Rhythm in Read British English: Interdialect Variability

Emmanuel Ferragne, François Pellegrino

Laboratoire Dynamique Du Langage, UMR 5596 CNRS/Université Lyon 2
Lyon, France
Emmanuel.Ferragne@univ-lyon2.fr
Francois.Pellegrino@univ-lyon2.fr

Abstract

Duration features have been thought to be the most obvious correlates of speech rhythm. Previous studies have shown that they can be used to distinguish among some world's languages. This paper investigates to what extent the methods employed in these studies can be applied to the dialects of British English. We have tested whether a set of variables derived from automatically extracted duration measurements constitute reliable predictors that could be used for automatic dialect identification. Preliminary results show that the automatic procedure, combined with the high interspeaker variability, yield overlapping rather than crisp dialectal categories.

1. Introduction

Over the past sixty years or so, endeavours have been made to investigate the speech signal in order to find empirical evidence for a perceived difference in the timing of speech among the world's languages. Early typological studies concentrated on the rhythm class hypothesis, according to which languages fell into either of two clear-cut – now well-known – categories: "stress-timed" or "syllable-timed", e.g., respectively, English and French ([1],[2]). The so-called "mora-timed" class was eventually added to account for languages such as Japanese (see [3], [4], [5] and [6] for reviews).

Nowadays, the study of speech rhythm has also become a key challenge in speech technology since most of automatic speech processing systems have to cope with the variability of speech rate and rhythm and their consequences both on the segmental units and suprasegmental organization of speech. Applications range from speaker adaptation of automatic speech recognition systems to automatic modelling of rhythm or prosody in a language identification perspective.

In recent studies, new approaches to gauging rhythmic differences across languages have been proposed. What emerges from the findings ([3] and [6]) is that a set of variables derived from duration measurements can indeed help discriminating between some languages. These methods have also been shown to be applicable to varieties of the same language ([7], [8]). Besides, the traditional, crisp rhythm categories have been seriously questioned ([3],[5]).

This paper is a preliminary report on our work on speech timing in relation to the dialects of British English within an automatic dialect identification framework. We address the question of whether the dialects of British English differ in terms of rhythm and whether we can automatically extract relevant acoustic cues that would account for this difference.

Further elements relative to the linguistic framework of this study are given in Section 2. The corpus and rhythm measurement algorithms are detailed in Section 3. Results and discussion appear in Section 4.

2. Scientific framework

2.1. Rhythm of the British dialects

Little is said about rhythmic differences in the literature that deals with traditional English dialectology, although it is widely acknowledged that such differences do exist between the dialects of British English. More specifically, it seems that some Northern British accents have a tendency to retain full vowels in some unstressed environments where other varieties have a reduced vowel [9]. Moreover, Scottish and Ulster English have no contrastive vowel length (recall however that the former is known to have context-conditioned vowel lengthening) [9]. Our hypothesis was that these phonetic peculiarities might well contribute to Northern speech exhibiting a different, more "syllable-timed", rhythm effect than Southern speech. We therefore hypothesize that:

(i) it is the variability of vowel duration that will more adequately capture rhythmic differences between our dialects,
(ii) the variability of consonant duration will provide poor discriminatory power since it may be more of a between-language feature.

2.2. Measuring rhythm

Several approaches to measuring rhythm have been described in the literature. In the case of British dialects, differences are likely to be fairly small since mutual intelligibility is preserved to a fairly high extent between these regional accents. For this reason, we select the PVI approach, described in [3], which is a more local index than those considered in [6]. To put it differently, as [10] points out, the measures proposed in [6] – especially the standard deviation of consonantal intervals – actually measure the overall phonic impression produced by varying syllabic complexity and diversity across languages; and it is highly questionable whether syllabic complexity varies greatly across the dialects of British English. Instead, we want an index that more adequately captures the *sequential* nature of speech rhythm. However, in another study concerned with rhythm differences between languages, we found [11] that the standard deviation of consonantal intervals was highly correlated with raw consonantal PVI, and the standard deviation of vocalic intervals is highly correlated with the raw vocalic PVI, which suggests that the two types of measures actually provide different ways of looking at the same thing.

For instance, the normalized vowel-duration pairwise variability index (nPVIv henceforth) measures the mean

difference in duration between two successive vowels over a whole utterance (or passage). We predicted that, in accordance with what has just been said about Northern British accents, this index would yield comparatively low values for Northern accents and high values for Southern accents.

3. Corpus and method

3.1. Corpus

For our measurements, we used a subset of the Accents of the British Isles (ABI) corpus [12]. The ABI database comprises recordings from 14 dialectal areas throughout the British Isles, and 20 speakers on average for each dialect (10 males and 10 females). The subset we used consisted of 3 short passages of read speech (approx. 40 seconds each) per speaker with all speakers and all dialects (i.e., 852 such passages). For the purpose of comparison, we also made measurements on a sample of French from the Eurom database (read speech, 10 speakers) [13]. In several cases, passages from the ABI database contain spurious and non-linguistic noises (laugh, cough, etc.) and some parts are repeated twice for a certain number of speakers. The corpus is therefore less constrained than other corpora available for similar purposes, which increases the variability of the phonetico-phonological content across speakers.

3.2. Method

3.2.1. Segmentation and labelling

The method we adopted required no prior hand- segmentation or labelling. The sound files were normalized for amplitude and automatically segmented. Then a speech activity detector and a vowel detection algorithm were applied to the data (see [14] for a comprehensive description of the algorithms). Here it should be noted that the algorithms are based on the detection of abrupt breaks occurring in the waveform. Consequently, vowel segments tend to be reduced to their steady part and transitions tend to be considered as consonants. In other words, vowel duration is underestimated and it is possible that long vowels are more affected by this bias than short reduced vowels (with very short transitions). Moreover, the vowel detection algorithm fails to detect unvoiced and very short vowels. The resulting consonant/vowel segmentation was then exported to Praat format and computations were carried out using Praat scripts.

3.2.2. Raw duration measurements

Vowel duration (Dv) and the duration of intervocalic intervals (Dc) were computed since these features have been shown to be useful for automatic language identification [15].

3.2.3. Pairwise variability indices

Following the method described in [3], adjacents segments of the same type (vowels or consonants) were merged into one single vocalic/consonantal interval for the computation of PVIs. The following indices were derived from the signal: the raw consonant-duration PVI (rPVIc) and the normalized vowel-duration PVI (nPVIv). We used the same formulas as [3]:

$$rPVIc = \left[\sum_{i=1}^{n-1} \left| Dc_i - Dc_{i+1} \right| / (n-1) \right] \quad (1)$$

$$nPVIv = 100 \times \left[\sum_{i=1}^{n-1} \left| \frac{Dv_i - Dv_{i+1}}{(Dv_i + Dv_{i+1})/2} \right| / (n-1) \right] \quad (2)$$

where n is the number of vocalic (v) or consonantal (c) intervals of a (vocalic or consonantal), D is the duration measured on the ith interval. Note that whenever c_i and c_{i+1} (or v_i and v_{i+1}) were separated by a pause, the pair was discarded. This was done in order to exclude extreme values resulting from phrase-final lengthening

4. Results

Grabe and Low [3] found that both rPVIc and nPVIv were relevant dimensions that separate some of the world's languages (especially the latter). As a preliminary test, we checked whether the two dimensions would provide a satisfactory picture for British dialects. Figure 1 shows the mean values for each speaker on the rPVIc/nPVIv plane (the abbreviations for the names of dialects appear in Table 1). For the purpose of readability, only 3 – randomly chosen – dialects (plus French) were plotted. Recall that we are working with automatically-segmented data and therefore, bearing in mind the limitations of the algorithms already mentioned, our absolute values cannot be directly compared with those obtained with manual segmentation. What stands out immediately is that speakers of the same dialect do not cluster together. Incidentally, rPVIc seems to be more relevant than nPVIv when it comes to discriminating between French and English. This is quite understandable because rPVIc can be regarded as an indicator of syllabic complexity.

Since it is quite obvious at a glance that the two dimensions considered in this preliminary check offer but poor – if any – discriminatory power, we decided to explore other variables (dimensions). What is more, given the various shapes of the underlying distributions for PVIs (non normality) we believed that using averages over all passages was not very informative, so we computed a series of Kruskal-Wallis tests (the non-parametric one-way ANOVA) for each dimension taken separately including *all* values for *each* pair of segments:

$$rPVIc' = \left| Dc_i - Dc_{i+1} \right| \quad (3)$$

$$nPVIv' = 100 \times \left| \frac{Dv_i - Dv_{i+1}}{(Dv_i + Dv_{i+1})/2} \right| \quad (4)$$

4.1. Consonantal duration and PVI

The Kruskal Wallis test was used to compare average Dc between dialects. The difference in Dc across dialects (plus French) was shown to be highly significant ($p<0.001$). Pairwise post-hoc testing reveals (as is obvious from Figure 2) that French is clearly separated from all English dialects. Now, focusing on the latter, it appears that some dialects exhibit Dc

values that are statistically different from the Dc values obtained in other dialects.

Concerning rPVIc', the test once again returns highly significant values ($p<0.001$). rPVIc' quite obviously appears to be a reliable cue for discriminating between English and French. The same conclusion as in the previous paragraph is reached after post-hoc testing (see Figure 3): rPVIc' allows robust between-language discrimination. This confirms hypothesis (ii) which stated that consonant duration is a reliable cue for the discrimination of languages.

4.2. Vocalic duration and PVI

Statistical differences also appear on the Dv dimension ($p<0.001$). However, as Figure 4 suggests, Dv fails to be a reliable between-language discriminator.

On the nPVIv' dimension, we achieve statistical significance ($p<0.001$). The results that we obtain between dialects suggest, as was the case for rPVIc' and Dc, that some dialects are distinct from others. Here, hypothesis (i) is confirmed.

4.3. Dialect identification

Now, turning to the question of British dialects, the analysis reveals that some of them exhibit differences with some others (on the dimensions studied). Although total discrimination has not been achieved and what we get, instead, is a continuum, we expected that a certain pattern – or at least tendencies – would emerge: for instance, we thought that Northern dialects would cluster together and so would Southern dialects in accordance with partitions often mentioned in the literature. As it turns out, no such rhythmic isogloss can be drawn. We will not go into too much detail here, but the Figures do not seem to support diachronic hypotheses: contemporary dialects from the same traditional Middle-English dialectal areas do not seem to pattern together. However, the question needs further examination.

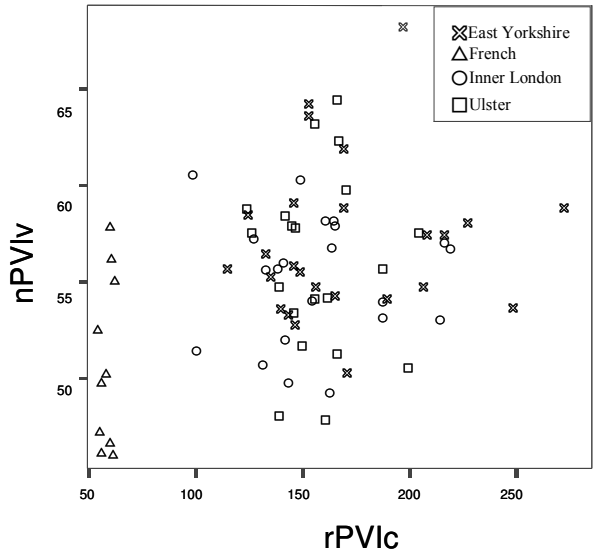

Figure 1: rPVIc/nPVIv plot .

5. Discussion and conclusion

Dc, rPVIc', and nPVIv' indisputably constitute reliable discriminators between French and English as has already been demonstrated in the literature [2]. This comes as no surprise since the two languages have always been described as prototypes of two different rhythm classes: English has vowel reduction whereas French does not, and English has more complex syllables than French. And this is precisely what the indices used in our study are measuring.

As for the measurement of rhythm, we believe that duration cannot be expected to tell us more than half the story. Therefore, it is indispensable to include other acoustic cues (such as intensity and pitch) to better comprehend what rhythm is and how it can be modeled for automatic dialect identification. PVIs undoubtedly constitute – intuitively – good indicators of what rhythm really is. Now, going back to hypotheses (i) and (ii), our results provide evidence that vowel duration and nPVIv' constitute adequate dimensions for both automatic dialect identification and automatic language identification whereas consonant duration and rPVIc' are not well-suited for the automatic identification of the dialects of British English. In other words, we observe greater between-dialect variability for the measures derived from vowel duration than those derived from consonant duration, which supports our prior expectations.

6. References

[1] K. L. Pike, *The intonation of American English*. Ann Arbor: University of Michigan Press, 1945.

[2] D. Abercrombie, *Elements of general phonetics*. Edimbourg: Edinburgh University Press, 1967.

[3] E. Grabe and E. L. Low, "Durational variability in speech and the rhythm class hypothesis," in *Papers in Laboratory Phonology 7*, C. Gussenhoven and N. Warner, Eds. Cambridge: CUP, 2002.

[4] A. Pamies Bertrán, "Prosodic typology: on the dichotomy between stress-timed and syllable-timed languages," *Language Design*, pp. 103-130, 1999.

[5] R. Dauer, M., "Stress-timing and syllable-timing reanalyzed," *Journal of Phonetics*, pp. 51-62, 1983.

[6] F. Ramus, M. Nespor, and J. Mehler, "Correlates of linguistic rhythm in the speech signal," *Cognition*, vol. 73, pp. 265-292, 1999.

[7] S. Ghazali, R. Hamdi, and M. Barkat-Defradas, "Speech rhythm variation in Arabic dialects," presented at Speech Prosody, Aix-en-Provence, France, 2002.

[8] D. Deterding, "The measurement of rhythm: a comparison of Singapore and British English," *Journal of Phonetics*, vol. 29, pp. 217-230, 2001.

[9] J. C. Wells, *Accents of English. The British Isles*, vol. 2. Cambridge: Cambridge University Press, 1982.

[10] P. S. Wagner and V. Dellwo, "Introducing YARD (Yet Another Rhythm Determination) and re-introducing isochrony to rhythm research," presented at Speech Prosody, Nara, Japan, 2004

[11] R. Hamdi, M. Barkat-Defradas, E. Ferragne, and F. Pellegrino, "Speech Timing and Rhythmic structure in Arabic dialects: a comparison of two approaches," Interspeech 2004, Jeju Island, Korea, submitted.

[12] 20/20 Speech, "Accents of the British Isles", www.aurix.com.

[13] E. Campione and J. Véronis, "A multilingual prosodic database," presented at ICSLP, Sydney, 1998. www.lpl.univaix. fr/projects/multext.
[14] F. Pellegrino and R. André-Obrechet, "Automatic language identification: an alternative approach to phonetic modelling," *Signal Processing*, vol. 80, pp. 1231-1244, 2000.
[15] J.-L. Rouas, J. Farinas, and F. Pellegrino, "Automatic modeling of rhythm and intonation for language identification," presented at ICPhs, Barcelona, Spain, 2003.

Table 1: *Dialects of the ABI database*

Location	Abbreviation
Birmingham	brm
Cornwall	crn
East Anglia	ean
East Yorkshire	eyk
Glasgow	gla
Inner London	ilo
Lancaster	lan
Liverpool	lvp
Newcastle	ncl
North Wales	nwa
Republic of Ireland	roi
Scottish Highlands	shl
Standard Southern English	sse
Ulster	uls

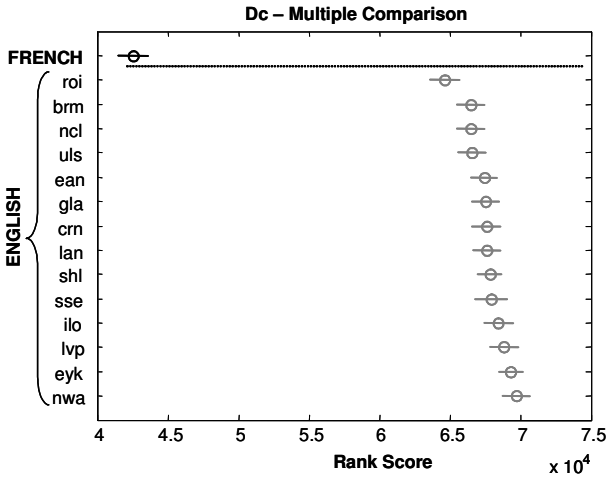

Figure 2: duration of consonantal intervals

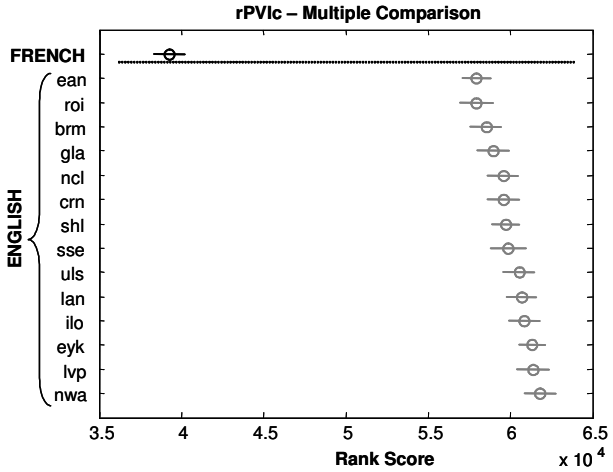

Figure 3: raw consonantal PVI

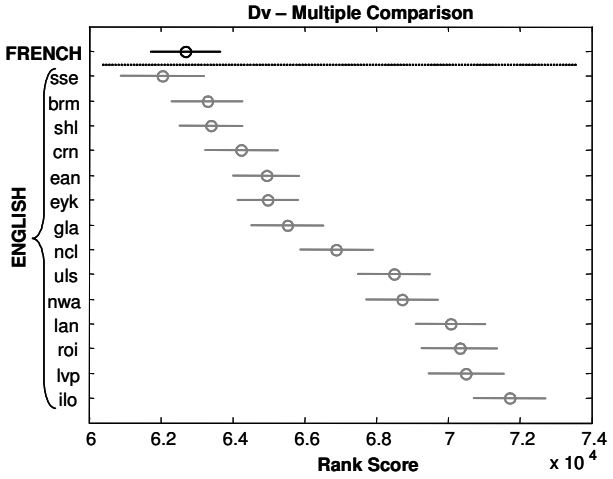

Figure 4: vowel duration

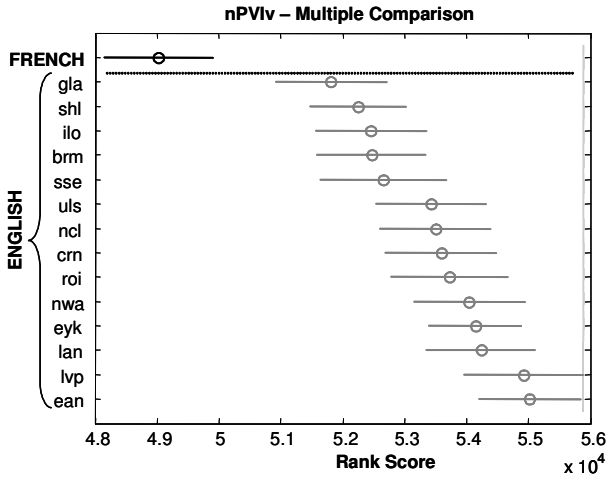

Figure 5: normalized vocalic PVI