pairwise comparisons failed to reach significance.

Most disturbingly perhaps, metric scores can be affected by the choice of materials, so that, independently of a language's accepted rhythmic type, more 'stress-timed' materials can yield scores that are closer to those of stress-timed languages and more 'syllable-timed' materials can yield scores closer to those of syllable-timed languages. This effect is illustrated in figure 6, which presents %V – ΔC scores for English,

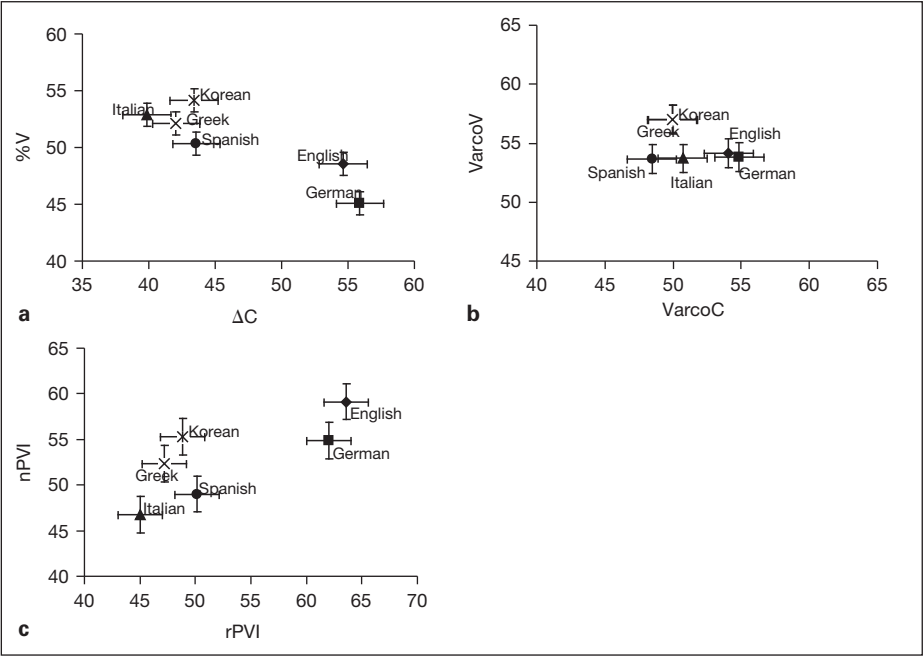


Fig. 4. **a** %V – ΔC scores for English, German, Italian, Korean, Greek and Spanish pooled over the entire corpus in the author’s study. **b, c** PVI and Varco scores for the same corpus.

Table 2. Euclidean distances of metric scores, pooled over all materials, using English as the point of reference (data from the author’s study in progress)

Metric	German	Italian	Korean	Spanish	Greek
%V – ΔC	3.8	15.4	12.6	11.2	13.1
Varcos	0.9	3.4	5.0	5.7	5.0
PVIs	1.5	20.2	14.8	14.6	16.6

Larger values indicate that a language is rhythmically less similar to English than smaller values.

Spanish and Greek, separately for each sentence subset. As can be seen, the scores of uncontrolled English materials and ‘stress-timed’ Spanish materials are virtually identical, while the scores of ‘stress-timed’ Greek materials and uncontrolled Spanish materials are very similar to each other, and quite close to the score of the ‘syllable-timed’ English materials. The effect of the materials on the scores of each of these languages can also be clearly seen in figure 6, where the spread of scores is comparable to that found in the scores of the eight languages examined by Ramus et al. [1999] and illustrated in figure 1c (for English ΔC , stressed-timed > syllable-timed, uncontrolled; for Spanish ΔC , stressed-timed > syllable-timed, uncontrolled, and uncontrolled > syllable-timed; for Greek ΔC , stressed-timed > syllable-timed, uncontrolled; for English

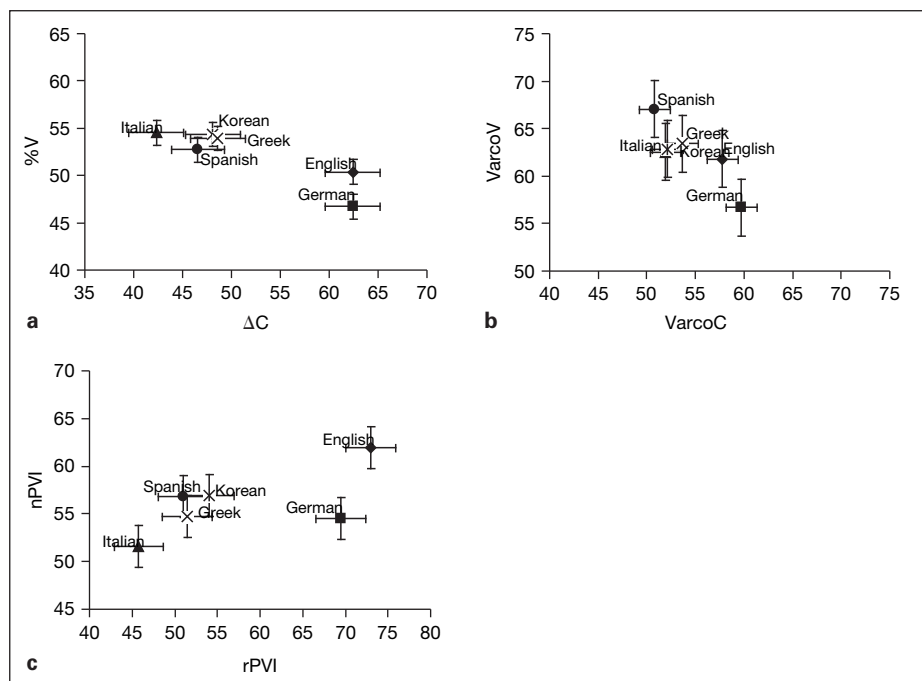


Fig. 5. **a** %V – ΔC scores for English, German, Italian, Korean, Greek and Spanish calculated over the spontaneous speech data from the author's study. **b, c** PVI and Varco scores for the same corpus.

%V, stressed-timed < syllable-timed, and syllable-timed < uncontrolled; for Spanish %V, stressed-timed < syllable-timed; for Greek %V, stressed-timed < syllable-timed).

Overall this review of results from previous studies and the study in progress briefly presented here shows that rhythmically prototypical languages can be separated along some dimension with varying degrees of success [Ramus et al., 1999; Grabe and Low, 2002; Lee and McAngus Todd, 2004; Rouas et al., 2005; White and Mattys, 2007]. On the other hand, the study of nonprototypical languages – such as Latvian [Stockmal et al., 2005], Tamil [Keane, 2006], Greek [Baltazani, 2007; the present study], Korean [Lee et al., 1994; the present study] and Bulgarian [Barry et al., 2003] seriously questions the ability of metrics to rhythmically classify all languages [for similar observations, see also Barry et al., 2003]. Finally, the elicitation and corpus manipulations presented here cast serious doubt on the overall robustness of metrics when faced with different types of materials: score differences among languages disappear in spontaneous speech, while the design of read materials can significantly affect scores and thus a language's rhythmic classification.

3. Reasons for the Lack of Metric Success

The poor results of metrics suggest that the conceptualization of rhythm they rest on may be flawed. In order to understand if this is so, it is essential to examine the

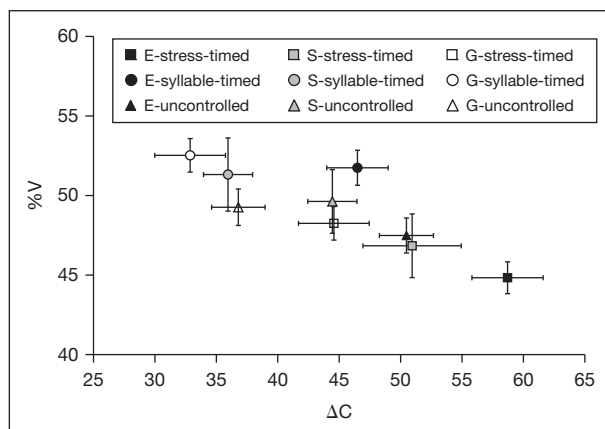


Fig. 6. %V – ΔC scores for English, Spanish and Greek separately for each sentence subset in the author's study. E = English, G = Greek, S = Spanish.

reasons behind the lack of metric success by scrutinizing their theoretical foundation. This foundation, according to Ramus et al. [1999] and Grabe and Low [2002], is the rhythmic classification parameters presented in Dauer [1983, 1987]. Thus, it is worth considering the extent to which these parameters are represented by metrics, but also, more generally, the ability of Dauer's [1983, 1987] parameters themselves to classify languages in terms of rhythm.

Dauer [1987] provided a list of eight parametric criteria to place languages on a rhythmic continuum. Crucially, her criteria were not meant to reflect durational variation per se but, rather, the extent to which a language has easily defined prominences [*stresses* in the terminology of Dauer, 1983; *accents* in that of Dauer, 1987]. This is evident in the fact that Dauer's list is not purely phonetic and includes both parameters that can be argued to be directly reflected in phonetic timing (e.g., the presence or absence of durational differences between stressed and unstressed syllables), and others where the relationship with timing, if existent, is much more indirect (e.g., the function of pitch in the language or the relationship, if any, between tone and stress in tone and pitch accent languages).

Although Dauer's [1987] criteria are now generally accepted, they have not been rigorously tested to see if they can provide a reliable rhythmic classification for a variety of languages. To the extent they have been tested, they appear to be successful at rhythmically distinguishing prototypical languages, but, like metrics, they falter when faced with less prototypical cases. For instance, English scores a plus for six of Dauer's eight criteria, while French scores five minuses for the same criteria. These scores place English and French at practically opposite ends of the continuum, but there is no compelling reason why Dauer's criteria will work synergistically in all languages [see, e.g., Barry et al., 2003, for a discussion of Dauer's criteria with respect to Bulgarian and German]. Indeed, studies show that languages said to belong to different rhythmic types do not show different patterns of reduction in running speech, as Dauer's criteria would lead one to expect [Roach, 1982; Barry and Andreeva, 2001]. In some cases, the

evidence even contradicts the criteria. Greek and Spanish are a case in point: Dauer [1983] places them toward the middle of her continuum, suggesting that stress should be rather weak in these languages. Yet Dauer [1983, p. 58] herself noted that Greek, Spanish and Italian have a ‘clearly discernible “beat”’, the main feature of stress-timed languages in her view, even though they have not been labeled stress-timed. Indeed studies show that stress has robust acoustic correlates in all three languages [on Greek, Botinis, 1989; Arvaniti, 1994, 2000; on Italian, Farnetani and Kori, 1990; D’Imperio and Rosenthal, 1999; on Spanish, Ortega-Llebaria and Prieto, 2007]. Thus the rhythmic classification of Italian, Spanish and Greek, at least, is at odds with the strength and importance of stress in these languages, and in turn these features are at odds with Dauer’s [1983, 1987] own conception of rhythmic types as resting on the strength of local prominences.

Given the mediocre success of Dauer’s criteria, it would be difficult for metrics to rhythmically classify nonprototypical languages, even if they did reflect all of the criteria. Even worse, metrics take into account only those criteria that are related to speech timing. In doing so, metrics are not much different from classic measures of isochrony, since they assume a simple and straightforward relationship between duration and abstract phonological categories, such as syllable structure, vowel weight and vowel reduction patterns.

But there are two problems with this assumption. First, it is well known that the duration of segments is affected by all sorts of factors. The most obvious, perhaps, is the presence of geminate consonants and/or of vocalic length distinctions in a language. But even if we consider only languages that do not use duration contrastively, several factors still influence segmental duration, including language-specific inherent durations, contextual effects due to voicing and syllabic position, stress-, accent- and focus-related lengthening, contextually and prosodically determined vowel reduction, phrase-final lengthening, allophonic changes of both consonants and vowels in stressed and unstressed syllables, and many more [for reviews, see Klatt, 1976, for English; Arvaniti, 2007, for Greek; Turk and Shattuck-Hufnagel, 2000, for cross-linguistic data]. Unstressed high vowel reduction in Greek illustrates this complexity. Since Dauer [1980], reduction has been assumed to be influenced by local context, namely the presence of abutting voiceless consonants. Yet, Tserdanelis [2005] convincingly showed that reduction is strongly affected by prosodic position: phrase-initial high vowels are elided, but phrase-final ones, found in the same local context, do not elide if they have to carry the intonational tune. Examples like this strongly suggest that rhythm metrics can at best provide crude measures of speech timing and variability; but they cannot reflect the origins of the variation they measure and thus they cannot convey an overall rhythmic impression (if we assume that rhythm strictly relies on timing, a point discussed in more detail below). This limitation of metrics is further illustrated by L2 results from the author’s study briefly presented earlier, which show that Korean and Spanish L2 speakers of English achieve similarly high scores of consonantal variability, which are not statistically different from those of L1 speakers of English. However, close scrutiny of the data reveals that these scores are due to very different strategies: Korean speakers show extreme phrase-final lengthening, while Spanish speakers show extreme lenition of intervocalic consonants. Although both practices contribute to consonantal variability and thus lead to high scores, their auditory effects are very different and could hardly be said to help speakers achieve a rhythm similar to that of L1 English.

Table 3. Differences in nPVI and rPVI scores between languages belonging to the same or different rhythmic types [after Grabe and Low, 2002]

	nPVI score differences	rPVI score differences
Stress-timed languages	$nPVI_{Th} - nPVI_{BE} = 8.6$	$rPVI_{BE} - rPVI_{Gm} = 8.8$
Syllable-timed languages	$nPVI_{Fr} - nPVI_{Mn} = 16.5$	$rPVI_{Jp} - rPVI_{Fr} = 12.1$
Languages of different types	$nPVI_{BE} - nPVI_{Fr} = 13.7$	$rPVI_{Sp} - rPVI_{Gm} = 2.4$
	$nPVI_{Gm} - nPVI_{Fr} = 16.2$	$rPVI_{Sp} - rPVI_{Th} = 1.2$

BE = British English, D = Dutch, Fr = French, Gm = German, Jp = Japanese, Mn = Mandarin, Sp = Spanish, Th = Thai.

Given the language-specific complexities of speech timing, it is unrealistic to expect that segmental duration will vary in a uniform manner in all languages or within rhythmic type. For instance, polysyllabic shortening applies both in English [e.g., Turk and Shattuck-Hufnagel, 2000] and Greek [Baltazani, 2007], but word-final and accentual lengthening apply only in the former [Turk and Shattuck-Hufnagel, 2000] but not in the latter [Botinis, 1989, chap. 2, on accentual lengthening; Arvaniti, 2000, on word-final lengthening]. Overall, then, it is not surprising that metrics fail to classify most languages. They are loosely based on a set of parameters that appear not to be very successful at rhythmic classification, and reflect but a small subset of these parameters that reduces rhythm to crude measures of timing.

It is tempting to assume that these problems can be resolved by devising more sophisticated metrics that take into account more factors affecting duration or normalize for unwanted effects such as speaking rate [as attempted by Frota and Vigário, 2001; Wagner and Dellwo, 2004; Dellwo, 2006]. However, the problem with metrics is more profound than the lack of an appropriate mathematical manipulation of duration. This is because, as mentioned, metrics equate rhythm with timing and by doing so the relationship between metrics and the definition of rhythm on which they rest is circular [see Bertinetto and Bertini, 2008, for similar remarks].

A result of this circularity is the lack of an independent measure of metric success and, by extension, of a clear method of interpreting differences and similarities between metric scores: typically, scores are judged to be similar or different depending on whether it is similarity or difference that supports the preexisting rhythmic classification of the languages under investigation. The interpretation of scores by Grabe and Low [2002] demonstrates this point: some score differences within a category can be as large as or even larger than differences between categories, while some of the differences do not go in the expected direction. Such results are illustrated in table 3: e.g. the difference in nPVI scores between French and Mandarin, both classified as syllable-timed by Grabe and Low [2002], is 16.5 points, while the nPVI difference between British English and French, the prototypical examples of stress- and syllable-timing, respectively, is 13.7 points; similarly, the rPVI difference between the scores of Spanish, which is classified as syllable-timed, and Thai, which is classified as stress-timed, is only 1.2 points, while the difference between Japanese and French, both classified as syllable-timed by Grabe and Low [2002], is 12.1 points. Similar discrepancies are found if Euclidean distances are calculated between scores, using English as the reference point. Thus, Euclidean distances between PVI scores suggest that Greek,

which is unclassified according to Grabe and Low [2002], is closer to English (9.6 points) than Dutch and Thai, which are both classified as stress-timed (10.7 and 11.5 points, respectively). Similarly, Euclidean distances calculated using the %V – Δ C scores presented in Grabe and Low [2002] for their set of data suggest that the language rhythmically closest to English is Japanese (4.5 points)!

Despite these inconsistencies, Grabe and Low's [2002] interpretation of their results relies squarely on preexisting classifications, not scores (since the scores are essentially uninterpretable without the preexisting classifications). This forces, e.g., Grabe and Low [2002] to conclude that Singapore English 'is not at all close to the traditional syllable-timed languages, French or Spanish' even though the nPVI score difference between French and Spanish is higher (13.8 points) than that between French and Singapore English (8.8 points), and the rPVI score difference between Spanish and Singapore English (10.5) is only three points larger than that between French and Spanish themselves (7.3 points).

4. Towards a New Conception of Rhythm

The above discussion clearly suggests that without independent criteria for language rhythm, it is not possible to establish an *independent* and *meaningful* measure of score similarity for any kind of rhythm quantification: independent criteria are required to avoid circularity and measures must be meaningful so that they avoid tautology and reflect differences in rhythm rather than purely statistical (or geometric) differences between scores.

The question that remains to be answered is whether such a quantification of rhythm is possible if we continue to rely on the notion of rhythmic categories and, by extension, on the equation of rhythm with timing. This is worth considering seriously because the phonetic conception of rhythm is psychologically questionable for a number of reasons. In psychological research, rhythm is defined as the perception of series of stimuli as series of groups of similar and repetitive pattern [Woodrow, 1951; Fraisse, 1963, 1982]. The grouping of stimuli relies not just on duration but on a host of factors, including relative intensity, relative and absolute duration and the temporal spacing of elements [Woodrow, 1951; Fraisse, 1963, 1982]. This definition of rhythm implies the presence of *meter*, which is distinguished from *grouping* itself: while grouping deals with phenomena that extend over time, meter is an abstract representation that relies on the alternation of strong and weak elements, not on absolute or relative durations [Lerdahl and Jackendoff, 1983].

A crucial implication of this conceptualization of rhythm is the implausibility of the notion of syllable-timing. Syllable-timing is often described as the regular repetition of syllables, that is, as a *cadence*, the simplest form of rhythm that 'is produced by the simple repetition of the same stimulus at a constant frequency' [Fraisse, 1982, p. 151]. One problem with this conceptualization of syllable-timing is the role of temporal spacing in the creation of rhythm. Specifically, in order for rhythm to be perceived, events must be sufficiently separated in time to be experienced as distinct entities – i.e., for *fusion* not to take place – but close enough that a number of them are perceived as occurring during the 'psychological' or perceptual present, so that grouping is achieved [Woodrow, 1951; Fraisse, 1963, 1982]. The psychological present is defined as 'the temporal extent of stimulations that can be perceived at a given time, without the intervention of rehearsal

during or after the stimulation [Fraisse, 1987, p. 205, cited in Clarke, 1999, p. 474] and is estimated to span 3–8 s [Clarke, 1999]. Now, for sounds, temporal spacing is expected to be at least 200 ms for fusion to be avoided [Fraisse, 1982]. But the typical speaking rate of many languages classified as syllable-timed, e.g., Greek, Spanish, Italian, is much faster than that. In Dauer [1983], Spanish, Greek and Italian had a pooled average speaking rate of 7.3 syllables/s. Even at this rate – which is probably rather low for these languages as it is based on laboratory speech [note, e.g., that Baltazani, 2007, reports a rate of 9.9–10.3 syllables/s in Greek] – syllables appear, on average, every 136 ms. This spacing suggests that the creation of a cadence, at least in these three languages, would be difficult at best.

But syllable-timing is also problematic for another reason: research shows that subjects presented with series of identical stimuli tend to impose rhythmic structure on them, by hearing some stimuli as more prominent than others (typically grouping stimuli into trochees or iambs, and larger constituents based on these basic units, with a slight preference for trochees) [Woodrow, 1951; Fraisse, 1963, 1982]; e.g., a typical cadence, such as the beating of a clock, is often interpreted as a series of trochees. It is difficult to reconcile this response with the equal prominence of all syllables expected in syllable-timed languages. The idea that all syllables are *perceived* as being the same is psychologically improbable by definition, as just noted. Alternatively, if we assume that syllable-timing is based on the more or less isochronous *production* of syllables, then syllable-timing suggests that speakers strive to produce an acoustic effect that their listeners will discard. But even if the notion of syllable-timing could be defended in some way, it still remains unsatisfactory: it implies that the speakers of some languages construe linguistic rhythm along different principles than in the rest of their experience with timing patterns, and along different principles than the speakers of stress-timed languages, since stress-timing relies on grouping and the alternation of more and less prominent syllables (in that feet contain both types), while syllable-timing does not.

The implausibility of syllable-timing suggests that models of rhythm may be more successful if they rely on the same psychologically plausible notions of grouping and relative prominence in all languages. To an extent, this is what Dauer [1983] was attempting to do. Dauer [1983] suggested that stress is the basis of rhythm in all languages, and for that reason proposed a rhythmic continuum that does not stretch from syllable- to stress-timing – as is sometimes claimed [e.g., Ramus et al., 1999, p. 269] – but, rather, from least to most stress-based.

Dauer's own data from several syllable-timed languages support her view. Dauer [1983] noted that stresses appear neither more regularly nor more frequently in stress-timed English, in which foot structure is meant to be regular, than in Spanish, Greek or Italian, in which no such tendencies are expected on the basis of rhythmic type. In her data, cross-linguistic comparisons of interstress interval durations showed no statistical differences; means ranged from 380 ms for Thai to 493 ms for English, with Italian, Spanish and Greek showing intermediate values (468 ms for Italian, 477 ms for Spanish, 483 ms for Greek). In other words, in these languages, stressed syllables appear approximately every half second, thereby allowing for 6–10 beats within the psychological present. This rate is not only sufficient for the perception of rhythm, it is in fact very close to the *preferred tempo* – the spacing of events every 0.5–0.6 s – that is judged by subjects as neither slow nor fast and is the pace most likely to induce good synchronization between external aural stimuli and human tapping, or *natural tempo* [Fraisse, 1982; Dowling and Harwood, 1986; Clarke, 1999].

It is important to note that this regularity may not be present in the acoustic signal at all times; e.g., interstress intervals in Dauer [1983] ranged from 300 to 700 ms. Yet, rhythmic patterning is still possible as a result of *anticipation* [Fraisie, 1982], the tendency to continue hearing a rhythmic pattern, in the absence of regularity, once the pattern is established. This psychological propensity has been successfully tested with linguistic materials by Dilley and McAuley [2008]. They showed that listeners were more likely to perceive the last two words in a sequence, such as *long hand shake*, as a compound (*handshake*), if the rhythm and/or intonation of a preceding list (*worthy vinyl life*) predisposed them to create trochees and thus link *life* with *long* (giving *worthy vinyl lifelong handshake*); if the established pattern predisposed them to put a break after *life*, then listeners preferred the grouping *longhand shake* instead.

In part, anticipation works because listeners do not require perfect regularity in the signal in order to extract rhythmic patterning. In order to capture this, psychologists make a distinction between (expressive) *timing* and rhythm: timing is concerned with the durational characteristics of events, while rhythm has to do with the pattern of periodicities that is extracted from these durations; this pattern can be arrived at, despite durational variation, because listeners assign durations to categories, while at the same time being sensitive to the nuances conveyed by the variation itself [Clarke, 1999, and references therein]. Although the distinction between timing and rhythm has been used primarily in research on music, the analogy with speech is evident, if we assume a conception of rhythm closer to Dauer's: listeners extract the rhythmic pattern of an utterance by classifying various prosodic entities (syllables, feet, phrases) as more or less prominent, on the basis of durational (and other) clues, while simultaneously noting that the speaker tends to draw out her stressed vowels, or 'swallow' the unstressed ones. The above suggests that durational variability may play only a small role in the creation of rhythm and that, therefore, the importance attributed to it in phonetic studies of rhythm may well be overestimated (note, for instance, that metrics lump together perceptible durational differences and those below the just noticeable difference of 10% or so) [Lehiste, 1970]. At the same time, the sensitivity of listeners to durational variation may explain the discrimination results obtained with rhythmically prototypical languages by Nazzi et al. [1998, 2000] and Nazzi and Ramus [2003].

It is also essential to recognize that the relatively regular spacing of prominences noted by Dauer [1983] is achieved in language-specific ways, such as differences in speaking rate and patterns of reduction [on the latter, see Barry and Andreeva, 2001, and Barry et al., 2003]. For example, in Dauer [1983], the average speaking rate for English was 5 syllables/s, while for Spanish, Greek and Italian the averages were 7.1, 7.5 and 7.3 syllables/s, respectively; t-tests run by the present author on Dauer's [1983] data show that the English rate was significantly lower than those of the other three languages ($p < 0.05$ in all cases). Independently of obvious reasons for these differences – such as vowel elision [Arvaniti, 1994, 2007] and syllable complexity, amply discussed in Dauer [1983] – the combination of relatively stable interstress intervals and different speaking rates suggests that languages like Greek, Spanish or Italian include more syllables in each interval, an observation confirmed by Dauer [1983]. Thus, one difference between languages called stress-timed and those called syllable-timed may have to do with the spacing of prominences, not in terms of duration but in terms of number syllables; in this respect, prominences may be sparser in syllable-timed languages. In turn, this difference could result in different rhythmic hierarchies and different degrees of flexibility in keeping prominences regular [Arvaniti, 1994, 2007].

The above discussion strongly suggests that Dauer [1983] was on the right track. One weakness of her reasoning was that, inevitably perhaps, it was still anchored on timing. She states, e.g., that in Spanish stresses are less prominent than in English because in the former the durational ratio between stressed and unstressed syllables is 1.3, while in the latter it is 1.5. Crucially, this statement reflects an additional assumption that is worth considering in some detail, namely that the correlates of prominence can be determined a priori and used as a guide to classifying the strength of a language's prominences. Yet research clearly shows that the notion of stress or prominence does not rest on one (or even a set of) universal acoustic parameter(s), i.e., stress does not have its own phonetic content [Beckman and Edwards, 1994, p. 14]. Languages use different parameters to make some syllables more prominent than others, and the same parameter may be allocated different degrees of importance in different languages. For instance, Sluijter and van Heuven [1996] show that spectral tilt is the primary acoustic parameter of stress in Dutch, while in English, spectral tilt differentiates only accented from unaccented (but stressed) syllables [Campbell and Beckman, 1997].

Crucially, such differences in the encoding of prominence result in different percepts of prominence for the same stimuli, depending on the native language of the listeners. Beckman [1986] has shown that English and Japanese listeners rely on their native set of parameters whether processing English or Japanese. Similarly, de Jong [1994], who asked speakers of English, Japanese and Korean to assign prominence to one syllable in Korean accentual phrases, found that each group responded differently: Korean listeners showed a preference for the second syllable, but also significant uncertainty; Japanese listeners strongly preferred the second syllable (which, in Korean, usually co-occurs with the F_0 peak of each accentual phrase), while English listeners paid equal attention to durational and F_0 differences. In other words, each group relied on the prosodic parameters that indicate prominence in their language; as a result, syllables that sounded prominent to one group were not perceived to be so by another.

These different ways of perceiving prominences in languages other than our own implies that rhythmic classification either on impressionistic grounds or along some acoustic dimension is bound to fail. Even trained phoneticians cannot easily transcend the way in which parameters such as duration, F_0 , amplitude and vowel quality function prosodically in their native language. This idea was first broached in Roach [1982], who suggested that English speakers hear French as syllable-timed because they are unaccustomed to full vowels in nonprominent syllables, and was amply supported by Miller [1984], who asked English and French phoneticians and naïve speakers to rhythmically classify Arabic, Yoruba, Polish, Spanish, Finnish, Japanese and Indonesian. Now, if there was any independent and objective basis to the stress-/syllable-timing distinction one would expect that the prototypical languages, at least, would be similarly classified by all four groups. Yet few of Miller's [1984] results reached statistical significance (most suggested that the subjects were guessing), and of those that did only the results for Arabic were unanimous (all groups classified it as stress-timed). In addition, Yoruba and Japanese were classified as syllable-timed by English phoneticians and nonphoneticians, respectively, while both groups of French listeners classified Indonesian as syllable-timed and Spanish as stress-timed. Results like these illustrate the futility of trying to classify languages in the manner so far employed, and explain why many native phoneticians, particularly of languages classified as syllable-timed, disagree with their language's rhythmic classification [see Dauer, 1983, and references therein].

On the whole, the above review suggests that we need to reconsider our view of speech rhythm so that it is less focused on timing, which should in principle be examined as a distinct phenomenon. Instead, it appears advantageous to adopt a conception of rhythm that goes beyond timing and rhythmic types but rests instead on grouping and patterns of prominence. In this respect, connecting phonetic research to models of rhythm that are widely accepted in phonology [e.g., Hayes, 1995] and closer to the psychological understanding of rhythm may also be beneficial.

This alternative view of rhythm advanced here has several implications for future research. First, it clearly suggests that focusing exclusively on durational measurements – and in particular on relative or absolute durational variability as metrics do – is misguided: as noted, local durational contrasts may not be as important for rhythm as the induction of meter, the abstract patterns of periodicity that may be the only mental representation of rhythm [Lerdahl and Jackendoff, 1983; Clarke, 1999, and references therein]. Further, the contribution of parameters other than duration to the creation of rhythm should be examined, since their role could be substantial [for a demonstration of the role of pitch and intonation, e.g., see Dilley and McAuley, 2008; Kohler, 2008]. Nevertheless, it should be understood that the role of parameters is language-specific and thus that some of them may contribute to rhythm in some languages but not in others. In this respect, it is imperative to place greater focus on nonprototypical languages and in particular to test the ideas advanced here with languages that challenge them, such as Korean, the native speakers of which do not have strong intuitions about prominence. Further this alternative conceptualization suggests that more studies are needed on the perception of rhythm rather than simply on its acoustic manifestation. Few studies so far have acknowledged this need [but see Beckman, 1992; Warner and Arai, 2001] and fewer have presented results that tap onto the perception of rhythm; those that do, such as Scott et al. [1985], support the ideas advanced here. Finally, native speakers should not be ignored, as has been done so far; rather, the ways they experience rhythm in their language should be seriously investigated, since it is evident that percepts of rhythm are not objective but language-dependent.

Conclusion

In order to understand rhythm it is necessary to decouple the quantification of timing from the study of rhythmic structure. Timing reflects segment durations and their interactions with a host of factors. Clearly timing is part of the study of rhythm, but it should not be the only part (and, in any case, it is not adequately quantified by any metrics proposed so far). The decoupling of rhythm and timing means abandoning the notion of rhythmic categories, in all its guises, since it rests squarely on timing and is, in addition, psychologically questionable. Instead, Dauer's idea that rhythm is stress-based in all languages is worth revisiting, provided a more liberal view of stress as prominence is adopted and the role of prominence for rhythm creation in a wide range of languages is investigated. The adoption of a conception of rhythm as the product of prominence and patterning is psychologically plausible and does not rely on a dubious and ultimately unsuccessful division between languages or the measuring of timing. However, in order to be successful, such a conception would also require that acoustic measurements be adapted to the prosody of each language and that the perception native speakers have of their language's rhythm be at last taken into account.

Acknowledgments

I warmly thank my students Tristie Ross and Naja Ferjan, who have worked tirelessly to gather and measure the data presented here and frequently discussed the ideas expressed in this article with me. Thanks are also due to Kris Phillips for writing the script to calculate rhythm metrics, to Nancy Gill, Nayoung Kwon, Sun-Ah Jun and Ken de Jong for help with the Korean materials, and to Klaus Kohler, Mary Beckman, and two anonymous reviewers for their helpful comments and suggestions.

References

- Abercrombie, D.: Elements of general phonetics (Edinburgh University Press, Edinburgh 1967).
- Arvaniti, A.: Acoustic features of Greek rhythmic structure. *J. Phonet.* 22: 239–268 (1994).
- Arvaniti, A.: The phonetics of stress in Greek. *J. Greek Ling.* 1: 9–38 (2000).
- Arvaniti, A.: Greek phonetics: the state of the art. *J. Greek Ling.* 8: 97–208 (2007).
- Baltazani, M.: Prosodic rhythm and the status of vowel reduction in Greek. 17th Int. Symp. on Theoret. and Appl. Ling., vol. 1, pp. 31–43 (Department of Theoretical and Applied Linguistics, Thessaloniki 2007).
- Barry, W.; Andreeva, B.: Cross-language similarities and differences in spontaneous speech patterns. *J. int. phonet. Ass.* 31: 51–66 (2001).
- Barry, W.J.; Andreeva, B.; Russo, M.; Dimitrova, S.; Kostadinova, T.: Do rhythm measures tell us anything about language type? *Proc. 15th ICPHS, Barcelona 2003*, pp. 2693–2696.
- Beckman, M.: Evidence for speech rhythms across languages; in Tohkura, Vatikiotis-Bateson, Sagisaka, Speech perception, production, and linguistic structure, pp. 458–463 (Ohmsha, Tokyo 1992).
- Beckman, M.E.: Stress and non-stress accent (Foris, Dordrecht 1986).
- Beckman, M.E.; Edwards, J.: Articulatory evidence for differentiating stress categories; in Keating, Phonological structure and phonetic form. *Papers in Laboratory Phonology*, vol. III, pp. 7–33 (Cambridge University Press, Cambridge 1994).
- Bertinetto, P.M.: Reflections on the dichotomy ‘stress’ vs. ‘syllable-timing’. *Revue Phonétique appl.* 91–93: 99–130 (1989).
- Bertinetto, P.M.; Bertini, C.: On modeling the rhythm of natural languages. *Proc. Speech Prosody, Campinas 2008*.
- Bolinger, D.L.: Pitch accent and sentence rhythm; in Abe, Kanekiyo, Forms of English: accent, morpheme, order, pp. 139–180 (Harvard University Press, Cambridge 1965).
- Botinis, A.: Stress and prosodic structure in Greek: a phonological, acoustic, physiological and perceptual study (Lund University Press, Lund 1989).
- Campbell, N.; Beckman, M.: Stress, prominence, and spectral tilt. *Proc. ESCA Workshop, Athens 1997*, pp. 67–70 (ESCA and University of Athens, Athens 1997).
- Clarke, E.F.: Rhythm and timing in music; in Deutsch, The psychology of music, pp. 473–500 (Academic Press, New York 1999).
- Coetzee, A.W.; Wissing, D.P.: Global and local durational properties in three varieties of South African English. *Linguistic Rev.* 24: 263–289 (2007).
- Cutler, A.; Mehler, J.; Norris, D.; Seguí, J.: The syllable’s differing role in the segmentation of French and English. *J. Mem. Lang.* 25: 385–400 (1986).
- Cutler, A.; Mehler, J.; Norris, D.; Seguí, J.: The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychol.* 24: 381–410 (1992).
- Cutler, A.; Otake, T.: Mora or phoneme? Further evidence for language-specific listening. *J. Mem. Lang.* 33: 824–844 (1994).
- Dauer, R.M.: The reduction of unstressed high vowels in Modern Greek. *J. int. Phonet. Ass.* 10: 17–27 (1980).
- Dauer, R.M.: Stress-timing and syllable-timing reanalyzed. *J. Phonet.* 11: 51–62 (1983).
- Dauer, R.M.: Phonetic and phonological components of language rhythm. *Proc. 11th ICPHS, Tallinn 1987*, pp. 447–449.
- de Jong, K.: Initial tones and prominence in Seoul Korean. *OSU Working Papers in Linguistics* 43: 1–14 (1994).
- Dellwo, V.: Rhythm and speech rate: a variation coefficient for deltaC; in Karnowski, Szigeti, Language and language processing. *Proc. 38th Linguistic Colloquium*, pp. 231–241 (Lang, Frankfurt 2006).
- Dilley, L.C.; McAuley, J.D.: Distal prosodic context affects word segmentation and lexical processing. *J. Mem. Lang.* 59: 294–311 (2008).
- D’Imperio, M.; Rosenthal, S.: Phonetics and phonology of main stress in Italian. *Phonology* 16: 1–28 (1999).
- Dowling, W.J.; Harwood, D.L.: Music cognition (Academic Press, Orlando 1986).
- Farnetani, E.; Kori, S.: Rhythmic structure in Italian noun phrases: a study of vowel durations. *Phonetica* 47: 50–65 (1990).
- Fraisse, P.: The psychology of time (Harper and Row, New York 1963).
- Fraisse, P.: Rhythm and tempo; in Deutsch, The psychology of music, pp. 149–180 (Academic Press, New York 1982).

- Fraisse, P.: A historical approach to rhythm as perception; in Gabriëlssohn, Action and perception in rhythm and music, pp. 7–18 (Royal Swedish Academy of Music, Stockholm 1987).
- Frota, S.; Vigário, M.: On the correlates of rhythmic distinctions: the European/Brazilian Portuguese case. *Probus* 13: 247–275 (2001).
- Grabe, E.; Low, E.L.: Acoustic correlates of rhythm class; in Gussenhoven, Warner, Laboratory Phonology, vol. 7, pp. 515–546 (Mouton de Gruyter, Berlin 2002).
- Hayes, B.: Metrical stress theory: principles and case studies (University of Chicago Press, Chicago 1995).
- Keane, E.: Rhythmic characteristics of colloquial and formal Tamil. *Lang. Speech* 49: 299–332 (2006).
- Klatt, D.H.: Linguistic uses of segmental duration in English: acoustic and perceptual evidence. *J. acoust. Soc. Am.* 59: 1208–1221 (1976).
- Kohler, K.: The perception of prominence patterns. *Phonetica* 65: 257–269 (2008).
- Lee, C.S.; McAngus Todd, N.P.: Towards an auditory account of speech rhythm: application of a model of the auditory ‘primal-sketch’ to two multi-language corpora. *Cognition* 9: 225–254 (2004).
- Lee, H.B.; Jin, N.; Seong, C.; Jung, I.; Lee, S.: An experimental phonetic study of speech rhythm in Standard Korean. *Proc. ICSLP, Yokohama 1994*, pp. 1091–1094.
- Lehiste, I.: Suprasegmentals (MIT Press, Cambridge 1970).
- Lehiste, I.: Isochrony reconsidered. *J. Phonet.* 5: 253–263 (1977).
- Lerdahl, F.; Jackendoff, R.: A generative theory of tonal music (MIT Press, Cambridge 1983).
- Low, E.L.; Grabe, E.; Nolan, F.: Quantitative characterisations of speech rhythm: ‘syllable-timing’ in Singapore English. *Lang. Speech* 43: 377–401 (2000).
- Miller, M.: On the perception of rhythm. *J. Phonet.* 12: 75–83 (1984).
- Murty, L.; Otake, T.; Cutler, A.: Perceptual tests of rhythmic similarity. I. Mora rhythm. *Lang. Speech* 50: 77–99 (2007).
- Nakatani, L.H.; O’Connor, K.D.; Aston, C.H.: Prosodic aspects of American English speech rhythm. *Phonetica* 38: 84–106 (1981).
- Nazzi, T.; Bertoni, J.; Mehler, J.: Language discrimination by newborns: toward an understanding of the role of rhythm. *J. exp. Psychol. hum. Percept. Perform.* 24: 756–766 (1998).
- Nazzi, T.; Jusczyk, P.W.; Johnson, E.K.: Language discrimination by English-learning 5-month-olds: effects of rhythm and familiarity. *J. Mem. Lang.* 43: 1–19 (2000).
- Nazzi, T.; Ramus, F.: Perception and acquisition of linguistic rhythm by Infants. *Speech Commun.* 41: 233–243 (2003).
- Nespor, M.; Vogel, I.: On clashes and lapses. *Phonology* 6: 69–116 (1989).
- Ortega-Llebaria, M.; Prieto, P.: Disentangling stress from accent in Spanish: production patterns of the stress contrast in deaccented syllables; in Prieto, Mascaró, Solé, Segmental and prosodic issues in Romance phonology, pp. 155–176 (Benjamins, Amsterdam 2007).
- Otake, T.; Hatano, G.; Cutler, A.; Mehler, J.: Mora or syllable? Speech segmentation in Japanese. *J. Mem. Lang.* 32: 358–378 (1993).
- Ramus, F.: Acoustic correlates of linguistic rhythm: perspectives. *Proc. Speech Prosody, Aix-en-Provence 2002*, pp. 115–120. Available at <http://www.isca-speech.org/archive>
- Ramus, F.; Nespor, M.; Mehler, J.: Correlates of linguistic rhythm in the speech signal. *Cognition* 73: 265–292 (1999).
- Roach, P.: On the distinction between ‘stress-timed’ and ‘syllable-timed’ languages; in Crystal, Linguistic controversies: essays in linguistic theory and practice in honour of F.R. Palmer, pp. 73–79 (Arnold, London 1982).
- Rouas, J.; Farinas, J.; Pellegrino, F.; André-Obrecht, R.: Rhythmic unit extraction and modelling for automatic language identification. *Speech Commun.* 47: 436–456 (2005).
- Scott, D.; Isard, S.D.; de Boysson-Bardies, B.: Perceptual isochrony in English and French. *J. Phonet.* 13: 155–162 (1985).
- Sluijter, A.M.C.; van Heuven, V.J.: Spectral balance as an acoustic correlate of linguistic stress. *J. acoust. Soc. Am.* 100: 2471–2485 (1996).
- Stockmal, V.; Markus, D.; Bond, D.: Measures of native and non-native rhythm in a quantity language. *Lang. Speech* 48: 55–63 (2005).
- Tserdanelis, G.: The role of segmental sandhi in the parsing of speech: evidence from Greek; doct. diss. Ohio State University (2005).
- Turk, A.E.; Shattuck-Hufnagel, S.: Word-boundary-related duration patterns in English. *J. Phonet.* 28: 397–440 (2000).
- Wagner, P.S.; Dellwo, V.: Introducing YARD (Yet Another Rhythm Determination) and re-introducing isochrony to rhythm research. *Proc. Speech Prosody, Nara 2004*. Available at <http://www.isca-speech.org/archive>.
- Warner, N.; Arai, T.: Japanese mora-timing: a review. *Phonetica* 58: 1–25 (2001).
- White, L.; Mattys, S.L.: Calibrating rhythm: first language and second language studies. *J. Phonet.* 35: 501–522 (2007).
- Woodrow, H.: Time perception; in Stevens, Handbook of experimental psychology, pp. 1224–1236 (Wiley, New York 1951).