DE GRUYTER MOUTON
DOI 10.1515/lp-2013-0012 — Laboratory Phonology 2013; 4(2): 339–382

Anastassia Loukina, Burton Rosner, Greg Kochanski,
Elinor Keane and Chilin Shih

What determines duration-based rhythm measures: text or speaker?

Abstract: Differences in rhythm between languages have been often attributed to differences in phonological properties such as syllable structure. This paper uses quantitative analyses to determine whether and how popular duration-based rhythm measures depend on the phonological structure of a language. Native speakers of five languages read a large corpus of comparable texts (approximately 371,000 syllables in total). Phonological properties of each language were specified as 11 variables, computed from the phonetic transcriptions. These variables were compared against published rhythm measures that captured variation in duration of consonantal and vocalic intervals. While the text-based measures discriminated well between languages, the values of rhythm measures overlapped substantially, showing that the languages are more alike in acoustic implementation than in their phonological description. Multilevel models demonstrated that the mapping between phonological properties and acoustics is much weaker than previously assumed: linear effects of the phonological variables explained less than a quarter of the total variance in rhythm measures. Instead, speaker was the main source of variation in those measures. Rhythm, in the sense of durational variability, depends far more on individual timing strategies than on the phonological structure of a language.

Anastassia Loukina: University of Oxford, UK. Currently at ETS, USA.
E-mail: anastassia.loukina@stx.oxon.org
Burton Rosner: University of Oxford, UK. E-mail: burton.rosner@phon.ox.ac.uk
Greg Kochanski: University of Oxford, UK. Currently at Google, Inc. E-mail: gpk@kochanski.org
Elinor Keane: University of Oxford, UK. E-mail: elinor@keane.org.uk
Chilin Shih: University of Illinois, USA. E-mail: cls@illinois.edu

1 Introduction

Attempts to classify languages based on their rhythm and experimental studies of linguistic rhythm go back to the early 20th century. These efforts began as a search for isochrony in particular units of speech (see, for example, Ramus 2002;

Kohler 2009; Arvaniti 2012 for the historical background). The search proved unsuccessful, leading Dauer (1983, 1987) and Roach (1982) to suggest in the 1980s that a language's rhythm arises from language-specific phonological structures. Dauer (1983) identified three particular aspects of phonological structure as affecting the overall rhythmic impression of a language: syllable structure, the nature of any vowel reduction, and the role of stress.

This new viewpoint heavily influenced attempts to find acoustic correlates of impressions of linguistic rhythm. Ramus, Nespor, and Mehler (1999: 270) introduced their influential rhythm measures (RMs) as "an implementation of the phonological account of speech rhythm". Low, Grabe, and Nolan (2000) devised another set of RMs that reflected durational alternations supposedly due to stress-related vowel reduction. All these measures became very popular and fostered a number of variants (for further discussion of rhythm measures in the context of rhythm typology see, for example, Arvaniti 2012).

The resulting widespread use of the term 'rhythm' to refer to durational variability is somewhat misleading. Experimental studies suggest that the perception of rhythm rests on a combination of different acoustic properties that go beyond duration. Barry, Andreeva, and Koreman (2009), for example, showed that changes in F_0 influence the perceived strength of rhythmicity. In another study, rate of spectral change proved the most robust property for distinguishing spoken poetry from prose (Kochanski et al. 2010). Other studies examined spectral properties (Low, Grabe, and Nolan 2000; Tilsen and Johnson 2008; Tilsen 2008), intensity (Lee and Todd 2004; Keane 2006), and modelled auditory prominence (Lee and Todd 2004). The foundations of linguistic rhythm remain an open question (see, for example, Kohler 2009 for further discussion).

Nevertheless, duration-based measures have been widely used for comparisons of different languages and varieties (see, for example, White and Mattys 2007; Arvaniti 2012 for an overview). Many of these past studies tended to use rhythm measures as a convenient way to quantify perceived similarities or differences between varieties or languages. The term 'rhythm' itself became a way to refer to these differences (cf., for example, Torgersen and Szakay [2012], who recently used RMs to study contact-induced changes in London English). The interpretation of the observed differences in RMs was not always clear and often invoked, among other factors, differences in syllable structure and other phonological properties of the studied varieties. For example, Torgersen and Szakay (2012) argued that different patterns of variation in vowel duration in London speech are partially caused by the ongoing monophthongisation of several diphthongs in parts of London.

Despite their initial success, the substantial variability of rhythm measures has aroused doubts about their effectiveness in separating languages or groups

of languages (Lee and Todd 2004; Keane 2006; Arvaniti 2009; Wiget et al. 2010; Loukina et al. 2011; Arvaniti 2012). For example, Loukina et al. (2011) showed that differences in RMs between languages could only be accurately detected on a sufficiently large corpus. Several studies attempted to test how RMs are affected by the choice of material that supposedly constituted the primary source of variability. Two studies tried to control syllable structure by constructing three sets of sentences: one set with sentences representative of a given language, and the other two sets designed to enhance or simplify syllable complexity. Prieto et al. (2010) found that while the three sets gave significantly different values for some RMs in English, Spanish, and Catalan, others were unaffected. Arvaniti (2012) used the same approach to study the effect of text properties in a multilingual corpus that included sentences in English, German, Greek, Italian, Korean, and Spanish. In agreement with Prieto et al. (2010), she reported inconsistent effects of sentence type across languages and metrics.

Wiget et al. (2010) adopted a different approach in investigating how the choice of sentence materials influences rhythm measures. They calculated a contrast regularity index, intended to represent the regularity of sentence stress patterns. They compared this phonological measure for the five English sentences used in their study against various RMs. Fairly close correlations emerged with the local measures developed by Low et al. (2000), but not with the global measures of Ramus et al. (1999).

These various results on the effects of sentence materials raise some important questions. On the one hand, the finding that some RMs may depend on syllable composition of the material or other text properties is in itself not surprising. After all, the RMs were designed to capture such differences and are often interpreted as a reflection of phonological differences. But the claim that RMs do not separate languages well enough because of their dependence on material (cf. Arvaniti 2012) has another consequence. It implies that while some RMs may successfully capture phonological properties, languages are not as different in phonology as has been previously assumed. Accordingly, overlap in phonology between different languages would result in overlap in acoustics.

On the other hand, the lack of any text dependence for other RMs implies that duration-based measures do not necessarily reflect differences in phonology. This in turn challenges the general assumption of a uniform mapping between phonology and phonetic implementation. This assumption underlies the expectation that phonological differences between texts in different languages or within the same language should be directly reflected in durational variation as captured by the RMs (see also Arvaniti [2009] for similar remarks). If this assumption is incorrect, observed differences in RMs cannot be explained by differences in syllable structure or changes in vowel quality.

Finally, the inconsistent results reported in these studies may be due to methodological shortcomings. Any text can be characterized by a variety of phonological properties that potentially could affect variation in durational patterns in the spoken version of the text. Previous studies either controlled for only one property (Low et al. [2000] looked at vowel quality; Wiget et al. [2010] looked at stress) or studied differences between sentences that combined several properties defined in impressionistic fashion. This leaves open the possibility that further differences in phonology between texts had confounded the analyses and obscured the results.

1.1 Purpose and design of experiment

The main aim of this paper is to establish just how tight a relationship actually holds between rhythm measures and phonological properties. Do languages differ in the phonological properties that supposedly underlie the values of RMs? Do RMs reflect these differences within and between languages? Furthermore, which phonological properties most affect the values of RMs and, conversely, which RMs show the highest dependence on phonology? Finally, are the relations between RMs and phonology uniform across different languages?

To attack these questions, we used a substantial corpus of existing texts in five languages, in preference to artificially constructed material. For each text, we computed numerous 'text phonological measures' (TPMs) that were intended to quantify aspects of its phonology. We first tested how clearly these TPMs separated the languages. Then we assessed how well the TPMs predicted the values of RMs derived from readings of the texts. We used all available RMs and numerous TPMs in order to ensure the best possible chance of finding the connection between rhythm and phonological structure in case such connection exists. Although we initially studied how variation in RMs depended on TPMs across the five languages, other sources of variability in RMs quickly became apparent. Foremost, they included speaker identity. Therefore, we assessed the relative importance for the RMs of speaker-dependence as against TPM-dependence within each language. We also examined possible language-specific differences in the relationships between RMs and TPMs.

In theory, duration-based rhythm measures rest purely on acoustic durational variability. In practice, though, their calculation requires prior linguistic segmentation and classification of the resulting segments as vowels or consonants. As Hirst (2009: 1520) notes, "neither of these operations is purely acoustic". Manual labelling depends on phonological interpretation and therefore introduces an intrinsic confound into cross-linguistic studies (see Loukina et al.

[2011] for further discussion). To avoid this confound, we used automatic segmentation that is blind to phonological interpretations and that applies consistent criteria across languages.

The paper has the following structure. First, Section 2 presents our methodology. Second, Section 3 contains the results of statistical analysis of our measurements. Third and finally, the implications of these results for rhythm studies and for linguistics in general are considered in Section 4.

2 Method

2.1 Selection of texts

2.1.1 Languages

The five languages represented in our corpus are English, Greek, Russian, French, and Mandarin. They are generally regarded as spanning a wide range of phonological characteristics that could affect rhythm. English and Russian have complex syllable structure and vowel reduction. Unlike English, Russian does not have distinctly short and long vowels. Compared to English and Russian, Greek and French have greater constraints on syllable structure and relatively little vowel reduction (although see Loukina 2009). The latter two languages, however, differ in their accentual systems. Mandarin shows the most constraints on consonant clusters. It is also a tone language, belonging to a completely different language family from the other four.

2.1.2 Text phonological measures

We developed 11 text phonological measures that the literature indicated should correlate with durational characteristics of speech and hence with the values of rhythm measures. Previous explanations of rhythmic differences between languages have invoked differences in syllable structure and in prosodic features such as stress, and we designed several TPMs to reflect such differences. However, the duration of consonantal and vocalic intervals may also be affected by other factors, such as differences in intrinsic durations between segments or proximity to pause. Although such factors have not been traditionally associated with rhythm, their effect on segmental durations could confound or mask the effect of more traditional 'rhythmic' factors. Conversely, it is not possible to assess

the effect of rhythm-related properties such as syllable structure without controlling for possible differences in other factors such as intrinsic durations. Our TPMs were designed to systematically reflect a wide range of factors that should affect the durations of vocalic and intervocalic intervals. This approach allows us to separate the effect of different properties of text on the variation in RMs.

A key consideration in designing the TPMs was consistent cross-linguistic application: every TPM had to be unambiguously defined in each language and defined identically in all five. Where possible, some TPMs were defined analogously to published RMs, to encourage comparability. Hence several TPMs have 'PVI variants' (pairwise variability index) that try to capture the importance of local differences. Each TPM also had to be automatically computable from broad phonetic transcriptions. Table 1 lists the 11 TPMs.

The Ccluster TPM and its PVI-variant were intended to capture differences between languages that show variation in syllable structure, such as Russian and English, as against languages with greater constraints on syllable structure, such as French or Mandarin. Therefore this TPM tests whether variation in the

Table 1: Text phonological measures (TPMs).

TPM	Description
Consonantal TPMs (intervocalic intervals)	
Ccluster	Mean number of consonants between successive vowels.
Ccluster-PVI	Mean-square difference between numbers of consonants in adjacent inter-vowel gaps.
Vocalic TPMs (vocalic intervals)	
CVVC	Across all syllable pairs, proportion containing a [C]V then a V[C], ignoring word boundaries and counting diphthongs as one vowel.
Highlow	Mean height index, after assigning 3 to each high vowel, 1 to each low vowel, 2 to other vowels, and giving a diphthong value for each of its two elements.
Highlow-PVI	Mean square difference in height index between adjacent vowels.
Diphthongs	Proportion of syllables containing diphthongs or a combination of a vowel and a glide.
Syllabic TPMs (both types of intervals)	
Sonority	Mean sonority, after giving values of 1 to stops, fricatives, and affricates, 2 to nasals and liquids, and 3 to glides and vowels.
Sonority-PVI	Mean square difference in sonority value between adjacent segments.
Voiced	Proportion of voiced segments across all segments.
Strong	Proportion of 'strong' syllables across all syllables. (See text for full explanation.)
Pauses	Mean expected pause incidence, after assigning 0 to each vowel and 1 to each punctuation mark.

number of intervocalic consonants is correlated with the variation in duration of intervocalic intervals. Similarly, CVVC and Diphthongs test whether the duration of vocalic intervals is directly related to the number of phonetic vowels in that interval.

The Strong TPM was designed to capture stress-related variation. It could not be defined identically across languages. We regarded it as important, however, and therefore fell back on language-specific definitions of strong syllables. For English, Greek, and Russian, strong syllables were those marked in the automatic transcription as bearing primary lexical stress. In French, we assigned one stressed syllable per phrase plus one per three words.[1] In Mandarin, stress was assigned to any syllable not bearing a neutral tone, since syllables with neutral tone have acoustic properties (short duration and unstable F_0 patterns) very similar to those of unstressed syllables in other languages (Chao 1968; Lin 1983).

Traditional explanations of rhythmic differences have focused on syllable structure and prosody. However, much variation in segmental timing is determined by differences in intrinsic duration of the segments (cf., for example, Klatt 1976; Van Santen 1992). To test whether such factors may partially explain the effect of text observed in previous studies, we included several TPMs designed to capture differences in intrinsic duration. These differences in intrinsic duration are rarely considered in rhythm studies (although see, for example, Dellwo, Fourcin, and Abberton 2007). However, they can substantially affect the duration of vocalic and intervocalic intervals and therefore can influence the values of RMs. For example, it is well established that in equal conditions low vowels have a longer duration than high vowels (cf. Peterson and Lehiste [1960] and Scherba [1912] for Russian; Fourakis, Botinis, and Katsaiti [1999] for Modern Greek). The Highlow TPM and its PVI variant were meant to capture such differences. The duration of consonants also depends on their articulatory characteristics and especially voicing (cf., for example, Port 1977; Maddieson 1997). The TPMs Voiced, Sonority, and Sonority-PVI reflect such differences in duration between voiced and voiceless segments as well as the difference between sonorants and obstruents.

Our selection of TPMs does not capture every factor that might influence duration. The emphasis on consistent unambiguous cross-linguistic application also means that most of our TPMs are phonetic in nature and are blind to phonological representation (for example, the Ccluster TPM does not take into account

1 We broadly follow Di Cristo (1998) in assigning a stressed (prosodically prominent) syllable per stress group; a stress group contains a content word and any related pro/en-clitics. Our formula gives a good mechanical approximation to the mean number of stressed syllables in French.

syllabification). We addressed this problem in the design of our statistical analysis. Our statistical model not only shows the effect of the selected TPMs on the variation in RMs, but it also tests how well these TPMs explain the effect of text. If the previously reported effect of text on RMs is due to properties that this set of TPMs does not capture, our statistical analysis would uncover that.

2.1.3 Selection of texts

We needed a set of texts in each language that showed a wide range of values for the TPMs. We also had to ensure that the selected paragraphs showed comparable distributions of TPM values in each language. To meet these requirements, we first compiled for each language a corpus of paragraphs from the original and translations of J.K. Rowling's novel *Harry Potter and the Chamber of Secrets*. We used paragraph-sized pieces of text; each text contained on average 184 syllables. We chose paragraphs that did not contain dialogue. The corpus for each language contained on average 129 paragraphs, except for Mandarin, where the paragraphs were much shorter than in the other languages. To compensate, we chose 520 paragraphs in Mandarin. The texts were machine-transcribed at the broad phonetic level.[2] The transcriptions reflected assimilations, vowel reduction, devoicing, and some major allophonic differences, such as clear vs. dark /l/ in English. All transcriptions were manually checked for accuracy and corrected if necessary. Appendix A provides examples of transcribed texts.

We computed the 11 TPMs for each transcribed paragraph. All values were multiplied by 100 for ease of presentation. The paragraphs in each language were ranked on the value of each TPM. Then for each language we selected 1 paragraph with a comparatively low value, 1 with a comparatively high value, and 20 others with values in the middle range of a given measure. We used the following criteria:

- The value of the given TPM for a selected paragraph must fall within the top 5% for 'maximal' values or within the bottom 5% for 'minimal' values.
- The value of that TPM for the other paragraphs must be as near average as possible.
- A paragraph must be chosen just once.

2 Unisyn (Fitt 2000) for English, where the dictionary was updated to include personal names, etc.; Mbrola (Bechet 2001) for French; and Shih and Sproat (1996) for Mandarin. Speech Technology Centre Ltd., St. Petersburg, Russia, and the Institute for Speech and Language Processing, Athens, Greece, produced the Russian and the Greek transcriptions, respectively.

Under these criteria, 22 paragraphs were chosen in each language that showed comparable variation in the values of a given TPM. For example, in each language there was a paragraph representing a near-maximal and a paragraph representing a near-minimal value of the TPM Voiced. For the remaining 20 paragraphs, the value of Voiced varied somewhat around average.

For each language, we also selected four 'normal' paragraphs where the value for each of the 11 TPMs was nearest to the median of its distribution. Finally, we chose two 'convergence' paragraphs for each language. The convergence paragraphs were intended to have average TPMs across all languages in contrast to the normal paragraphs that represented each language's typical TPM values. The first set of convergence paragraphs had values that were as similar as possible in all five languages for the four TPMs related most closely to syllable structure: Ccluster, Ccluster-PVI, CVVC, and Diphthongs. The second set had very similar values for TPMs that reflected segmental characteristics: Highlow, Highlow-PVI, Sonority, Sonority-PVI, and Voiced.

In total, then, 28 paragraphs were chosen for each of the five languages. To compensate for the short paragraph lengths in the translations, an additional 28 Mandarin paragraphs were selected under the same procedures. Figure 1 illustrates the selection of paragraphs. The distributions of the TPM Voiced across all selected paragraphs are shown for three of the five languages. The locations of particular types of paragraphs, such as 'near-maximum' and 'normal', are also indicated.

2.2 Recordings

Native speakers of each language read the paragraphs selected for that language. Speakers were 20–32 years old. All had been born to monolingual parents and had grown up in their respective countries. At the time of the recordings, they were residents in Oxford. Non-British participants had lived outside their home country for less than 4 years; median length of residence in the UK was 1 year.

Sixty-three participants were recorded: 24 British English speakers (Southern England), 10 Russian speakers (Moscow or St. Petersburg), 10 Taiwanese Mandarin speakers (Taipei), 9 Greek speakers (Athens), and 10 French speakers (Paris). This resulted in 2,016 recorded paragraphs (approximately 371,000 syllables) for our use. The recordings constitute part of the "Oxford Aesop Corpus", which is available online at www.phon.ox.ac.uk/corpus. Loukina et al. (2011) used a subset of the corpus and give further details on the recording procedures. All experimental procedures complied with relevant laws and institutional

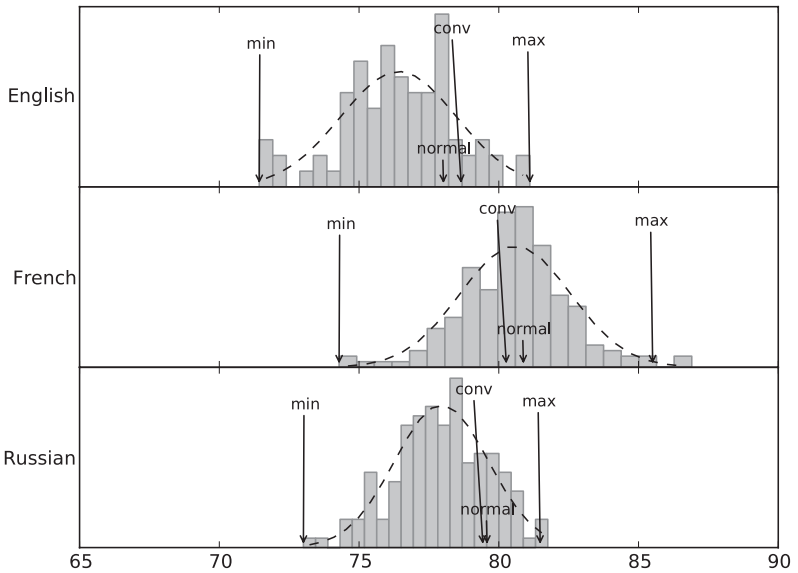

Fig. 1: Distribution of values for TPM Voiced in three languages over all selected paragraphs. Arrows indicate values for near-maximum ('max') and near-minimum ('min') paragraphs, for a 'normal' paragraph with values for every TPM near the centre of its distribution, and for a 'convergence' paragraph ('conv') where all languages show similar values for a given subset of TPMs. See text for further details.

guidelines and were approved by the Central University Research Ethics Committee, University of Oxford.

2.3 Segmentation and rhythm measures

The recordings were automatically segmented into vocalic and consonantal intervals. The segmentation algorithm was built from the standard HTK toolkit (Young et al. 2006). It splits speech into vowel-like, consonant-like, and silent intervals. The algorithm is similar to one described by Loukina et al. (2011), which showed substantial agreement with human labellers.

Fifteen different RMs were calculated for each recorded paragraph. (As with the TPMs, all RM values were multiplied by 100.) Table 2 lists the RMs. Each rests entirely on the durations of the vocalic and consonantal intervals resulting from the automatic segmentation. Previous investigators have computed RMs over different stretches of speech. Loukina et al. (2011) found that these procedural differences had no significant effect on the values of the RMs. The present computation

Table 2: Rhythm measures.

RM	Description	Reference
%V	Percentage of vocalic intervals	Ramus, Nespor, and Mehler (1999)
Vdur/Cdur	Ratio of total vowel durations to total consonant durations	Barry and Russo (2003)
ΔV	Standard deviation of vocalic intervals	Ramus, Nespor, and Mehler (1999)
VarcoΔV	ΔV/mean vocalic duration	Dellwo (2006)
VnPVI	Normalised pairwise variability index (PVI) of vocalic intervals	Grabe and Low (2002)
medVnPVI	VnPVI computed using median value	Ferragne and Pellegrino (2004)
ΔC	Standard deviation of consonantal intervals	Ramus, Nespor, and Mehler (1999)
VarcoΔC	ΔC/mean vocalic duration	Dellwo (2006)
CrPVI	Raw PVI of consonantal intervals	Grabe and Low (2002)
CnPVI	Normalised PVI of consonantal intervals	Grabe and Low (2002)
medCrPVI	CrPVI computed using median value	Ferragne and Pellegrino (2004)
PVI-CV	PVI of consonant+vowels groups	Barry et al. (2003)
nCVPVI	Normalised PVI of consonant+vowels groups	Asu and Nolan (2005)
VI	Variability index of syllable durations	Deterding (2001)
YARD	Variability of syllable durations	Wagner and Dellwo (2004)

of RMs excluded the final syllable in each interpause stretch (see Loukina et al. 2011 for details).

3 Results

3.1 Do TPMs differentiate between languages?

According to traditional phonological descriptions, the five languages in our corpus differ in various properties such as the complexity of consonant clusters. We first had to demonstrate that our TPMs capture these differences. To do this, we compared inter-language with intra-language variability in TPMs by applying machine classifier techniques (cf. Kochanski and Orphanidou 2008). Loukina et al. (2011) used this computational procedure to determine how well RMs could differentiate between languages. We applied exactly the same machine algorithm here, evaluating how well a linear classifier can use TPMs to identify the language of a paragraph. Our particular classifiers decided which language was most likely to have produced a particular text, given one or more TPM values computed from that text (see Loukina et al. 2011 for further details).

A classifier's performance is measured by the probability of correctly identifying the language of a given paragraph. If this is large (i.e., near 1.0), then the data from the various languages can be separated into distinct groups bounded by straight lines. This constitutes a test of the hypothesis that TPMs for different languages form separate groups. An identification probability near chance means that the TPM values from different languages largely overlap, so that the TPMs are insensitive to phonological differences between the languages. We excluded the property Strong from classifier analysis, since it was defined in a language-specific way.

To allow simple comparisons, we report both the proportion of correct identifications *P(C)* and a figure of merit designated *K*. This was computed as $K = \frac{P(C) - chance}{1.00 - chance}$, where 1.00 represents perfect performance. Therefore, *K* varies between 0 for classifiers that perform at chance and 1 for perfect classifiers. We used *z*-tests to assess the significance of differences both between *P(C)* and chance for each classifier and the significance of differences in *P(C)* between classifiers.[3] Throughout this paper, the significance threshold α is set conservatively at .01.

As in Loukina et al. (2011), we first ran 10 classifiers based on single TPMs and then 45 classifiers based on all possible pairs of those TPMs. All classifiers based on single TPMs identified the five languages in our corpus at a level above chance. The average *P(C)* was .52 (chance = .2), and the average *K* was .4. Table 3 presents success data for each single-TPM classifier. All of them produced a statistically significant value for *K*.

Compared to the previously studied classifiers based on single RMs, classifiers based on single TPMs were substantially more successful: Loukina et al. (2011) had found that only 8 out of the 15 individual RMs in Table 3 separated languages at a level above chance. The average *P(C)* and average *K* were just .33 (chance = .23) and .12, respectively.

Loukina et al. (2011) also found that maximum identification rates were achieved by classifiers using various combinations of three rhythm measures. The average success rate reached .44 with *K* at .27[4]. Similarly, combinations of two TPMs here yielded classification rates above those produced by single TPMs. Of the 45 possible classifiers based on pairs of TPMs, 18 achieved remarkably high

3 Different splits into training sets and test sets are not independent and therefore do not satisfy the independence assumption of standard *z*-tests. To correct for this, the boundaries for confidence intervals were calibrated using a Monte Carlo simulation. We generated 200 different random data samples from a known distribution and performed a classifier analysis on these samples. We then used the results of this analysis to compute the adjustment parameter necessary to obtain accurate significance levels ($f = 0.71$).

4 This is comparable to human performance; see Loukina et al. (2011) for references.

Table 3: *P(C)* and *K* for classifiers based on individual TPMs. Chance in all cases was .2. All single *K* values are significantly greater than 0 (one-tailed comparison), that is, classifiers based on each of these TPMs performed above chance level. Between-classifier differences in *K* that exceed .15 are significant (two-tailed comparison).

TPM	*P(C)*	*K*
Sonority-PVI	.31	.13
Pauses	.40	.25
CVVC	.44	.29
Ccluster-PVI	.44	.30
Voiced	.49	.37
Ccluster	.52	.40
Highlow-PVI	.58	.48
Highlow	.61	.52
Sonority	.68	.60
Diphthongs	.76	.70

success rates. Average *P*(C) over these 18 classifiers was .83, with average *K* at .8. No significant differences in success rates appeared between any of these 18 classifiers based on pairs of TPMs. These rates for classifiers based on two TPMs substantially exceed the maximum success rates (given above) for classifiers based on combinations of three RMs.

Further increase in the number of TPMs used by a classifier was not feasible due to a restriction on data size. The TPMs are properties of texts, independent of speakers. Therefore, having ten speakers read a given text merely gives ten copies of the same TPMs rather than ten independent samples of each TPM. Hence the amount of available data was only one-tenth that of the speech data used in Loukina et al. (2011).

Even with the restriction of data size, the classifier analyses based on pairs of TPMs makes two points clear. First, the TPMs succeeded well in picking up phonological differences yielded by traditional descriptions of the five languages. Second, the overlap in TPMs between languages was much smaller than that for RMs.

3.2 Differences in TPMs between languages

Having established that languages in our corpus differ in the values of TPMs, we examined the nature of those differences. Mean values and standard deviations for all TPMs appear in Appendix B. Some between-language differences in

those values agreed with our expectations. Accordingly, English showed a higher average number of consonant clusters (Ccluster) than other languages, while this number was lowest for French and Greek (post-hoc Tukey test,[5] adjusted p ranged from 1.94×10^{-14} to 1.9×10^{-6}). In Mandarin the average number of consonants was higher than in French or Greek, due to the frequent occurrence of nasal codas that created multiple clusters of two consonants.

The properties that separated languages most successfully, however, were related to segment quality such as Highlow, Sonority, and Voicing. We found that Russian and English had the highest occurrence of high vowels (such as [i] and [u]), while in French and Greek the number was the lowest, so that those languages had a higher proportion of low vowels (adjusted p ranged from 1.93×10^{-14} to 8.46×10^{-6}). At the same time, in English and Greek, high vowels often alternated with low vowels giving rise to higher Highlow-PVI, while in Russian and French vowels of the same height frequently occurred together, leading to lower values of Highlow-PVI. Amongst all five languages, Mandarin showed the biggest difference in height between adjacent vowels (adjusted p ranged from 1.93×10^{-14} to 1.67×10^{-10}).

Greek, Mandarin, and French showed higher proportions of sonorant segments such as vowels and glides than did Russian, with Sonority lowest for English (adjusted p ranged from 1.93×10^{-14} to 3.68×10^{-9}). This result mirrors previously reported differences in the number of consonant clusters. At the same time, the proportion of voiced segments was highest in French, followed by Russian and English, and lowest in Greek and Mandarin (adjusted p ranged from 1.93×10^{-14} to 3.54×10^{-7}). This discrepancy between sonority and voicing is due to different distribution of voiced stops in languages in our corpus.

3.3 Multidimensional scaling

Given these favourable results with TPMs, we now can confront the main issue of this paper: the relationships between RMs and TPMs. Many rhythm measures listed in Table 2 above involve similar computations and therefore should be correlated. Bivariate correlations between the 105 pairs of RMs confirmed this. For each language, we made a histogram of the values of r^2, treating positive and negative correlations in the same way. The results were much the same for all languages. The histograms were positively skewed, with medians varying from

5 The values of TPMs do not vary between speakers; therefore, one value was used for each text. Shapiro-Wilks tests showed that the data were normally distributed in all cases.

.06 to .10 and means varying from .17 to .19. Most importantly, the ranges of r^2 were very large (.89 to .95), showing considerable variation in the correlations; r^2 even approached unity for some pairs. A Friedman two-way ANOVA on the r^2 values revealed no differences between languages.

Since some of the 15 RMs were highly correlated, systematic analyses using all RMs would be wasteful. To avoid this, multi-dimensional scaling (MDS) was performed with PROXSCAL in order to identify a compact subset of RMs for use in further analyses. We used $1 - r$ as an ordinal measure of dissimilarity between each pair of RMs.[6] Languages were treated as separate sources in PROXSCAL.

A three-dimensional solution emerged, with an acceptable normalized raw stress of .03.[7] Figure 2 shows how the RMs grouped in this 3-D space. The first dimension seems related to Normalization, with the unnormalized ΔC at one end, the ratio measure %V at the other, and normalized measures such as YARD in between. The second dimension reflects Type of Interval (consonant vs. vowel) and differentiates between measures using consonantal durations and those using durations of vocalic intervals. Finally, the third dimension seems sensitive to Computational Span, such as the use of standard deviation over the whole paragraph as against local PVI-like measures.

In agreement with the negative result of the Friedman ANOVA reported above, MDS weights differed only modestly between languages (see Table 4). For all languages, Computational Span had the least weight. For Russian, English, and Greek, Normalization had greater weight than Type of Interval. For French and Mandarin, the reverse was true.

From the MDS results we selected four representative measures: CrPVI, %V, VarcoΔV, and YARD. These measures span the three-dimensional common space

Table 4: MDS dimensional weights for each language.

Language	Normalization	Type of interval	Computation
Russian	.43	.41	.33
English	.44	.41	.31
Greek	.46	.39	.32
Mandarin	.40	.45	.32
French	.39	.43	.35

6 Note that anti-correlated RMs are treated as more different than uncorrelated RMs.

7 The scree plot for French showed a small drift towards 5D, but the additional decrease in stress values was small.

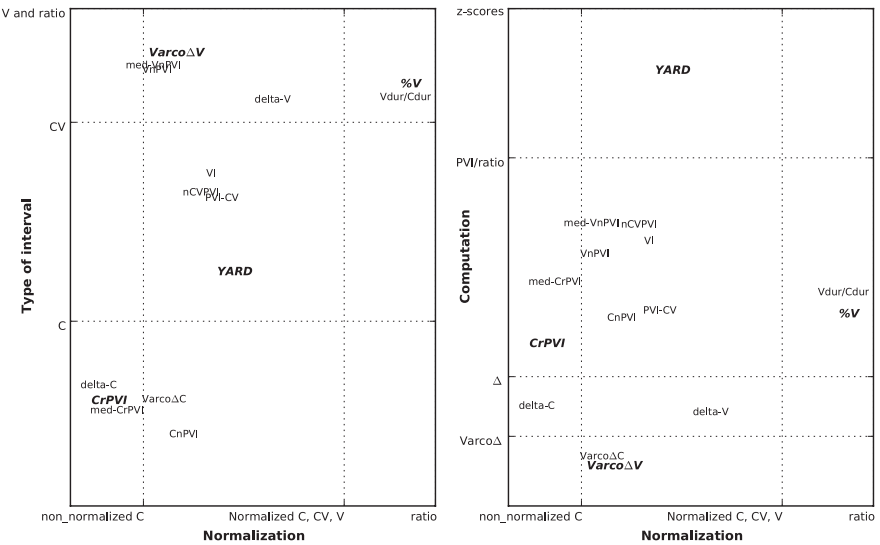

Fig. 2: Location of RMs in the common MDS space, for dimensions 1 and 2 (left) and dimensions 1 and 3 (right). Normalized raw stress = .03. Dimensions are interpreted as Normalization, Computation, and Type of Interval. Measures selected for further analysis appear in boldface.

occupied by all 15 RMs and appear in boldface in Figure 2.[8] All further analyses with RMs used only the selected four, leaving us with 20 possible combinations of RMs and languages. Other choices of RMs, of course, are possible. But whatever the choice, any results that would be obtained on the remaining RMs should be predictable from the findings on four measures chosen to cover the MDS space. Notice that to be worthwhile, any attempt to develop a new rhythm measure would have to demonstrate that the proposed RM captures new information, adding a fourth dimension to the MDS space. This requirement should prevent duplication of effort on similar RMs.

We could not conduct MDS on TPMs with languages treated as separate sources, since Strong was not defined in a language-independent way. We therefore ran a separate MDS analysis on the 11 TPMs for each individual language. Dimensionality differed somewhat between languages, and the optimally placed TPMs differed across languages. Given these differences, we chose the conserva-

8 In Loukina et al. (2011), we reported that all these measures apart from CrPVI could diferentiate between the five languages in our corpus with moderate success (see 3.1). None of the consonantal measures in that study could separate languages above chance.

tive path of retaining all 11 TPMs for further analyses, along with the four selected RMs.

3.4 Explaining the variability in RMs

To evaluate the contribution of TPMs and three other possible factors for variation in RMs, we performed a sequence of statistical analyses with mixed linear models. The analyses assessed how accurately the value of an RM for a given paragraph is predicted by the values of TPMs for that paragraph and by its speaker, text, and language. To obtain directly comparable coefficients and meaningful intercepts, all TPM values underwent *z*-transforms. The transforms were done twice, independently: first, within each language for language-specific analyses (Sections 3.4.2–3.4.3 below) and second, across all languages for cross-linguistic analyses (Section 3.4.5).

3.4.1 General description of the model

The individual recordings in our study are not independent. They group naturally by language, speaker, and text. Within a language, our data follow a balanced cross-classified design. Each individual recording is nested within a particular speaker of that language and a particular text in that language. Figure 3 represents this schematically (cf., for example, Rasbash et al. 2009: 272).

This nesting implies that individual RMs will cluster by speaker or text. In that case, single-level models such as classical linear regression would underestimate standard errors and increase the probability of Type I error (cf., for example, Pinheiro and Bates 2009: 6). Adding speaker and text as fixed factors would generate a large number of additional parameters. Furthermore, such a model would only apply to a specific set of texts and speakers. It could not be generalized to other speakers or texts.

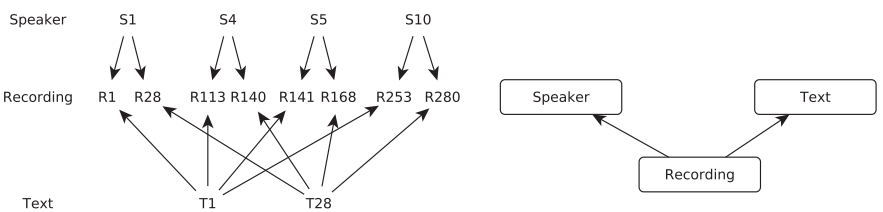

Fig. 3: Classification diagram (left) and unit diagram (right) for recordings within a language cross-classified by speaker and text.

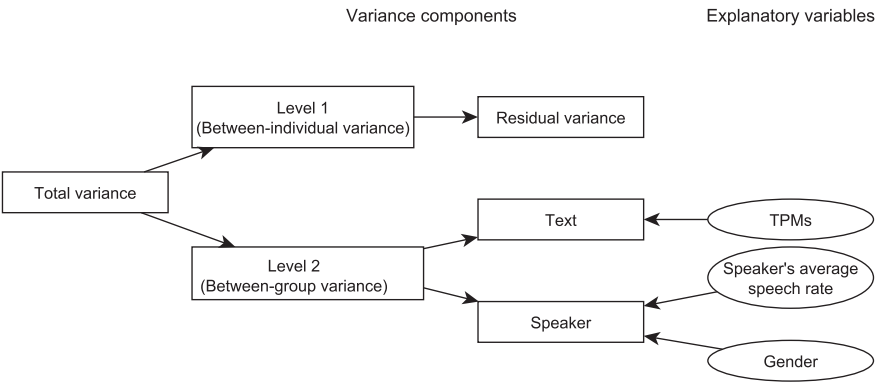

Fig. 4: The partitioning of variance in linear mixed models used here. Ellipses show explanatory variables for different components.

We therefore treated speakers and texts as random factors (see Baayen 2008: 169–275, for an extensive discussion of the advantages of such a treatment over the frequent approach of fitting separate models for individual subjects and texts). The total variance of an RM in our mixed models consists of the two components shown schematically in Figure 4. Level 1, or between-individual variance, arises from variation in an RM at the level of individual recordings. Level 2 is between-group variance and can be further partitioned into variance between speakers and variance between texts (for further details on variance partitioning see, for example, Snijders and Bosker 2012: 18).

The variance explained by TPMs appears as an ellipse connected to between-text variance. Note that the text variance places an upper limit on the variance attributable to the TPMs. Finally, two ellipses connected to speaker variance represent gender and speech rate, which comprise level-2 variables that might affect between-speaker variance. We will treat those two variables later.

Our analysis proceeded in three main steps. First, for each language, we established the partitioning of variance in each RM between the three sources, shown in Figure 4 by the rightmost rectangles. This step determined how far the values of RMs differed across speakers and across texts. Second, we estimated how much between-text variance could be attributed to linear effect of TPMs, treated as fixed factors. Third and finally, we ascertained whether and how these effects differed across the five languages in our corpus.

All mixed models were fitted with the lme4 package in R (Bates, Maechler, and Bolker 2011). Statistical parameters were estimated using maximum likelihood (ML). This procedure allowed us to compare models with different fixed

parts using likelihood ratio tests (see below and Snijders and Bosker [2012: 97] for further details).[9]

3.4.2 Partitioning of the variance in RMs

We first tested for clustering of each RM by speaker and text in the data for each language. With four RMs and five languages, there were 20 such tests. Such tests required the initial fit of a baseline single-level model (1), lacking any clustering, where RM_i is the value of a particular RM for recording i, β_0 is the grand mean across all speakers and texts, and e_i is the residual for recording i.

$$RM_i = \beta_0 + e_i \quad (1)$$

Next, a first-order multilevel model (2) was fitted to the data. The model includes a random effect of speaker and text on an RM and represents the cross-classified structure shown in Figure 3.

$$RM_{i(jk)} = \beta_0 + u_j + u_k + e_{i(jk)} \quad (2)$$

Here, $RM_{i(jk)}$ is the value of an RM for recording i, now tagged as obtained from speaker j reading text k. As before, β_0 is the grand mean. The new terms u_j and u_k represent the random variation in intercept due to speaker j and to text k, respectively. The modified component $e_{i(jk)}$ is the individual-level random departure in intercept. Notice in particular that u_j, u_k, and $e_{i(jk)}$ are all assumed to be normally distributed random variables, each with a mean of 0. Since both u_j and u_k are latent variables, the model is not defined by their individual values but instead by their variance. The latter is estimated by an iterative algorithm that optimizes the likelihood of the model parameters given the observed data (see Bates 2011 for a detailed description of the algorithm used in lme4).

9 ML and the other common estimation method, residual maximum likelihood (REML), differ in their estimates of variance components. This difference is important when the number of cases *(N)* is close to the number of level-2 explanatory variables *(q)*. Snijders and Bosker (2012: 60) suggest as a rule of thumb that the two methods become essentially equivalent when $N - q - 1 \geq 50$. In all our models $N - q - 1 \geq 240$, making the difference in estimates between REML and ML immaterial. The ratio of any REML estimates for our data to the actual ML estimates should be $\frac{N-q-1}{N}$ that is, between 0.95 and 0.98 for separate languages in our corpus, and about 0.99 for the cross-linguistic analysis in Section 3.4.5.

If there were no significant effect of text and speaker, the models in (1) and (2) would fit the data equally well. If speaker or text had a significant effect, however, the model in (2) would give a better fit. For each combination of *RM* and language, two likelihood ratio tests (otherwise known as 'deviance tests') were conducted to evaluate the significance of any effect of text and of speaker.

We first broke out any effect of speaker by comparing the empty model in (1) to a limited version of (2) that included a single level-1 variable, u_j, the random effects of speaker.[10] The ratio of the likelihoods of the two models was computed (see, for example, Snijders and Bosker 2012: 96–97, 206–208). This test revealed that the effect of speaker was significant for all four RMs within each of the five languages (p ranged from 5.1×10^{-137} to .0047).

We next determined whether model (2), which adds text as a random factor, produced an even better fit to the data. All 20 likelihood ratio tests showed that addition of the text factor significantly improved the model's fit (p ranged from 4.05×10^{-97} to .00078).

Given these positive results, we calculated the partitioning of total variance between speaker and text. We used the variance partition coefficient *(VPC)*, which is the ratio of variance attributed to a particular random factor to the total variance. Table 5 presents these coefficients for speaker (VPC_{sp}) and for text (VPC_{txt}) for each combination of RM and language.[11]

The mean and standard deviation of VPC_{sp} were .35 and .21, respectively. Friedman tests showed that VPC_{sp} varied significantly across RMs ($p = .005$) but not across language ($p = .02$). The statistic was largest for %V and CrPVI and

Table 5: The partitioning of level-2 (between-group) variance between speaker and between text by RM and language, presented as proportion of total variance.

Language	%V		CrPVI		VarcoΔV		YARD	
	VPC_{sp}	VPC_{txt}	VPC_{sp}	VPC_{txt}	VPC_{sp}	VPC_{txt}	VPC_{sp}	VPC_{txt}
Russian	.72	.16	.72	.04	.22	.38	.34	.19
Greek	.57	.29	.47	.11	.23	.38	.15	.19
French	.43	.34	.55	.10	.22	.27	.20	.11
Mandarin	.29	.45	.38	.14	.07	.39	.03	.27
English	.66	.19	.43	.24	.17	.35	.10	.33

10 The model is reduced to $RM_{i(jk)} = \beta_0 + u_j + e_{i(jk)}$.

11 For simple multilevel models, *VPC* is equal to the intra-class correlation coefficient (cf. Snijders and Bosker 2012: 39).

reached .72 in Russian, showing that almost three-quarters of the total variance in these two measures could be attributed to between-speaker differences.

Mean VPC_{txt} was .25 with a standard deviation of .12. As noted above, this already places a limit on the maximum possible effect of the text and therefore of the TPMs. According to Friedman tests, VPC_{text} did not vary significantly across RMs ($p = .03$) or across language ($p = .12$).

In all languages, %V showed substantial variation due to speakers and to texts. In each case, however, some portion of the total variance was not attributable to speaker or text and therefore remained unexplained. The share of unexplained variance in %V averaged only 18%.[12] For YARD, in contrast, total level-2 variance was low, so that variance was concentrated at the level of individual recordings; on average the share of unexplained total variance was 60%. Finally, CrPVI and VarcoΔV fell between these extremes: average unexplained total variance came to 36% and 46%, respectively. A Friedman test revealed that the share of unexplained variance differed significantly across RMs ($p = .003$). This effect, however, was due principally to significant differences in speaker effects across RMs. There was no significant difference between languages ($p = .011$).

Note that our first-order model assumes that text and speaker effects were additive. It is entirely possible, however, that speakers do not treat each text uniformly; for instance, speakers could use a different approach to dramatic passages. This would give rise to a speaker by text interaction. We did not have the data to evaluate it, but such an interaction could underlie part of the unexplained variance.

In summary, variation between speakers and between texts accounted for some but not all variance in each of our four rhythm measures. This finding confirms that our data are clustered, making multi-level models necessary.

3.4.3 Between-texts variance: RMs and TPMs

We now could ascertain how much between-text variance was explained by linear effects of the 11 TPMs. This test used the second-order model shown in (3). It simply adds the TPMs as linear predictors to the model in (2).

$$RM_{i(jk)} = \beta_0 + u_j + u_k + \beta_1 * \text{Ccluster}_k + \cdots + \beta_{11} * \text{Pauses}_k + e_{i(jk)} \quad (3)$$

12 This percentage of unexplained total variance is easily computed from Table 5 as $100 * (1 - (VPC_{sp} + VPC_{txt}))$.

In the new equation, $Ccluster_k$, $Ccluster\text{-}PVI_k$, . . . , and $Pauses_k$ are the values of TPMs for text k, and $\beta_{1–11}$ are the coefficients for the TPMs. The TPMs are level-2 explanatory variables, as their values do not vary between recordings of the same text read by different speakers. All other terms remain as in (2).

The addition of TPMs significantly improved the fits for %V in all languages but English. The fits for VarcoΔV were better for all languages but French. For CrPVI, the fit significantly improved only for Greek, while the fits improved for YARD in Mandarin and English. In short, the TPMs improved the fit in 11 out of 20 cases, with p ranging from 1.4×10^{-11} to .0061.

Given this positive result, we computed the fraction of variance explained by TPMs (R^2_{tpm}) as the proportional reduction in variance due to their inclusion as explanatory variables (Snijders and Bosker 1994, 2012). Equation (4) shows the computation.

$$R^2_{tpm} = \frac{var(RM_{ijk}) - var(RM_{ijk} - \sum_h \beta_h * TPM_{hk})}{var(RM_{ijk})} \tag{4}$$

Here, R^2_{tpm} is the variance in an RM explained by the linear effect of TPMs. The first term $var(RM_{ijk})$ is the total variance in that RM; that term is the sum of level-1 (residual) and level-2 variances (speaker and text). The second term $var(RM_{ijk} - \sum_h \beta_h * TPM_{hk})$ is the smaller variance left in the RM values after removing the linear effects of TPMs (see (3)). The difference between the two terms is divided by the first term to give the fractional variance.

Table 6 shows three components of the original variance for each combination of RM and language. The components are: (a) the fraction of variance explained by the linear effects of the 11 TPMs (R^2_{tpm}); (b) the share of original variance in the value of each RM attributed to between-speaker effects (Var_{sp}); and (c) the net share due to between-text variance after removing the effects of the TPMs (Var_{tn}). Equation (5) gives an example of this computation for Var_{sp}.

$$Var_{sp} = \frac{r^2_{sp}}{\sum r^2_E + \sigma^2_E} \tag{5}$$

Here, r^2_{sp} is between-speaker variance returned by the model in (3), $\sum r^2_E$ is the sum of level-2 variances in the first-order model in (2) and σ^2_E is its residual variance.[13]

13 This approach is only appropriate for models that do not include level-1 explanatory variables (see, for example, Snijders and Bosker [2012: 114–117] for further discussion of variance components).

On average, linear effects of TPMs explained about one-fifth of the variance in RMs. The TPMs explained larger fractions of the variances in %V and VarcoΔV than in the other two rhythm measures. For %V, they explained about one-third of variance in Mandarin, French, and Greek. For VarcoΔV, they explained one-third of variance in Russian and Greek and one-quarter of variance in English.

The analysis of coefficients showed that only a small number of TPMs gave slopes that differed significantly from zero.[14] The TPMs that had significant linear effects on the values of RMs were mainly those that indicate the quality of segments rather than syllable structure or stress patterns. Thus %V in French, Greek, and Russian showed a positive linear effect of voiced, while in French and Mandarin it was also affected by Highlow. French alone showed a weak effect of CVVC on %V. VarcoΔV was primarily affected by Highlow-PVI and (surprisingly) by pauses. Only YARD in English showed any dependency on Ccluster and Ccluster-PVI. Full results of fitting the model in (3) appear in Tables A2–A5 in Appendix C.

Since all TPMs were level-2 variables, their addition to the model did not decrease the residual variance (level 1) in the model in (3) compared to that in (2). The average proportion of residual variance across all languages and RMs remained at .41. Since the distribution of TPM values is identical across all speakers, no change could occur in the amount of level-2 variance attributed to between-speaker differences (compare Tables 5 and 6). Indeed, each pair of components R^2_{tpm} and Var_{tn} in Table 6 must sum to the corresponding value of VPC_{txt} in Table 5.

Table 6: The fraction of original variance in RMs attributed to the fixed effect of TPMs (R^2_{tpm} in italics) and to random effects of speaker (Var_{sp}), and net variance due to text after removing effects of TPMs (Var_{txtn}). Boldface indicates a significant effect.

	%V			CrPVI			VarcoΔV			YARD		
	R^2	Var		R^2	Var		R^2	Var		R^2	Var	
Language	*tpm*	sp	txt_n	*tpm*	sp	txt_n	*tpm*	sp	txt_n	*tpm*	sp	txt_n
Russian	***.14***	**.72**	**.03**	*.02*	**.72**	.02	***.26***	**.22**	**.12**	*.08*	**.33**	**.11**
Greek	***.29***	**.56**	.01	***.12***	**.47**	.00	***.31***	**.22**	**.08**	*.14*	**.14**	.05
French	***.34***	**.42**	.00	*.09*	**.55**	.02	*.20*	**.22**	**.08**	*.06*	**.19**	.05
Mandarin	***.36***	**.28**	**.10**	*.05*	**.37**	**.09**	***.18***	**.06**	**.21**	***.14***	**.04**	**.13**
English	*.12*	**.66**	**.07**	*.15*	**.43**	**.09**	***.24***	**.16**	**.11**	***.25***	**.10**	**.08**

14 There are different ways to estimate *p* values for coefficients in multi-level models. We used the pvals.fnc function in the LanguageR package (Baayen 2008: 248), which estimates *p* values using Markov chain Monte Carlo sampling. The number of samples in all cases was set to 10,000.

The addition of TPMs reduced the level-2 variance attributed to the effect of text: across all languages and RMs, average Var_{txtn} decreased from .25 to .07. This shows that our chosen TPMs captured most of the random effect of text. To test whether the remaining net random effect of text was significant, we removed the text effect (u_k) from (3) and used deviance tests to compare the new model with the original one. We found that our TPMs captured the full effect of the text for the 7 combinations of languages and RMs marked by non-significant values of Var_{txtn} in Table 6. For the other 13 combinations, text had an additional small effect beyond the linear effect of TPMs (p varied between 7.9×10^{-44} and .0008).

3.4.4 Between-speaker variance: the effect of gender and speech rate

Having examined TPMs as a level-2 source of between-text variance, we next investigated two possible level-2 factors behind between-speaker variance. First, Benton and Dockendorf (2008) reported that gender affected the values of RMs in their corpus. We therefore tested whether adding gender to (3) would reduce the between-speaker variance. We found no significant effect of gender for any combination of language and RM (average $p = .6$).

Second, we determined whether differences in speech rate explained any between-speaker variance. For each speaker, we calculated average syllable duration (D_{syll}) as the total duration of the speech signal, after excluding pauses, divided by the total number of vowels in transcription. The values of D_{syll} in our corpus varied between 120 and 217 ms. We then added this level-2 explanatory variable to (3), as shown in (6). As with the TPMs, all values of D_{syll} were centred within each language using z-transforms.

$$RM_{i(jk)} = \beta_0 + u_j + u_k + \beta_1 * \text{Ccluster}_k + \cdots + \beta_{12} * D_{syll} + e_{i(jk)} \quad (6)$$

Speech rate had a significant effect (p ranged from .002 to .009) only for CrPVI and %V in English and CrPVI in French, where it explained 17%, 4%, and 33% of between-speaker variance, respectively.[15] For most combinations of languages and RMs, differences in speech rate did not explain the variance between speakers.

The variance components after fitting the final model in (6) appear in Figure 5. In most cases, differences between speakers constituted the main source of variation in RMs: average Var_{sp} across all RMs and languages was .35. For a small

15 Although speech rate appeared to explain one third of the variance in CrPVI in Russian, the likelihood ratio test gave $p = .02$ which is below the level of $\alpha = .01$ used throughout this paper.

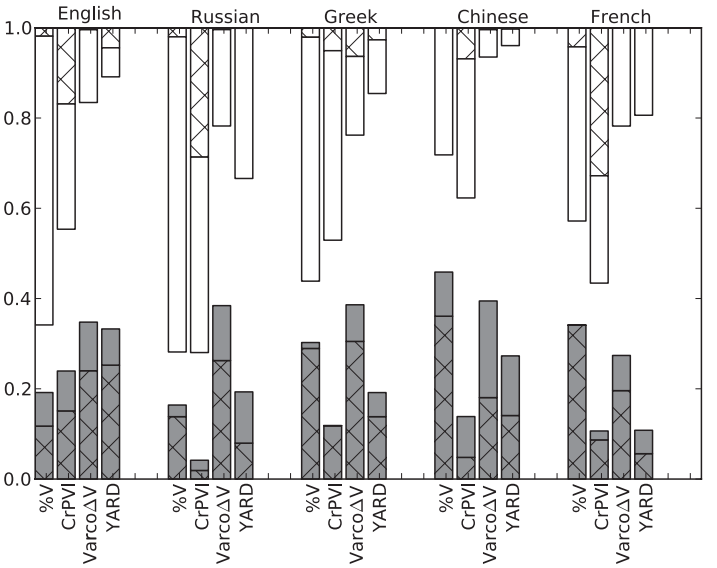

Fig. 5: Partitioning of level-2 variance level by text (grey boxes) and speaker (white boxes) from fitting the model in (3). The vertical space between the boxes corresponds to residual variance (level 1). The grey and white hatched boxes respectively show the fraction of level 2-variance explained by the fixed factors of TPMs and speech rate.

number of combinations of languages and non-normalized RMs, differences in speech rate accounted for some between-speaker variance. Between-text differences underlay about one-quarter of total variance in RMs. Linear effects of TPMs explained most of that variance; removing those effects reduced the average Var_{txtn} from .25 to .07.

3.4.5 The effect of language

We turn finally to the relationships between TPMs and RMs across languages. The phonological account of rhythm holds that cross-linguistic differences in RMs result directly from phonological differences. This theory assumes that the relationships between TPMs and RMs are invariant over all languages.

To test that assumption, we pooled the data for each RM over all five languages. We first fitted the pooled data with the first-order model (2) that only includes random effects of text and speaker. In line with our previous findings (Table 5), substantial between-speaker variance (level 2) explained more than

half of the total variance in %V and CrPVI, a third of the total variance in VarcoΔV, but only one-fifth of the total variance in YARD. Level-2 between-text variance accounted for about 20% of the total variance in each RM.

We then compared these results to the second-order model in (3) that includes fixed linear effects of the TPMs along with random effects of text and speaker. As the phonological account of speech rhythm requires, this model assumes that TPMs have the same effects on RMs across languages. Adding the TPMs improved the fit to the pooled data for each of the four RMs (p ranged from 2.4×10^{-33} to .00025). We calculated R^2_{tpm} through the approach shown in (4). Overall, R^2_{tpm} equalled .25, so that the linear effects of TPMs explained approximately one quarter of the variation in RMs across languages. In particular, the TPMs explained almost 40% of the cross-linguistic variation in %V and VarcoΔV, 17% in CrPVI, and 9% in YARD. These findings resemble the within-language results presented earlier in Table 6.

In the within-language model, values for TPMs did not vary across speakers of a given language. Hence, the addition of TPMs reduced between-text variance but could not affect between-speaker variance. In contrast, values for TPMs in the present cross-linguistic analysis can differ between speakers of different languages. Adding in the effects of TPMs in this analysis did reduce variance between speakers as well as between texts.

So far, the cross-linguistic models have not permitted differences between RMs across languages. To examine this possibility, we developed a final model that allows different mean values of RMs across languages; this model appears in (7). We compared it to the previous model in (3). In (7), language is treated as a level-2 fixed factor, since our sample of languages is not random (see, for example, Snijders and Bosker [2012: 44–48] for an overview of fixed and random factors).

$$RM_{i(jk)} = u_j + u_k + \beta_1 * \text{Ccluster}_k + \cdots + \beta_{11} * \text{Pauses}_k + \beta_{12} * \textit{Russian} + \cdots + \beta_{15} * \textit{Mandarin} + e_{i(jk)} \quad (7)$$

Since we were interested in the effect of individual languages, the model did not include an overall intercept. Instead, it included four dummy variables for Russian, Greek, French, and Mandarin. English was chosen as a reference category. Suppose that differences in an RM across languages were due exclusively to differences in TPMs. There would only be a cross-language linear effect of TPMs on that RM. Then the final model in (7) would not produce a better fit than (3) that contains no language-specific differences in RMs.

Adding the language-specific intercepts slightly reduced both between-speaker variance and between-text variance. It led to a better fit for %V (R^2_{lang} = .07,

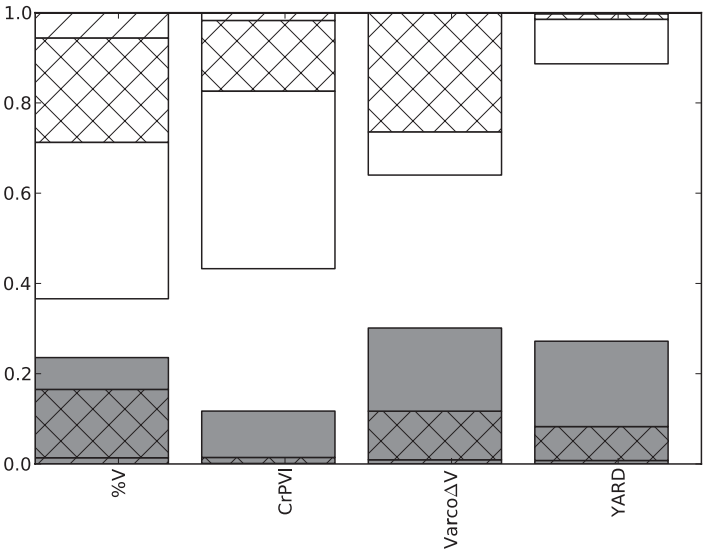

Fig. 6: Partitioning of level-2 variance for text (grey boxes) and speaker (white boxes), from fitting the model in (7) to data from all languages. The vertical space between the boxes corresponds to residual variance (level 1). The hatched boxes show the fraction of level-2 variance explained by fixed factors: diagonally lined boxes show the effect of language and cross-hatched boxes show the effect of TPMs. Note that TPMs and languages varied between both speakers and texts in this model. Thus, the addition of fixed factors reduced both between-speaker and between-text variance.

$p = 5.3 \times 10^{-7}$) but gave no improvement for the other three rhythm measures. The partitioning of variance after fitting the model in (7) appears in Figure 6.

Table 7 gives the coefficients obtained from fitting the model in (7). The phonological properties that had the most effect on cross-linguistic variation on RMs were Voiced, Highlow, Ccluster, and Pauses. The mean values of %V in French and Mandarin differed significantly from predictions based on the value of this measure in English (the reference category) plus the linear effect of TPMs.

In summary, constant cross-language linear effects of TPMs explained some differences in RMs between the five languages in our corpus. The success of this phonologically based explanation, however, varied across the RMs. The TPMs explained about 40% of the cross-linguistic variance in %V and VarcoΔV but only moderate amounts of variation in CrPVI and YARD. In addition, languages showed further differences in mean values of %V beyond those explained by constant cross-language linear effect of TPMs.

Table 7: Parameters obtained from fitting the model in (7). For fixed effects, the numbers in brackets show standard error. Coefficients in bold differed significantly from zero at α = .01. For random effects, numbers in brackets indicate the range of 95% highest posterior density intervals for variances, constructed using the pvals.inc function.[16] Since the reference category for languages is English, Intercept corresponds to the mean value for English.

	%V	Varco ΔV	CrPVI	YARD
Fixed effects				
(Intercept)	**56.09** *(1.15)*	**58.69** *(2.99)*	**5.58** *(0.36)*	**124.02** *(3.34)*
Ccluster	−0.31 *(0.44)*	**−3.5** *(1.28)*	−0.06 *(0.14)*	1.21 *(1.44)*
Ccluster-PVI	−0.01 *(0.23)*	0.26 *(0.68)*	0.00 *(0.08)*	−1.00 *(0.77)*
CVVC	0.28 *(0.17)*	0.76 *(0.49)*	−0.03 *(0.05)*	0.30 *(0.55)*
Sonority	−0.11 *(0.95)*	−4.54 *(2.78)*	−0.45 *(0.31)*	0.58 *(3.11)*
Sonority-PVI	−0.49 *(0.26)*	−1.49 *(0.77)*	−0.18 *(0.09)*	−1.14 *(0.87)*
Voiced	**2.37** *(0.29)*	**2.06** *(0.85)*	0.06 *(0.09)*	−1.90 *(0.95)*
Highlow	**−0.9** *(0.20)*	0.65 *(0.57)*	0.14 *(0.06)*	−0.63 *(0.65)*
Highlow-PVI	0.20 *(0.22)*	−1.03 *(0.64)*	0.02 *(0.07)*	**2.21** *(0.72)*
Strong	−0.84 *(0.96)*	−0.21 *(2.80)*	−0.34 *(0.31)*	−1.56 *(3.14)*
Diphthongs	−0.45 *(0.71)*	4.25 *(2.07)*	−0.12 *(0.23)*	−1.69 *(2.32)*
Pauses	−0.15 *(0.15)*	**−2.4** *(0.43)*	0.06 *(0.05)*	**2.16** *(0.48)*
Russian	−0.11 *(1.42)*	**6.81** *(3.07)*	−0.28 *(0.42)*	−3.43 *(3.42)*
Greek	−0.58 *(2.12)*	**13.36** *(5.49)*	−0.05 *(0.66)*	−7.50 *(6.13)*
French	**−4.5** *(2.03)*	9.53 *(5.20)*	0.50 *(0.63)*	−0.62 *(5.82)*
Mandarin	**12.12** *(2.58)*	1.19 *(7.03)*	0.55 *(0.82)*	−3.13 *(7.85)*
Random effects				
Subject	7.17 *(0.86)*	6.46 *(3.84)*	0.53 *(0.13)*	7.34 *(5.30)*
Text	1.46 *(0.40)*	12.49 *(3.33)*	0.14 *(0.05)*	14.13 *(5.28)*
Residual	2.69 *(0.43)*	23.00 *(3.26)*	0.43 *(0.06)*	45.95 *(6.41)*

3.5 The effect of segmentation

Our results rest on automatic segmentation of the speech signal into vowel-like and consonant-like intervals. The segmentation algorithm is a modification of a previous one for which Loukina et al. (2011) had found substantial agreement with human labellers. As in that work, we checked that the current results are not artefacts of inaccurate automatic segmentation.

16 These confidence intervals merely provide information about the spread of distribution. They are not used to ascertain the significance of any effect, which is done by a deviance test. For further discussion of confidence intervals for random factors, see Snijders and Bosker (2012: 100–101) and Baayen (2008: 248–257).

To do this we obtained manual labels from a group of trained phoneticians for a subset of data consisting of 25 files. Each phonetician was fluent in the language of the recording and labelled the data independently. Eight files were segmented by four labellers each, and another five were segmented by two labellers each. The other twelve files were segmented by one phonetician each. Labelling was at the phone level. We converted the labels into 'V', 'C', and 'S'. Sonorants and approximants were converted into 'V'.

We also obtained labels on another five paragraphs from phoneticians who had no access to the transcriptions and reported no familiarity with the language; this restriction confined the texts to Mandarin and Modern Greek. Two phoneticians labelled each file. They had to segment the file into 'obstruents', 'sonorants and approximants', and 'vowels'.

Cohen's kappa was computed to quantify the agreement between the automatic tags of 'V', 'C', and 'S' and the three recoded categories of human labels. The procedure described in Loukina et al. (2011) was followed.[17] Automatic segmentation and human labelling agreed well. The median K was .74. This is only slightly lower than the agreement between labels from the phoneticians who knew a given language and those from the phoneticians who did not (median $K = .80$).[18]

4 Discussion

Our aim was to test how tightly duration-based rhythm measures are bound to differences in phonology within and between languages. To do this, we examined the quantitative relationships between rhythm measures (RMs) derived from the durational properties of the speech signal and quantified phonological properties of the underlying text in five different languages. The values of these text phonological measures (TPMs) were computed from the transcription of each text. The TPMs were intended to represent various aspects of phonology, such as

17 We treated labels at each 10 ms epoch as separate observations, giving 4000–7000 observations for each test paragraph. We excluded initial and final silences where they were labelled both automatically and manually.

18 Trained phoneticians unfamiliar with a language still differ in many aspects from a segmentation machine. They are experienced in segmental analysis of natural languages and have expectations about the phonological structure of any language. For example, phoneticians expect consistent alternation between vocalic and intervocalic intervals and at least one vocalic interval between two pauses. Furthermore, human labellers may have an implicit understanding about the variation existing between natural languages that would influence their assignment of labels. Our automatic segmentation has no access to such information.

syllable structure. Automatic segmentation was applied to each recording of a text, splitting the recording into vowel-like and consonant-like intervals. The durations of these intervals were used to compute RMs for that recording.

The rhythm measures discussed in this paper had been originally designed to capture phonological differences between languages. Previous studies of the relations between RMs and text (Prieto et al. 2010; Wiget et al. 2010; Arvaniti 2012), however, have reported conflicting results, possibly due to methodological shortcomings. Our analysis addressed such shortcomings in four ways. First, our text phonological measures allowed us to quantify different aspects of phonology and to study any effect of each of these properties separately. Second, our statistical analysis allowed us to separate the effect of text in general from linear effects of individual properties. Third, we used a corpus that is substantially larger than any in previous work, allowing robust statistical analysis and complex models. Fourth and finally, our sampling strategy ensured that the texts selected in each language covered the observed range of values for the text phonological measures. This approach let us determine how well those properties separated the five languages. We could then compare how well the text phonological measures predicted the rhythm measures.

4.1 Language identification by TPMs and RMs

We first turned our attention to how much the values of TPMs and RMs differed between the languages in our corpus. Application of linear discriminant classifiers (see Loukina et al. 2011) based on the TPMs showed that these measures actually reflected differences between our five languages. Furthermore, this effect was large enough so that the classifiers successfully identified the languages more than 80% of the time, with chance at 20%. This result confirms the intuitive assumption that languages can be separated by quantitative phonological properties computed from traditional phonemic transcriptions for each language.

An analysis of individual properties, however, yielded a somewhat surprising result: syllable structure did not constitute the biggest phonological difference between the languages in our corpus. This shows that paradigmatic differences such as the number of intervocalic consonants allowed in a given language do not necessarily lead to what can be called syntagmatic differences, for example, a higher average number of intervocalic consonants in a paragraph. Instead, we found that languages differed most in properties that describe the quality of segments. Examples are Sonority, which distinguishes between obstruents, sonorants, and vowels, and Highlow, which distinguishes between high, low, and middle vowels.

The high rates of language classification (around 80%) by TPMs clearly exceed the results for classifiers based on duration-based RMs in Loukina et al. (2011). The latter classifiers gave correct identifications only about 55% of the time; chance was 23%. Loukina et al. (2011) pointed out that this relatively low success rate is comparable to that reported for humans.

We know that our TPMs do not take into account many phonological factors such as syllable boundaries as well as phonetic factors such as contextual effects. Despite their limitations, the TPMs explained most of the variance in RMs between texts. Our TPMs seem to represent the first systematic attempt to quantify the segmental phonological properties of large stretches of text. They challenge some of the intuitions common both within and outside the field of rhythm studies.

In short, properties computed from phonemic transcriptions allow good separation of different languages. Differences in acoustic durational properties seem less sharp. Languages appear to be more alike when viewed from the standpoint of acoustic implementation than when their phonological structures are considered. This difference arises at least partly from the fact that individual speakers deploy different strategies in the phonetic realization of any language. Conversely, abstract phonological representations may exaggerate differences actually present in speech acoustics or even introduce new differences where the acoustics is very similar.

4.2 TPMs and the variance of RMs

We used mixed linear models to evaluate the contribution of the TPMs to the values of the RMs. These models also let us quantify the effects of different speakers and of different languages on the RMs. We first investigated how much variance in RMs within each language can be attributed to differences between the texts read by the speakers from whose speech the RMs were computed. On the scale of a paragraph, transcribed text accounted, on average, for a quarter of the variation in acoustically based rhythm measures. This total amount of between-text variance places an upper limit on how much variation could be explained by text-based phonological properties.

Our TPMs succeeded in explaining a large share of the between-text variance in RMs. The phonological properties that exerted the most consistent effect on RMs were Voiced and Highlow. Rhythm measures seem more dependent on intrinsic properties of segments than on syllable structure. This is consistent with the fact that the former properties also captured more differences between languages than did the latter and challenges the established view about the nature

of rhythmic differences. For example, syllable structure is generally considered one of the main factors underlying variation in rhythm measures. Therefore, previous studies (Prieto et al. 2010; Arvaniti 2012) constructed sentences with different types of syllables in order to study the effect of text on RMs. Our results show that rhythm measures are in fact more sensitive to quality of the segments such as vowel height or voicing. This may be one of the reasons why previous studies reported conflicting results. Finally, contrary to Wiget et al. (2010), no evidence emerged of a strong connection between rhythm measures and stress patterns as reflected by the TPM Strong.[19]

A more complex model of durational variability (cf. Klatt 1976) might explain between-text variance in RMs in cases where our TPMs fell short (e.g. VarcoΔV in Mandarin). Similarly, the level of detail or choice of alternative pronunciations in transcription could influence the values of TPMs, leading to a slightly different result. The performance of more complex TPMs or TPMs based on different transcriptions would still be limited, however, by the total amount of between-text variance which only averaged about a quarter of the total variance in RMs here. Within a language, variation in the phonological properties of the material does not translate directly into differences in RMs.

The results discussed so far came from treating each language separately, yet RMs were designed to capture differences between languages. To evaluate possible cross-language differences, further analyses were done on the data for each RM pooled across languages. Some cross-linguistic differences in RMs could indeed be explained by the linear effect of TPMs. Differences in VarcoΔV, for example, could partially be explained by difference in Ccluster, so that this RM to some extent reflected cross-linguistic differences in syllable structure. Several other RMs reflected different distribution of high and low vowels as captured by Highlow and Highlow-PVI, and voiced and voiceless segments as captured by Voiced.

These results highlight the role of phonetic factors in the patterns of durational variation across languages. While the fact that segments differ in their intrinsic durations is well known and accepted in phonetic studies, it is often ignored in phonological analysis. We found that segments with different intrinsic durations have different distribution across languages. This is one of the main factors causing between-text variation in rhythm measures. Any future research which aims to study more closely the effect of syllable structure or stress on rhythm must control for such effects of intrinsic durations.

---

19 Note that for reasons explained earlier, we used a different measure of stress than the one described in Wiget et al. (2010).

The relationships between TPMs and RMs across languages, however, were much weaker than usually assumed. While the TPMs explained about 40% of cross-linguistic variation in VarcoΔV and %V, they accounted for only a moderate amount of variation in CrPVI and YARD. Furthermore, %V and VarcoΔV showed an additional difference between languages that went beyond the linear effect of TPMs. This should be taken into account when interpreting the differences in RMs in future studies.

Since languages were treated as fixed factors in these analyses, we emphasize that the results apply only to the five particular languages in our corpus. Our sample is heavily biased towards Indo-European languages, and five languages is too few to be read as applying to all languages. Similar reservations should be applied when interpreting the results of previous rhythm studies, which always have treated language as a fixed factor.

4.3 Speaker effects

While the effect of text on the values of RMs was moderate, we found a substantial effect of speaker. For all measures apart from VarcoΔV, the speaker effect clearly exceeded the effect of text. Variation between speakers sometimes accounted for up to 72% of variation in a given RM. These between-speaker differences were not artefacts of differences in speech rate (cf., for example, Ramus 2002). We found that speech rate was a significant influence on only 3 out of 20 combinations of RMs and languages. Furthermore, this effect was confined to RMs that are not normalized for rate.

Rhythm measures rest on acoustic segmentation of spoken text into vocalic and consonantal intervals. These measures can only reflect phonological properties through the filter of phonetic implementation. Our results show that this mapping is nowhere near as direct and strong as the phonological account of rhythm maintains.

Our results conform to predictions by Arvaniti (2009). Arvaniti argued that it is unlikely that rhythm measures will show straightforward relationships with abstract phonological categories, not only because of the multiple factors mentioned earlier that affect segmental durations, but also because of language-specific complexities of speech timing. While the effect of language in our study was relatively limited, we found that different individuals used different strategies for phonetic implementation of the same text. Therefore, better predictive power could only be achieved by a model that does not assume uniform implementation of phonology in acoustic outputs across all speakers.

Previous studies support this view. For example, Krause and Braida (2004) found that speakers use different strategies for producing clear normal speech (cf. also Kohler [2009] who noted that individual variability in what he calls "rhythmicity of speech" may affect the values of RMs). The substantial effect of speaker would also explain the large variation in the results of previous studies on the reliability of rhythm measures and on their dependence on text. It also would explain why RMs have proved to discriminate poorly between languages.

This result goes beyond the field of rhythm studies. Many phonological theories see phonetic implementation of phonological structures as an automatic process governed by universal and sometimes language-specific constraints. Our results do not support that view. Individual strategies clearly play an important role in that implementation.

Finally, our study has a broad methodological implication. We have demonstrated that the values of RMs are clustered by speaker and text. Therefore, future studies that attempt to model RM – or indeed any other speech measures – should use multilevel models. Inaccurate results will necessarily flow from classical linear regression and other tests that assume independence of data generated by the same speaker or text.

5 Conclusion

The differences in rhythm measures between languages and dialects are often interpreted as reflections of differences in syllable structure or degree of vowel reduction. Our results show that this assumption is incorrect. We found that while languages indeed differed substantially in their phonological properties measured from transcription, these differences did not translate directly into differences in RMs. Languages appeared to be most disparate in properties such as distribution of vowels or voiced consonants rather than in syllable structure, traditionally seen as one of the main phonological bases of rhythmic differences. Text-based phonological properties did explain most of the between-text variance. Such variance, however, accounted for only a quarter of the variation in rhythm measures. This number places an upper limit on the performance of any text-based measures or measures based on different transcriptions, as long as these transcriptions do not vary between speakers. Finally, between-speaker differences gave rise to far more variance in rhythm measures than did phonological properties. These results were consistent across five languages. Different speakers seem to use different strategies for mapping between text and acoustics. This could well be one reason why previous studies in this area have disagreed in their findings.

Acknowledgments: The authors would like to thank John Coleman, Carlos Gussenhoven, Jennifer Cole, and two anonymous reviewers for useful discussions. We also thank Speech Technology Center Ltd. (St. Petersburg, Russia) and the Institute for Speech and Language Processing (Athens, Greece) for their help with automatic transcription of the data. Finally, we thank all the speakers and transcribers for their help with this study. This project was supported by the Economic and Social Research Council (UK) [grant number RES-062-23-1323] and National Science Foundation grants [numbers IIS-0623805, IIS-0534133] awarded to Dr Shih.

References

Arvaniti, Amalia. 2009. Rhythm, timing and the timing of rhythm. *Phonetica* 66(1–2). 46–63.

Arvaniti, Amalia. 2012. The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics* 40(3). 351–373.

Asu, Eva Liina, & Francis Nolan. 2005. Estonian rhythm and the pairwise variability index. In Anders Eriksson & Jonas Lindh (eds.), *Proceedings of FONETIK 2005, Göteborg, 25–27 May 2005,* 29–32. Göteborg: Department of Linguistics, Göteborg University.

Baayen, R. Harald. 2008. Analyzing linguistic data: a practical introduction to statistics using R. Cambridge: Cambridge University Press.

Barry, William, Bistra Andreeva, Michella Russo, Snezhina Dimitrova, & Tanja Kostadinova. 2003. Do rhythm measures tell us anything about language type? In Maria Josep Solé, Daniel Recasens, & Joaquín Romero (eds.), *Proceedings of the 15th ICPhS 2003, Barcelona,* 2693–2696. Barcelona: Causal Productions Pty Ltd.

Barry, William, & Mich Russo. 2003. Measuring rhythm: is it separable from speech rate? In Amina Mettouchi & Gaelle Ferré (eds.), *Actes des interfaces prosodiques,* 15–20. Nantes: Université Nantes.

Barry, William, Bistra Andreeva, & Jacques Koreman. 2009. Do rhythm measures reflect perceived rhythm? *Phonetica* 66(1–2). 78–94.

Bates, Douglas. 2011. Computational methods for mixed models. http://cran.r-project.org/web/packages/lme4/vignettes (accessed 19 January 2012).

Bates, Douglas, Martin Maechler, & Ben Bolker. 2011. *lme4: Linear mixed-effects models using S4 classes.* R package version 0.999375-42.

Bechet, Frederic. 2001. Lia phon: Un syst'eme complet de phon'etisation de textes. *Traitement Automatique des Langues* 42(1). 47–67.

Benton, Matthew, & Liz Dockendorf. 2008. A comparison of two acoustic measurement approaches to the rhythm continuum of natural Chinese and English. In *Proceedings of Interspeech 2008, 9th Annual Conference of the International Speech Communication Association, Brisbane, Australia, September 22–26,* 772–775.

Chao, Yuen Ren. 1968. A grammar of spoken Chinese. Berkeley & Los Angeles: University of California Press.

Dauer, Rebecca. 1983. Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics* 11. 51–62.

Dauer, Rebecca. 1987. Phonetic and phonological components of language rhythm. In *Proceedings of the 11th International Congress of Phonetic Sciences*, 447–450. Tallinn, Estonia: Academy of Sciences of the Estonian S.S.R.

Dellwo, Volker. 2006. Rhythm and speech rate: a variation coefficient for ΔC. In *Language and Language Processing: Proceedings of the 38th Linguistic Colloquium, Piliscsaba 2003,* 231–241. Frankfurt: Peter Lang.

Dellwo, Volker, Adrian Fourcin, & Evelyn Abberton. 2007. Rhythmical classification of languages based on voice parameters. In *Proceedings of the International Congress of Phonetic Sciences (ICPhS) XVI, Saarbrücken, 6–10 August, 2007,* 1129–1132.

Deterding, David. 2001. The measurement of rhythm: a comparison of Singapore and British English. *Journal of Phonetics* 29. 217–230.

Di Cristo, Albert. 1998. Intonation in French. In Daniel Hirst & Albert Di Cristo (eds.), *Intonation Systems: a Survey of Twenty Languages,* 195–218. Cambridge: Cambridge University Press.

Ferragne, Emanuel, & François Pellegrino. 2004. A comparative account of the suprasegmental and rhythmic features of British English dialects. In *Actes de Mod'elisations pour l'Identification des Langues, Paris, 29–30 novembre 2004,* 121–126. Paris.

Fitt, Susan. 2000. Documentation and user guide to UNISYN lexicon and post-lexical rules. http://www.cstr.ed.ac.uk/projects/unisyn/ (accessed 3 February 2011).

Fourakis, Marios, Antonis Botinis, & Maria Katsaiti. 1999. Acoustic characteristics of Greek vowels. *Phonetica* 56(1–2). 28–43.

Grabe, Esther, & Ee Ling Low. 2002. Durational variability in speech and the rhythm class hypothesis. In Carlos Gussenhoven & Natasha Warner (eds.), *Laboratory Phonology 7,* 515–546. Berlin: Mouton de Gruyter.

Hirst, Daniel. 2009. The rhythm of text and the rhythm of utterances: from metrics to models. In *Proceedings of Interspeech 2009, 6–10 September, Brighton, UK,* 1519–1522. Brighton.

Keane, Elinor. 2006. Rhythmic characteristics of colloquial and formal Tamil. *Language and Speech* 49. 299–332.

Klatt, Dennis H. 1976. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America* 59(5). 1208–1221.

Kochanski, Greg, Anastassia Loukina, Elinor Keane, Chilin Shih, & Burton Rosner. 2010. Long-range prosody prediction and rhythm. In *Proceedings of Speech Prosody 2010, Chicago* 100222. 1–4.

Kochanski, Greg, & Christina Orphanidou. 2008. What marks the beat of speech? *Journal of the Acoustical Society of America* 123(5). 2780–2791.

Kohler, Klaus J. 2009. Rhythm in speech and language: a new research paradigm. *Phonetica* 66(1–2). 29–45.

Krause, Jean C., & Louis D. Braida. 2004. Acoustic properties of naturally produced clear speech at normal speaking rates. *Journal of the Acoustical Society of America* 115(1). 362–378.

Lee, Christopher S., & Neil P. McAngus Todd. 2004. Towards an auditory account of speech rhythm: application of a model of the auditory 'primal sketch' to two multi-language corpora. *Cognition* 93(3). 225–254.

Lin, Tao. 1983. A preliminary experimental study on the neutral tone in Beijing Mandarin. *Journal of Linguistic Study* 10. 16–37.

Loukina, Anastassia. 2009. Phonetic variation in spontaneous speech. Vowel and consonant reduction in Modern Greek dialects. In Elinor Payne & 'Ōiwi Parker-Jones (eds.), *Oxford University Working Papers in Phonetics 12,* 36–56. Oxford: University of Oxford.

Loukina, Anastassia, Greg Kochanski, Burton Rosner, Chilin Shih, & Elinor Keane. 2011. Rhythm measures and dimensions of durational variation in speech. *Journal of the Acoustical Society of America* 129(5). 3258–3270.

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

Maddieson, Ian. 1997. Phonetic universals. In W. J. Hardcastle & J. Laver (eds.), *The Handbook of Phonetic Sciences,* 619–639. Oxford: Blackwell.

Peterson, Gordon E., & Ilse Lehiste. 1960. Duration of syllable nuclei in English. *Journal of the Acoustical Society of America* 32(6). 693–703.

Pinheiro, José C., & Douglas M. Bates. 2009. *Mixed-Effects Models in S and S-PLUS. Statistics and Computing.* New York: Springer.

Port, Robert F. 1977. *The Influence of Speaking Tempo on the Duration of Stressed Vowel and Medial Stop in English Trochee Words.* Bloomington: Indiana University Linguistics Club.

Prieto, Pilar, Maria del Mar Vanrell, Lluïsa Astruc, Elinor Payne, & Brechtje Post. 2010. Speech rhythm as durational marking of prosodic heads and edges. Evidence from Catalan, English, and Spanish. *Speech Prosody* 100951. 1–4.

Ramus, Franck. 2002. Acoustic correlates of linguistic rhythm: perspectives. In *Speech Prosody 2002, Aix-en-Provence,* 115–120.

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

Rasbash, Jon, Fiona Steele, William J. Browne, & Harvey Goldstein. 2009. *A user's guide to MLwiN.* Version 2.10. Bristol: Centre for Multilevel Modelling, University of Bristol.

Roach, Peter. 1982. On the distinction between "stress-timed" and "syllable-timed" languages. In David Crystal (ed.), *Linguistic Controversies,* 73–79. London: Edward Arnold.

Scherba, Lev Vladimirovich. 1912. Russkiie glasnye v kachestvennom i kolichestvennom otnoshenii [The quality and quantity of the Russian vowels]. St. Petersburg: Tipografiia Yu.N. Erlikh.

Shih, Chilin, & Richard Sproat. 1996. Issues in text-to-speech conversion for Mandarin. *Computational Linguistics and Chinese Language Processing* 1(1). 37–86.

Snijders, Tom A. B., & Roel J. Bosker. 1994. Modeled variance in two-level models. *Sociological Methods and Research* 22(3). 342–363.

Snijders, Tom A. B., & Roel J. Bosker. 2012. Multilevel analysis: an introduction to basic and advanced multilevel modeling (2nd ed.). London: Sage.

Tilsen, Sam. 2008. Relations between speech rhythms and segmental deletions. *Proceedings from the Annual Meeting of the Chicago Linguistic Society* 44(1). 211–223.

Tilsen, Sam, & Keith Johnson. 2008. Low-frequency Fourier analysis of speech rhythm. *Journal of the Acoustical Society of America* 124(2). EL34–39.

Torgersen, Eivind Nessa, & Anita Szakay. 2012. An investigation of speech rhythm in London English. *Lingua* 122(7). 822–840.

Van Santen, Jan P. H. 1992. Contextual effects on vowel duration. *Speech Communication* 11(6). 513–546.

Wagner, Petra, & Volker Dellwo. 2004. Introducing YARD (yet another rhythm determination) and reintroducing isochrony to rhythm research. In Bernard Bel & Isabelle Marlien (eds.), *Speech Prosody 2004, Nara, Japan,* 227–230.

Wells, John. 1995. Computer-coding the IPA: A proposed extension of Sampa. Unpublished manuscript available at http://www.phon.ucl.ac.uk/home/sampa/ipasam-x.pdf (accessed 30 November 2012).
White, Laurence, & Sven L. Mattys. 2007. Calibrating rhythm: First language and second language studies. *Journal of Phonetics* 35(4). 501–522.
Wiget, Lukas, Laurence White, Barbara Schuppler, Izabelle Grenon, Olesya Rauch, & Sven L. Mattys. 2010. How stable are acoustic metrics of contrastive speech rhythm? *The Journal of the Acoustical Society of America* 127(3). 1559–1569.
Young, Steve J., Gunnar Evermann, Mark J. F. Gales, Thomas Hain, Dan Kershaw, Xunying Liu, Gareth Moore, Julian J. Odell, Dave G. Ollason, Dan Povey, Valtcho Valtchev, & Phil C. Woodland. 2006. *The HTK book*. Version 3.4. Cambridge, UK: Cambridge University Engineering Department. http://htk.eng.cam.ac.uk/docs/docs.shtml (accessed 20 November 2009).

Appendix A

This appendix gives examples of broad phonetic transcription used to select texts and compute TPMs described in Section 2.1.2. As described in 2.1.3, the transcription was generated using automatic transcribers and manually corrected for accuracy. Refer to the main text (Section 4.2) for discussion of the effect of transcription on the results of the analysis.

For automatic analysis we used X-SAMPA, the machine-readable version of IPA (Wells 1995). For this paper the transcriptions were converted to standard IPA. The stress mark where applicable is placed before the stressed vowel. All texts have been adapted from J.K. Rowling, *Harry Potter and the Chamber of Secrets,* and from the corresponding translations.

English

Harry missed the castle, with its secret passageways and ghosts, his classes, the mail arriving by owl, eating banquets in the Great Hall, sleeping in his four-poster bed in the tower dormitory, visiting the gamekeeper, Hagrid, in his cabin next to the Forbidden Forest in the grounds, and, especially, Quidditch, the most popular sport in the wizarding world. All Harry's spellbooks, his wand, robes, cauldron, and top of the line Nimbus Two Thousand broomstick had been locked in a cupboard under the stairs by Uncle Vernon the instant Harry had come home. The Dursleys were what wizards called Muggles, and as far as they were concerned, having a wizard in the family was a matter of deepest shame. Uncle Vernon had even padlocked Harry's owl, Hedwig, inside her cage, to stop her from carrying messages to anyone in the wizarding world.

h'æɹɪ mɪst ðə k'ɑːsɬ | wɪð ɪts s'iːkɹət p'æsɪʤweɪz ənd g'əʊsts | hɪz kl'ɑːsɪz | ðə m'eɪɬ əɹ'aɪvɪŋ b'aɪ 'aʊɬ | 'iːtɪŋ b'æŋkwɪts ɪn ðə gɹ'ert h'ɔːɬ | sl'iːpɪŋ ɪn hɪz fɔːp'əʊstə b'ɛd ɪn ðə t'aʊə d'ɔːmɪtəɹ | vɪzɪtɪŋ ðə g'eɪmkˌiːpə | h'ægɹɪd | ɪn hɪz k'æbɪn n'ɛkst tə ðə fəbɪdn̩ f'ɒɹɪst ɪn ðə gɹ'aʊndz | ənd | ɛsp'əʃəlɪ | kwɪdɪʧ | ðə m'əʊst p'ɒpjuːlə sp'ɔːt ɪn ðə wɪzədɪŋ w'ɜːɬd | 'ɔːɬ h'æɹɪz sp'ɛɬbʊks | hɪz w'ɒnd | ɹəʊbz | k'ɔːɬdɹən | ənd t'ɒp əv ðə laɪn nɪmbəs t'uː θaʊzənd bɹ'uːmstɪk həd b'ɪn l'ɒkt ɪn ə k'ʌbəd 'ʌndə ðə st'eəz b'aɪ 'ʌŋkl̩ v'ɜːnən ðɪ ɪnstənt h'æɹɪ həd k'əm h'eʊm | ðə d'ɜːslɪz wə w'ɒt wɪzədz k'ɔːɬd m'ʌgɬz | ənd əz f'ɑːɹ əz ðeɪ wə kənsˈɜːnd | hævɪŋ ə wɪzəd ɪn ðə f'æmɪliː wəz ə m'ætəɹ əv d'iːpɪst ʃ'eɪm' | 'ʌŋkɬ v'ɜːnən h'æd 'iːvn̩ p'ædlɒkt h'æɹɪz 'aʊɬ | h'ɛdwɪg | ɪns'aɪd hə k'eɪʤ | tə st'ɒp hə fɹəm k'æɹiːɪŋ m'ɛsɪʤɪz tə 'ɛnɪwˌʌn ɪn [ðə wɪzədɪŋ w'ɜːɬd |

Mandarin

兩個小孩順著腳印一步一步的走過兩排點亮的火把進入大廳，耳朵裡只聽到腳步聲在空曠的大廳裡回響。他們甚至不敢抬頭交換一下眼神。廳裡瀰漫著食物的香味。他們穿過溫暖光亮的大廳走向旁邊通往地牢的狹窄石梯，最後在辦公室門前停住了。

ljàŋ kʌ ʨjàʊh ăɪ ʂwânʧʂʌ ʧjàʊîn í pû í pû tʌ ʦòʊ kwô ljàŋ pʰăɪ tjɛ̀nljâŋ tʌ hwŏpà ʧînzû tâtʰ íŋ | ɜ˨two lĭ ʧʂɨ˩ tʰíŋtâʊ ʧjàʊpûʂʌ́ŋ ʦâɪ kʰóŋkʰwân tʌ tâtʰíŋ lì hwĕɪçjàŋ | tʰámʌn ʂânʧʂɨ˩pû kàn tʰăɪtʰǒʊ ʧjáʊhwân íçjû jɛ̀nʂʌ̌n | tʰíŋ lì mĭmân ʧʂʌ ʂɨ˩û tʌ ʨjâŋwêr | tʰámʌn ʧʰwánkwô wʌ́nnwàn kwáŋljân tʌ tâtʰíŋ ʦòʊ ʨjâŋpʰăŋpjén tʰóŋwàŋ tîlăʊ tʌ ʨjă̋ʧʂàɪ ʂɨ˩tʰí | ʦwêɪhôʊ ʦâɪ pânkóŋʂɨ˩ mʌ̌n ʧʰjĕn tʰĭŋʧʂû lʌ |

French

Pendant environ cinq minutes, ils suivirent les bruits de pas qui résonnaient un peu plus loin, puis, soudain, Jedusor s'immobilisa, l'oreille tendue. Harry entendit alors une porte grincer et quelqu'un parler dans un murmure rauque. La voix parut familière aux oreilles de Harry. Tout à coup, Jedusor se précipita en avant. Harry le suivit et vit la silhouette sombre et massive d'un jeune homme accroupi devant une porte ouverte. Une grosse boîte était posée sur le sol.

pãdã ãviʁõ sɛ̃ minyt | il sɥiviʁ le bʁɥi də pa ki ʁezɔnɛtɔ̃e pø ply lwɛ̃ | pɥi | sudɛ̃ | ʒedyzɔʁ simɔbiliza | lɔʁɛj tãdy | aʁi ãtãditalɔʁ yn pɔʁt gʁɛ̃se e kɛlkɔ̃e paʁle dãzɔ̃e myʁmyʁ ʁok | la vwa paʁy familjɛʁ ozɔʁɛj daʁi | tutaku | ʒedyzɔʁ sə pʁesipita ãnavã | aʁi lə sɥivi e vi la silwɛt sõbʁ e masiv dɔ̃ʒœnɔm akʁupi dəvãtyn pɔʁt uvɛʁt | yn gʁos bwat etɛ poze syʁ lə sɔl |

Russian

Гарри вышел через заднюю дверь. День был чудесный, солнечный. Мальчик прошелся по аккуратно подстриженной лужайке, плюхнулся на садовую скамейку и тихонько запел. Ни открыток, ни подарков, и вообще он проведет вечер, притворяясь, будто его не существует. Он горестно уставился на живую изгородь. Больше всего из оставленного в «Хогварце», больше даже, чем по квидишу, Гарри скучал по своим лучшим друзьям, Рону Уэсли и Гермионе Грэнжер. А вот они, как оказалось, совершенно по нему не скучали. За все лето он не получил от них ни строчки, хотя Рон обещал пригласить Гарри к себе погостить.

g'arʲɪ v'ɨʃɨl ʧ'erʲɪz z'andʲuju dvʲ'erʲ | dʲ'enʲ b'ɨl ʧudʲ'esnɨj | s'olnʲrʧnɨj | m'alʲɨʧɨk prʌʃ'olsʲə pə ʌkur'atnə pʌt͡strʲ'iʒɨnej luʒ'ajkʲɪ | plʲ'uxnulsʲə nə sʌd'ovuju skʌmʲ'ejku ɪ tʲɪx'onʲkə zʌpʲ'el | nʲɪ ʌtkr'ɨtək | nʲɪ pʌd'arkəf | ɪ vəʌpʃ:ʲ'e 'on prəvʲɪdʲ'ot vʲ'eʧɪr | prʲɪtvʌrʲ'ajɪsʲ | b'utə jɪv'o nʲɪ suʃ:ɪstv'ujɪt | 'on g'orʲɪsnə ust'avʲɪlsʲə nə ʒɨv'uju 'izgərətʲ | b'olʲʃɨ fsʲɪv'o ɪz ʌst'avlʲɪnəvə f xʌgv'ar͡tsɨ | b'olʲʃɨ d'aʒɨ ʧ'em pə kvʲɪdʲ'iʃu | g'arʲɪ skuʧ'al pə svʌ'im l'uʧʃɨm druzʲj'am | r'onu u'eslʲɪ ɪ gʲɪrmʲɪ'onʲɪ gr'enʒɨr | ʌ v'ot ʌnʲ'i | k'ak ʌkʌz'aləsʲ | səvʲɪrʃ'enə pə nʲɪm'u nʲɪ skuʧ'alɪ | zʌ fsʲ'o lʲ'etə 'on nʲɪ pəluʧ'il ʌt nʲ'ix nʲɪ str'oʧkʲi | xʌtʲ'a r'on ʌbʲɪʃ:ʲ'al prʲɪglʌsʲ'idʲ g'arʲɪ k sʲɪbʲ'e | pəgʌsʲɪtʲ'itʲ |

Greek

Η Ερμιόνη έγινε κατακόκκινη κι άρπαξε από τα χέρια του το ωρολόγιο πρόγραμμα της. Όταν τέλειωσαν το μεσημεριανό τους, βγήκαν στο προαύλιο. Ο ουρανός ήταν συννεφιασμένος. Η Ερμιόνη κάθισε σε ένα πέτρινο πεζούλι κι έχωσε πάλι τη μύτη της στις περιηγησεις με τους Βρικόλακες. Ο Χάρι και ο Ρον έπιασαν κουβέντα για το κουίντιτς, μέχρι που ο Χάρι ένιωσε ένα επίμονο βλέμμα καρφωμένο πάνω του. Γύρισε και είδε το κοντό μελαχρινό αγόρι που είχε δει χθες το βράδυ να βάζει το καπέλο για την επιλογή, να τον κοιτάζει μαγνητισμένο. Κρατούσε σφιχτά κάτι που έμοιαζε με κοινή φωτογραφική μηχανή των μαγκλ.

iermi'oni 'eʝine katak'ocini c'arpakse ap'o ta ç'erʝa tu to orol'oʝio pr'oɣrama tis | 'otan t'eʌosanto mesimerʝan'o tus | vʝ'ikan sto pro'avlio | o uran'os 'itan sinefçazm'enos | i ermi'oni k'aθise se 'ena p'etrino pez'uli c 'exose p'ali ti m'iti tis stis periʝ'isis me tus vrik'olaces | o x'ari ce o ron 'epçasan kuv'end ʝa to ku'idits | m'exri pu o x'ari 'eɲose 'ena ep'imono vl'ema karfom'eno p'ano tu | ʝ'irise ce 'iðe to kond'o melaxrin'o aɣ'ori pu 'içe ð 'i xθ'es to vr'aði na v'azi to kap'elo ʝa tin

epiloj'i |na ton cit'azi maɣnitizm'eno |krat'use sfixt'a k'ati pu 'emɲaze me cin'i fotoɣrafic'i mixan'i ton mangl |

Appendix B

Table A1: Mean values of Text Phonological Measures for languages in our corpus. For ease of interpretation, the table shows raw values computed according to 1. As described in Section 2.1.2, for the rest of the paper the values were multiplied by 100 and centered using *z*-scores. The numbers in brackets indicate standard deviations.

	Russian	**English**	**Greek**	**Mandarin**	**French**
Ccluster	1.43 *(0.04)*	1.58 *(0.04)*	1.30 *(0.04)*	1.40 *(0.11)*	1.32 *(0.06)*
Ccluster-PVI	0.36 *(0.03)*	0.42 *(0.02)*	0.28 *(0.03)*	0.32 *(0.07)*	0.31 *(0.04)*
CVVC	0.05 *(0.01)*	0.04 *(0.01)*	0.10 *(0.02)*	0.04 *(0.03)*	0.06 *(0.02)*
Highlow	2.32 *(0.04)*	2.24 *(0.04)*	2.05 *(0.07)*	2.14 *(0.12)*	1.97 *(0.05)*
Highlow-PVI	0.16 *(0.01)*	0.22 *(0.02)*	0.22 *(0.01)*	0.28 *(0.05)*	0.19 *(0.02)*
Diphthongs	0.14 *(0.04)*	0.27 *(0.04)*	0.03 *(0.01)*	0.49 *(0.08)*	0.12 *(0.04)*
Sonority	2.10 *(0.02)*	2.00 *(0.02)*	2.13 *(0.03)*	2.22 *(0.05)*	2.16 *(0.03)*
Sonority-PVI	0.64 *(0.02)*	0.64 *(0.02)*	0.67 *(0.02)*	0.67 *(0.05)*	0.68 *(0.03)*
Voiced	0.78 *(0.02)*	0.76 *(0.02)*	0.71 *(0.02)*	0.71 *(0.03)*	0.80 *(0.02)*
Strong	0.36 *(0.02)*	0.45 *(0.02)*	0.30 *(0.01)*	0.90 *(0.04)*	0.28 *(0.02)*
Pauses	0.08 *(0.01)*	0.08 *(0.02)*	0.05 *(0.01)*	0.10 *(0.03)*	0.07 *(0.01)*

Appendix C

Tables A2–A5 below show the parameters obtained from fitting the model in (3). For fixed effects, the numbers in brackets show standard error. The coefficients indicated in bold were significantly different from zero at α = .01. For random effects, the numbers in brackets indicate the range of confidence intervals (see footnote to Table 7).

Table A2: %V.

	Mandarin	English	French	Greek	Russian
Fixed effects					
(Intercept)	**63.92** *(0.76)*	**56.61** *(0.61)*	**57.03** *(0.72)*	**56.29** *(0.82)*	**57.65** *(1.02)*
Ccluster	−0.75 *(0.60)*	−0.70 *(0.36)*	−0.61 *(0.46)*	−0.70 *(0.45)*	0.02 *(0.61)*
Ccluster-PVI	−0.13 *(0.34)*	−0.12 *(0.28)*	0.31 *(0.23)*	−0.08 *(0.25)*	0.25 *(0.27)*
CVVC	0.42 *(0.27)*	0.03 *(0.33)*	**0.51** *(0.13)*	0.22 *(0.20)*	0.02 *(0.16)*
Sonority	−1.18 *(1.07)*	0.24 *(0.43)*	−0.91 *(0.52)*	−1.05 *(0.57)*	0.82 *(0.57)*
Sonority-PVI	−1.67 *(0.94)*	−0.55 *(0.32)*	−0.71 *(0.36)*	−0.62 *(0.22)*	0.01 *(0.37)*
Voiced	1.38 *(0.71)*	0.54 *(0.22)*	**2.10** *(0.16)*	**1.73** *(0.29)*	**0.78** *(0.21)*
Highlow	**−0.92** *(0.27)*	0.21 *(0.22)*	**0.76** *(0.14)*	−0.43 *(0.17)*	−0.35 *(0.21)*
Highlow-PVI	−0.01 *(0.34)*	0.28 *(0.24)*	0.08 *(0.11)*	0.19 *(0.15)*	0.05 *(0.19)*
Strong	−0.12 *(0.22)*	0.06 *(0.22)*	0.86 *(0.27)*	−0.11 *(0.12)*	−0.28 *(0.23)*
Diphthongs	0.55 *(0.70)*	−0.59 *(0.36)*	0.07 *(0.26)*	0.02 *(0.19)*	**−0.77** *(0.27)*
Pauses	−0.33 *(0.25)*	−0.04 *(0.34)*	−0.52 *(0.27)*	−0.39 *(0.16)*	−0.07 *(0.18)*
Random effects					
Subject	5.33 *(4.09)*	8.08 *(0.94)*	4.56 *(2.29)*	5.96 *(1.30)*	10.24 *(1.40)*
Textfile	1.85 *(1.22)*	0.92 *(0.85)*	0.00 *(0.49)*	0.14 *(0.51)*	0.37 *(0.66)*
Residual	4.91 *(1.38)*	1.84 *(0.52)*	2.55 *(1.06)*	1.46 *(0.71)*	1.68 *(0.83)*

Table A3: VarcoΔV.

	Mandarin	English	French	Greek	Russian
Fixed effects					
(Intercept)	**55.09** *(0.94)*	**62.08** *(0.56)*	**68.76** *(1.20)*	**70.16** *(1.10)*	**68.13** *(0.93)*
Ccluster	−3.19 *(1.80)*	−0.49 *(0.67)*	−5.69 *(2.24)*	−2.40 *(1.87)*	3.46 *(1.87)*
Ccluster-PVI	−0.58 *(1.02)*	−1.36 *(0.53)*	0.93 *(1.13)*	1.50 *(1.05)*	−1.06 *(0.83)*
CVVC	1.57 *(0.82)*	0.57 *(0.61)*	−0.13 *(0.65)*	−0.20 *(0.82)*	1.01 *(0.50)*
Sonority	−4.86 *(3.23)*	0.26 *(0.80)*	−5.09 *(2.52)*	1.67 *(2.38)*	0.91 *(1.76)*
Sonority-PVI	−0.03 *(2.84)*	0.29 *(0.59)*	−4.73 *(1.76)*	0.28 *(0.90)*	−0.22 *(1.14)*
Voiced	4.82 *(2.12)*	0.80 *(0.42)*	0.15 *(0.78)*	−1.44 *(1.19)*	−0.11 *(0.64)*
Highlow	0.55 *(0.82)*	0.37 *(0.40)*	1.05 *(0.68)*	−1.32 *(0.69)*	−1.21 *(0.65)*
Highlow-PVI	−1.45 *(1.03)*	0.42 *(0.46)*	**−2.23** *(0.54)*	0.75 *(0.64)*	**−2.52** *(0.57)*
Strong	0.09 *(0.66)*	−0.48 *(0.41)*	0.55 *(1.32)*	−1.13 *(0.50)*	−1.47 *(0.72)*
Diphthongs	4.26 *(2.11)*	0.93 *(0.68)*	2.97 *(1.27)*	1.12 *(0.79)*	−1.02 *(0.83)*
Pauses	**−2.36** *(0.76)*	−1.49 *(0.63)*	−2.11 *(1.30)*	**−2.57** *(0.67)*	−0.16 *(0.56)*
Random effects					
Subject	5.04 *(13.79)*	4.45 *(3.96)*	10.76 *(18.82)*	9.25 *(12.06)*	6.77 *(9.24)*
Textfile	16.87 *(10.54)*	2.92 *(4.13)*	3.85 *(8.35)*	3.46 *(5.79)*	3.81 *(4.80)*
Residual	42.56 *(11.50)*	13.17 *(3.07)*	25.12 *(10.39)*	16.02 *(6.62)*	12.45 *(4.94)*

Table A4: CrPVI.

	Mandarin	English	French	Greek	Russia
Fixed effects					
(Intercept)	**4.85** *(0.23)*	**6.35** *(0.16)*	**6.03** *(0.27)*	**5.61** *(0.18)*	**5.98** *(0.28)*
Ccluster	0.07 *(0.17)*	0.17 *(0.13)*	0.15 *(0.23)*	−0.08 *(0.14)*	0.15 *(0.19)*
Ccluster-PVI	0.03 *(0.10)*	0.24 *(0.10)*	−0.17 *(0.11)*	0.21 *(0.08)*	0.02 *(0.08)*
CVVC	−0.05 *(0.08)*	0.10 *(0.11)*	0.08 *(0.07)*	−0.14 *(0.06)*	−0.02 *(0.05)*
Sonority	0.26 *(0.31)*	−0.05 *(0.15)*	−0.36 *(0.25)*	−0.03 *(0.17)*	0.22 *(0.18)*
Sonority-PVI	0.01 *(0.27)*	−0.01 *(0.11)*	−0.12 *(0.18)*	−0.05 *(0.07)*	0.06 *(0.11)*
Voiced	0.04 *(0.20)*	0.07 *(0.08)*	−0.04 *(0.08)*	−0.01 *(0.09)*	−0.07 *(0.06)*
Highlow	0.04 *(0.08)*	0.18 *(0.08)*	−0.11 *(0.07)*	0.16 *(0.05)*	−0.08 *(0.07)*
Highlow-PVI	0.08 *(0.10)*	0.10 *(0.09)*	0.00 *(0.05)*	−0.07 *(0.05)*	0.02 *(0.06)*
Strong	−0.08 *(0.06)*	−0.05 *(0.08)*	0.04 *(0.13)*	0.11 *(0.04)*	0.02 *(0.07)*
Diphthongs	−0.40 *(0.20)*	0.03 *(0.13)*	0.05 *(0.13)*	0.18 *(0.06)*	−0.03 *(0.08)*
Pauses	0.04 *(0.07)*	**0.34** *(0.12)*	−0.03 *(0.13)*	0.03 *(0.05)*	0.02 *(0.06)*
Random effects					
Subject	0.52 *(0.57)*	0.50 *(0.19)*	0.63 *(0.38)*	0.28 *(0.23)*	0.78 *(0.23)*
Textfile	0.13 *(0.13)*	0.11 *(0.13)*	0.02 *(0.09)*	0.00 *(0.03)*	0.02 *(0.08)*
Residual	0.67 *(0.18)*	0.38 *(0.09)*	0.39 *(0.17)*	0.25 *(0.10)*	0.26 *(0.11)*

Table A5: YARD.

	Mandarin		English		French		Greek		Russian	
Fixed effects										
(Intercept)	**122.33**	*(1.01)*	**123.50** *(0.59)*		**121.17**	*(1.28)*	**119.74**	*(0.87)*	**118.03**	*(1.36)*
Ccluster	4.20	*(2.04)*	**−2.62** *(0.79)*		1.20	*(2.48)*	−1.17	*(1.78)*	−0.23	*(2.30)*
Ccluster-PVI	−1.72	*(1.16)*	**3.46** *(0.61)*		0.50	*(1.25)*	0.85	*(1.00)*	−0.15	*(1.02)*
CVVC	−0.69	*(0.93)*	1.58 *(0.71)*		0.54	*(0.72)*	−0.08	*(0.78)*	−0.08	*(0.61)*
Sonority	3.26	*(3.67)*	−1.33 *(0.93)*		1.93	*(2.79)*	−2.74	*(2.27)*	0.77	*(2.16)*
Sonority-PVI	−4.21	*(3.22)*	0.17 *(0.70)*		1.36	*(1.95)*	−0.97	*(0.85)*	−0.42	*(1.40)*
Voiced	−4.51	*(2.41)*	−0.96 *(0.49)*		−1.50	*(0.86)*	0.43	*(1.13)*	−0.58	*(0.78)*
Highlow	−0.53	*(0.93)*	−0.39 *(0.47)*		−0.46	*(0.75)*	1.72	*(0.66)*	−0.53	*(0.80)*
Highlow-PVI	**4.35**	*(1.17)*	−0.14 *(0.53)*		−0.61	*(0.60)*	−0.57	*(0.61)*	0.38	*(0.70)*
Strong	−0.07	*(0.75)*	0.56 *(0.47)*		1.65	*(1.46)*	**1.54**	*(0.47)*	0.31	*(0.88)*
Diphthongs	−5.02	*(2.39)*	0.31 *(0.80)*		0.07	*(1.41)*	0.84	*(0.75)*	0.59	*(1.01)*
Pauses	**3.35**	*(0.87)*	**3.23** *(0.74)*		−0.88	*(1.44)*	0.20	*(0.64)*	0.65	*(0.69)*
Random effects										
Subject	5.31	*(18.40)*	4.42 *(5.42)*		12.17	*(27.11)*	5.28	*(12.88)*	15.79	*(17.60)*
Textfile	17.94	*(18.04)*	3.68 *(6.72)*		3.26	*(14.24)*	2.00	*(7.37)*	5.37	*(8.26)*
Residual	93.30	*(24.22)*	25.54 *(5.78)*		43.83	*(17.54)*	24.70	*(10.02)*	22.36	*(8.78)*