How far can phonological properties explain rhythm measures?

Elinor Keane[1], Anastassia Loukina[1], Greg Kochanski[1], Burton Rosner[1], Chilin Shih[2]

[1] Oxford University Phonetics Laboratory, UK
[2] EALC/Linguistics, University of Illinois, Urbana-Champaign, USA

Presented at the 2010 Colloquium of the British Association of Academic Phoneticians, 28-31 March 2010, London, UK. This can be downloaded from the Oxford Research Archive, http://ora.ouls.ox.ac.uk , or http://kochanski.org/gpk/papers/2010/BAAP/Keane .

Introduction

Phonological properties

?

↔

Rhythm measures

e.g.

complexity of consonant clusters

?

↔

variability in duration of consonantal intervals

Introduction

UNIVERSITY OF OXFORD

How far can rhythm measures (RMs) be predicted from the phonological properties of texts?

How closely are phonological properties and RMs correlated?

→ text-dependence of RMs

How much variability in RMs is there between speakers?

→ speaker-dependence of RMs

How much variability in RMs is there between languages?

→ language-dependence of RMs

Corpus: speakers

- 62 speakers:
  - 24 British English (Southern England)
  - 10 Russian (Moscow/St.-Petersburg)
  - 10 Taiwanese Mandarin (Taipei)
  - 9 Modern Greek (Athens)
  - 9 French (Paris)
- 20-28 years old
- <4 years outside their home country

Corpus: texts

UNIVERSITY OF OXFORD

- 40 short texts for each language:
  - paragraphs from '*Harry Potter and the Chamber of Secrets*'
  - Aesop's fables
  - children's poetry

= 2730 recorded paragraphs

- sentences from '*Harry Potter and the Chamber of Secrets*' for each language

= 22,899 recorded sentences

Segmentation

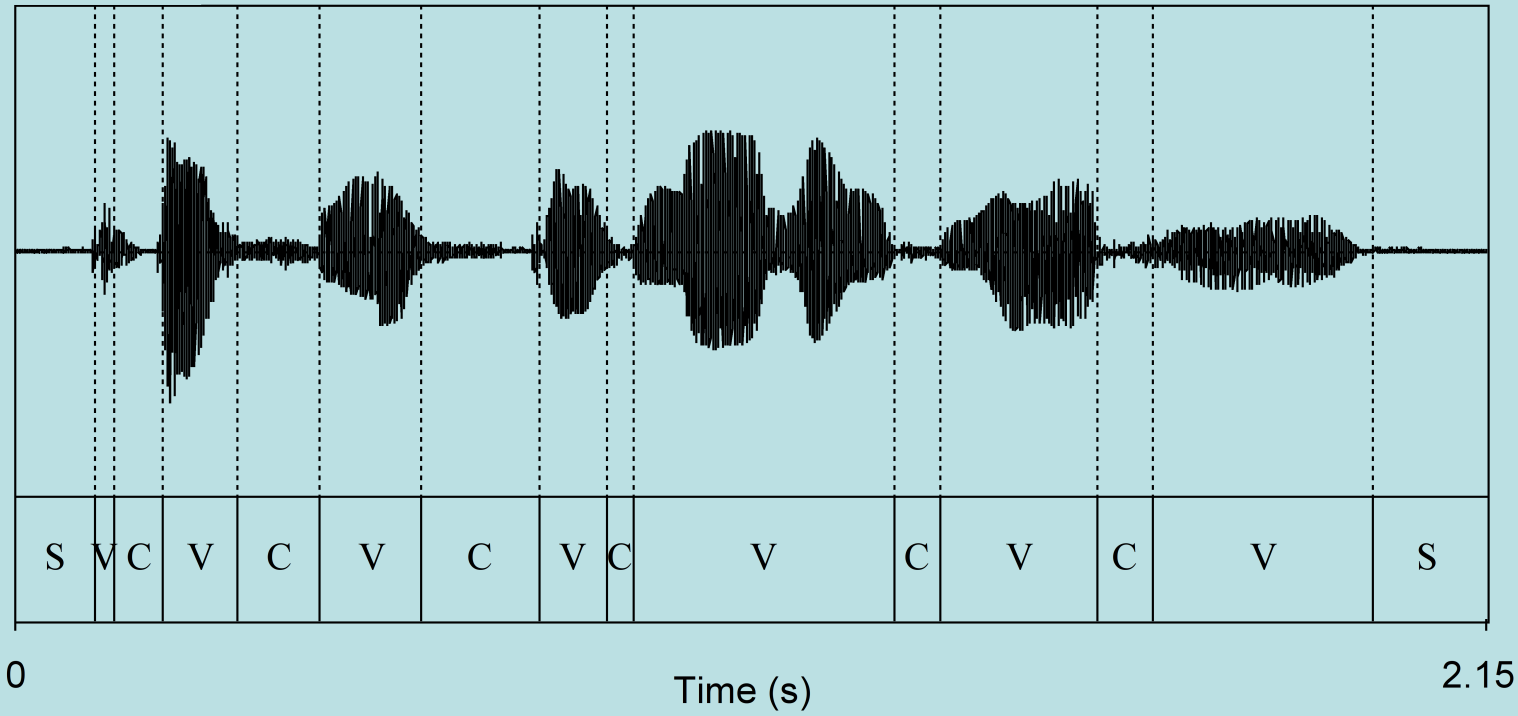
UNIVERSITY OF OXFORD

- automatic, using the HTK toolkit
- cross-linguistic
- divided speech into consonant-like, vowel-like and silent intervals

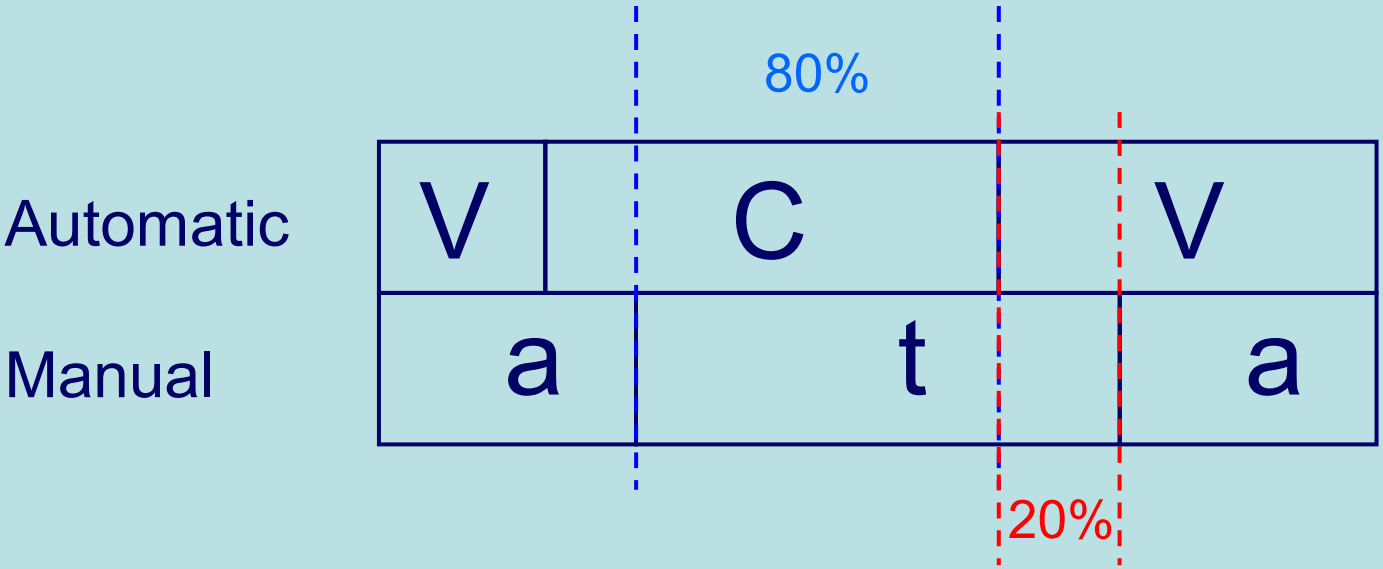
Segmentation

UNIVERSITY OF OXFORD

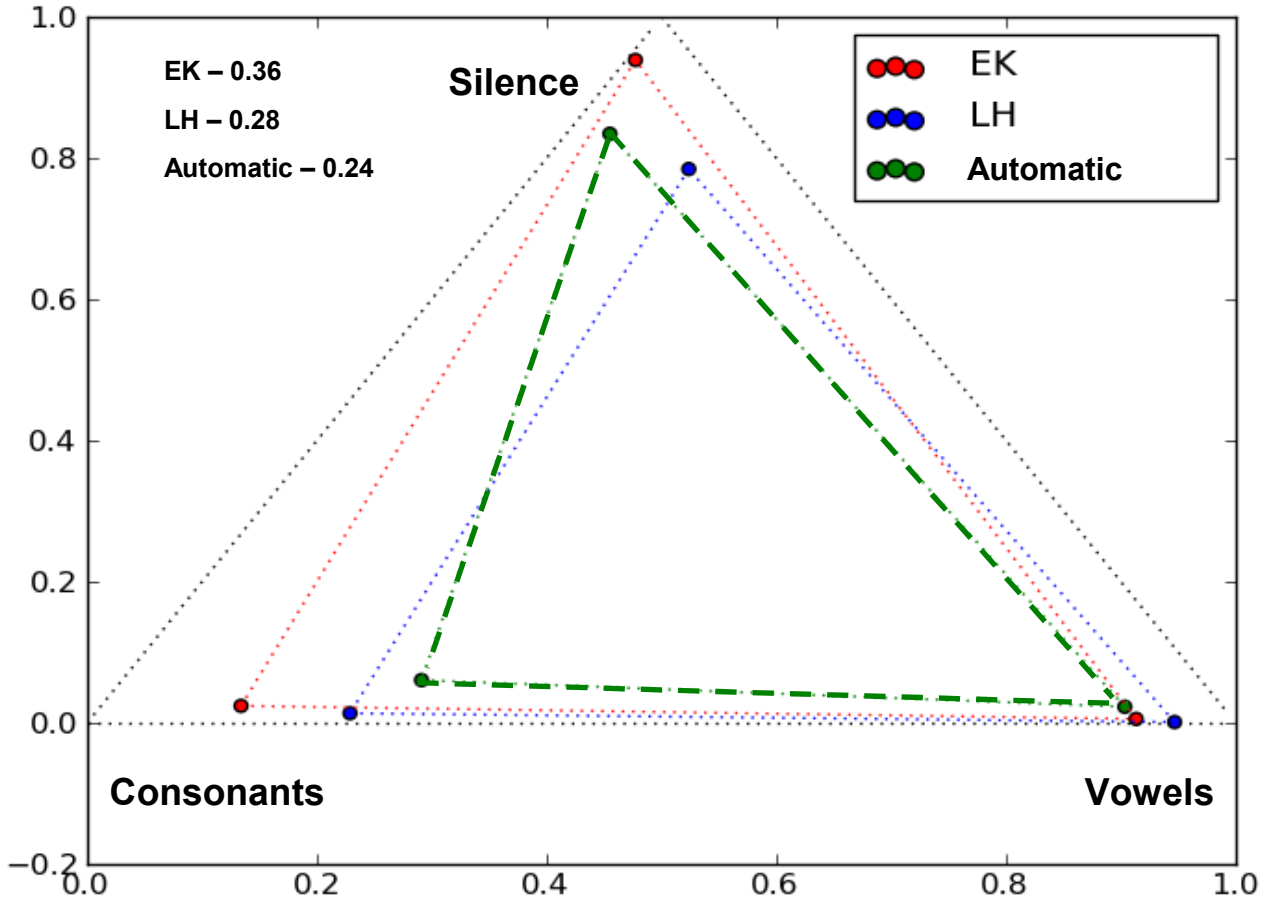
‘The basilisk was moving toward Harry.’



Comparing segmentations



Comparison between segmentations



Rhythm measures

UNIVERSITY OF OXFORD

- statistical indices based on temporal properties (Ramus et al. 1999; Grabe & Low 2002)

- acoustically defined

- typically rely on manual segmentation into vocalic and consonantal intervals

Examples

UNIVERSITY OF OXFORD

- global measures, e.g:
  - %V – percentage of vocalic intervals
  - ΔC – standard deviation of consonantal intervals
  - Vdur/Cdur – ratio of vowel to consonant duration
- PVI-like measures, e.g:
  - CrPVI – raw consonantal pairwise variability index
  - PVI-CV – PVI of consonant+vowel group

Phonological properties

- computed from roughly phonemic transcriptions of the texts
- cross-linguistically defined
- plausibly have a direct effect on rhythm measures

Examples

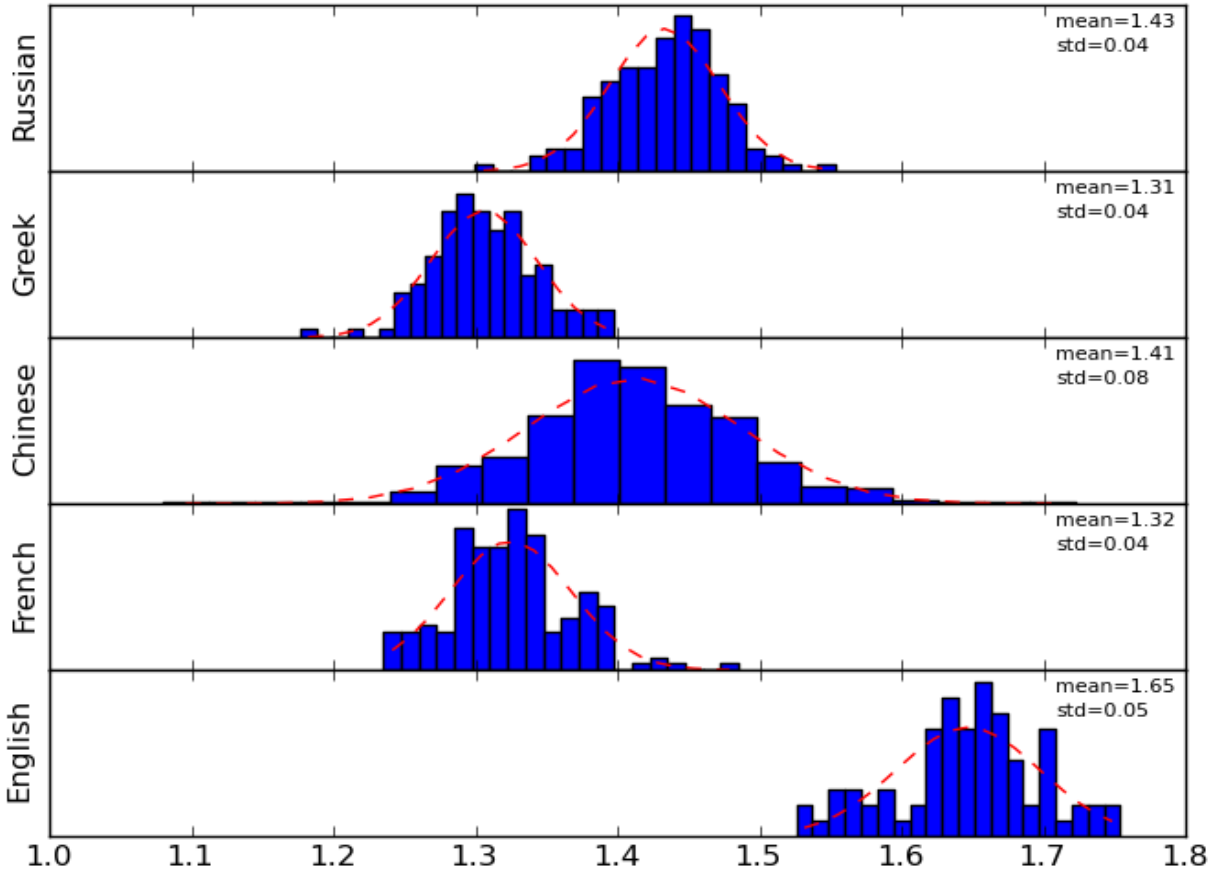
UNIVERSITY OF OXFORD

- **ccluster** – mean number of consonants between vowels
- **voiced** – fraction of voiced segments in total number of segments
- **sonority** – mean sonority index based on scale:

  [stops, fricatives, affricates] – 1
  [nasals, liquids] – 2
  [glides, vowels] – 3

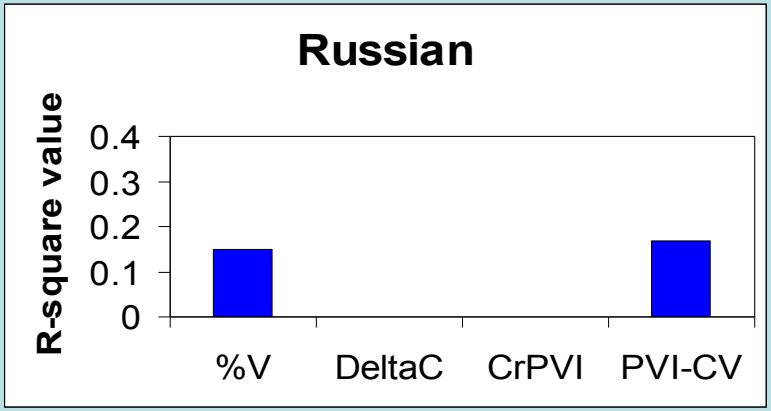
- **pvi-variants**, e.g. ccluster_pvi – mean-square difference between numbers of consonants on adjacent inter-vowel gaps

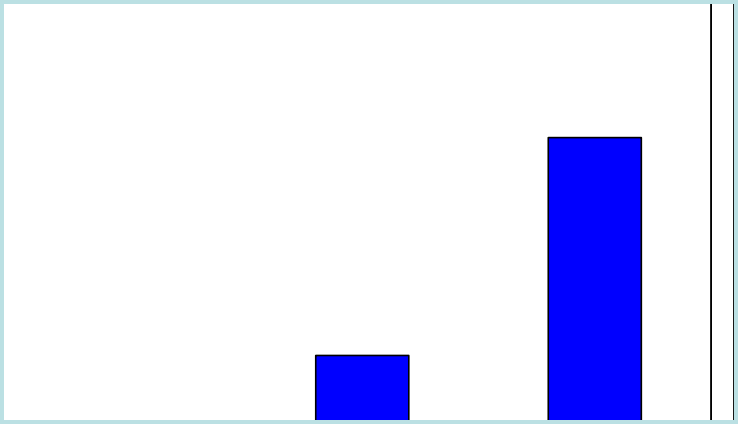
Ccluster for ‘Harry Potter’ paragraphs

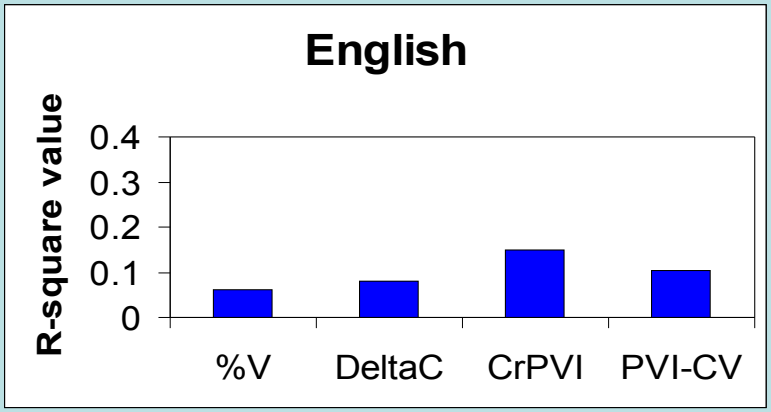


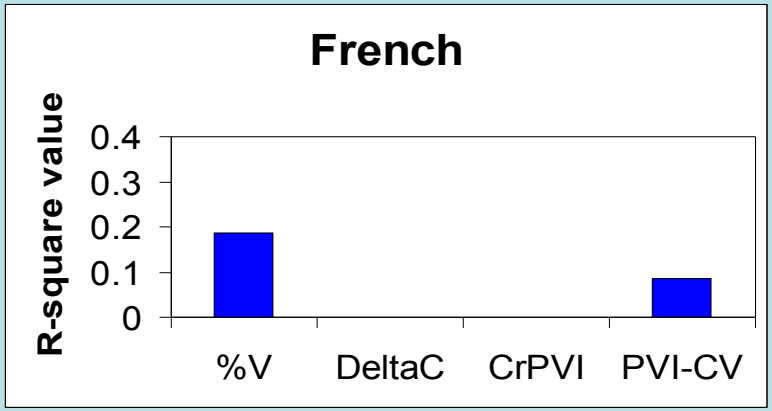
Results

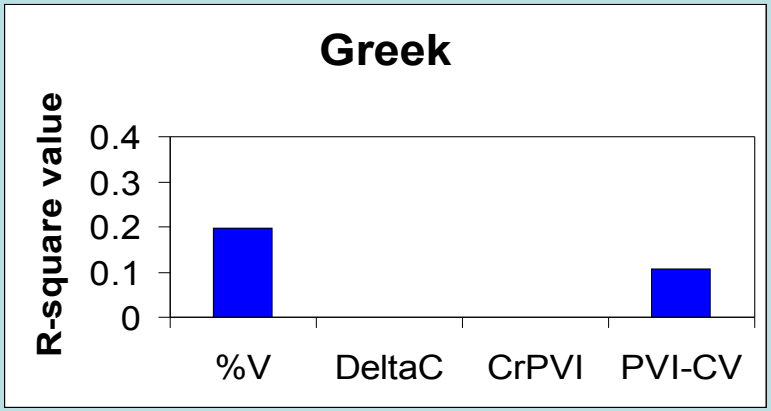
Multiple regression analyses of rhythm measures on all 11 phonological properties







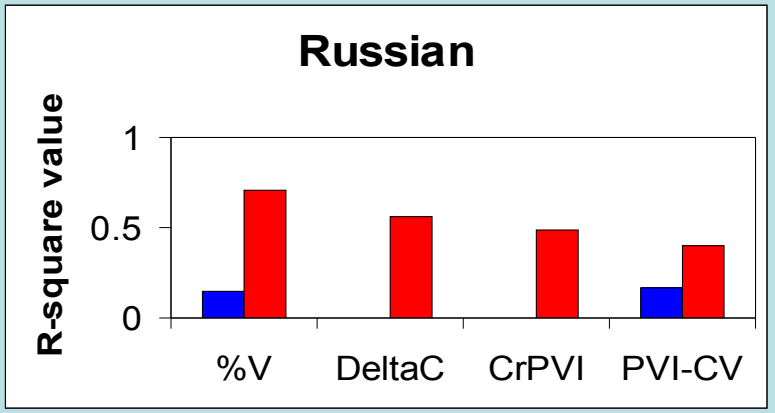


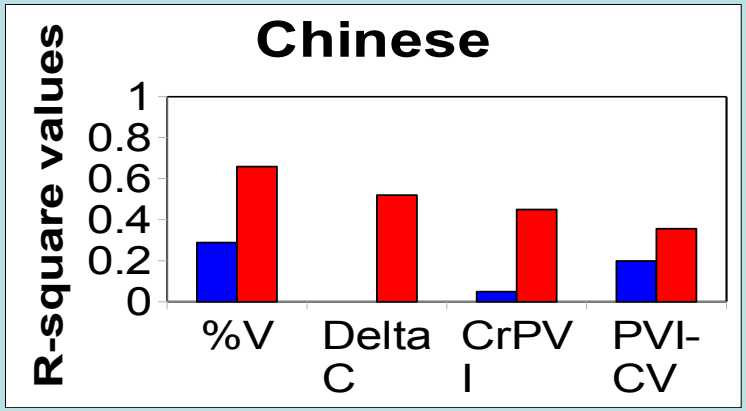


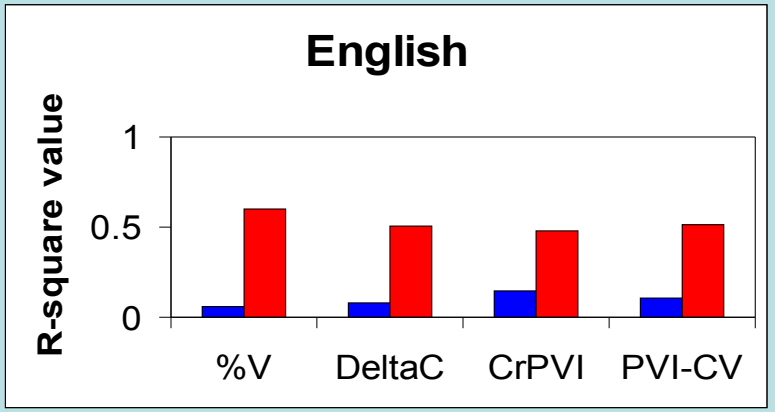
Results

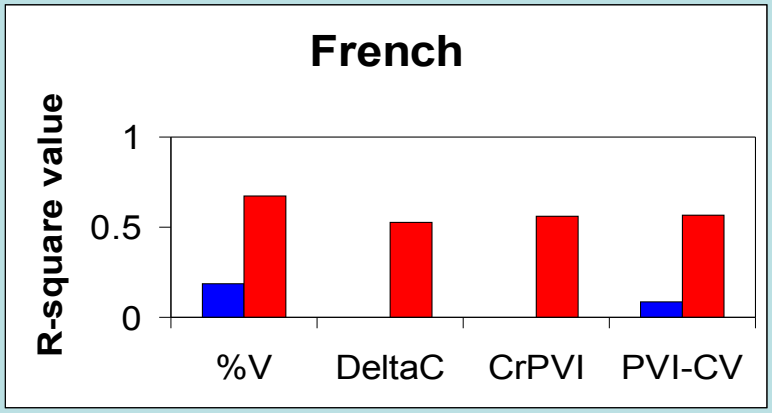
Multiple regression analyses on:

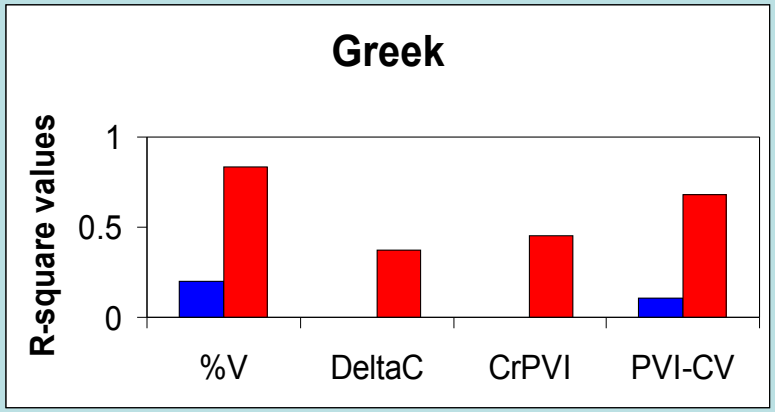
- phonological properties only
- phonological properties & speaker





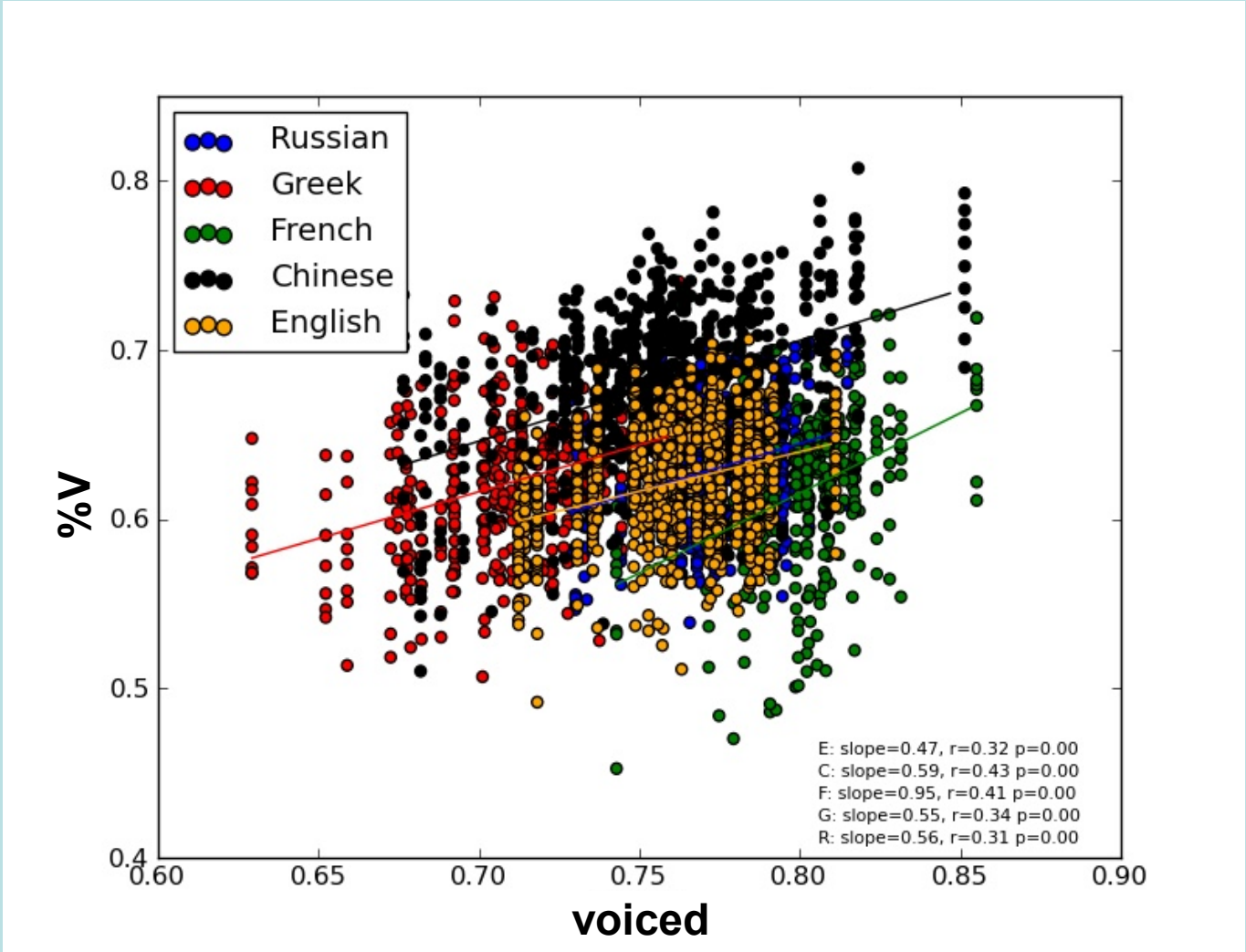






Results

UNIVERSITY OF OXFORD



Conclusions

UNIVERSITY OF OXFORD

How far can rhythm measures (RMs) be predicted from the phonological properties of texts?

Text-dependence: low

Speaker-dependence: high

Acknowledgements

Our project *'Comparing Dialects and Languages using Statistical Measures of Rhythm'* is supported by the ESRC via research grant RES-062-23-1323.

The support for Dr Shih was provided by NSF IIS-0623805 and NSF IIS-0534133.

Thank you!

How far can phonological properties explain rhythm measures?

Elinor Keane[1], Anastassia Loukina[1], Greg Kochanski[1], Burton Rosner[1], Chilin Shih[2]

[1] Oxford University Phonetics Laboratory, UK
[2] EALC/Linguistics, University of Illinois, Urbana-Champaign, USA

Presented at the 2010 Colloquium of the British Association of Academic Phoneticians, 28-31 March 2010, London, UK. This can be downloaded from the Oxford Research Archive, http://ora.ouls.ox.ac.uk , or http://kochanski.org/gpk/papers/2010/BAAP/Keane .

ABSTRACT

Speech rhythm has long been thought to reflect the phonological structure of a language (e.g., Roach 1982; Dauer 1983, 1987). Syllable structure is a key example: languages that allow complex consonant clusters would have a rhythm characterized by much more variability in consonant length than a language like Mandarin where consonant clusters are rare. We explored this experimentally by seeing how well a range of popular rhythm measures were predicted by the phonological properties of the text.

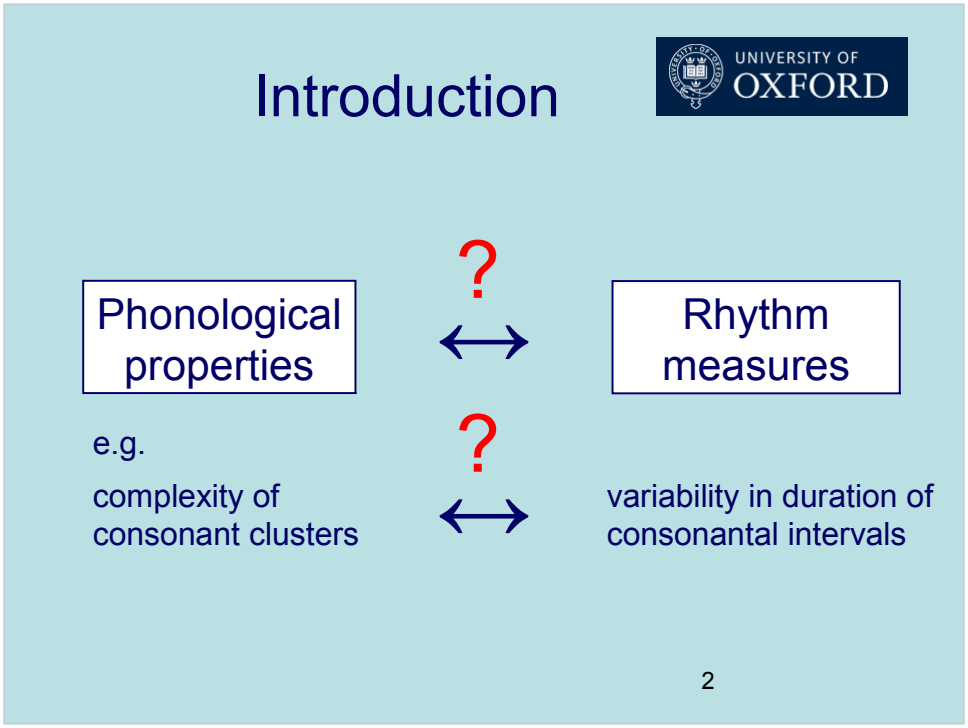
The results are based on 3059 paragraphs read by 62 native speakers of English, Greek, French, Russian and Mandarin. The paragraphs were selected from the novel *Harry Potter and the Chamber of Secrets*, to represent the full range of phonological variation existing in each language. They included pairs of paragraphs chosen for particularly high and particularly low values of eleven different phonological properties. These were calculated from the expected transcription and included the average complexity of consonant clusters, percentage of diphthongs in the text and average sonority (assigning a sonority level of 0 to obstruents, 1 to sonorants and 2 to vowels).

First, we confirmed that languages indeed have different phonotactics, based on the expected transcription. For example, the complexity of consonant clusters in the English data was significantly greater than in the Mandarin data. A classifier based on a pair of averaged phonological properties (e.g. mean consonant cluster length and mean sonority) would correctly identify the language of 70% to 87% of the paragraphs (1Q-3Q range, depending on the pair of properties, chance=20%).

The recorded speech was divided into vowel-like and consonant-like segments using a language-independent automatic segmenter, trained on all five languages. From this, we computed 15 statistical indices proposed as rhythm measures in the literature, e.g. %V, VnPVI (references in Loukina et al. 2009): all were devised to capture durational variability between languages. In contrast to the classifiers based on phonological properties, we found large overlap between languages.

Phonological properties were found to predict paragraph-to-paragraph differences in rhythm measures rather poorly. The largest correlations involved the percentage of voiced segments in speech vs. the percentage of voiced segments in text, but these only explained 9% of the variance in Russian and 18% in Mandarin. Instead, interspeaker differences accounted for much more of the variation in the rhythm measures in a linear regression analysis. For example, for Russian, the average adjusted r^2 across different rhythm measures was .112 for regressions against phonological properties, but .295 for regressions against speakers. The corresponding values for English were .139 and .335.

These results indicate that differences in timing strategies between speakers, even within the same language, are at least twice as important as the average phonological properties of the paragraph. It suggests that rhythm, in the sense of durational variability, is determined more by performance differences between individuals than differences in the phonological structure of languages.



The particular issue that I'm going to address is the relationship between the phonological properties of a language and its rhythm, as calculated by various acoustically-based measures proposed over the last decade or so. The idea that phonological structure and rhythm are related is hardly new. There are, for instance, papers by Dauer and Roach from the 1980s arguing that the rhythmic impression of a language emerges from its phonological characteristics.

Syllable structure is a key example. Take the complexity of consonant clusters allowed by a language. It makes intuitive sense that a language which freely allows complex clusters, like English, should give a different rhythmic impression from one that does not and that's reflected in various different rhythm measures based on variability in the duration of consonantal intervals. So just how closely are the phonological property on the one hand and the specific rhythm measure on the other related?

Introduction

UNIVERSITY OF OXFORD

How far can rhythm measures (RMs) be predicted from the phonological properties of texts?

How closely are phonological properties and RMs correlated?

→ text-dependence of RMs

How much variability in RMs is there between speakers?

→ speaker-dependence of RMs

How much variability in RMs is there between languages?

→ language-dependence of RMs

3

More generally, how far can rhythm measures be predicted from the phonological properties of the texts to which they're applied?

First I'll give you a very brief overview of how we went about tackling this question and what we hoped to find, then the rest of the talk will flesh that out and discuss what we actually found.

In a nutshell, we approached this by designing a large corpus of texts, for each of which a set of phonological properties was computed. We recorded a large number of speakers reading those texts and then applied the rhythm measures to the resulting audio.

We aimed to establish three main results:

•firstly (and this should answer our key question), how closely are the phonological properties and rhythm measures correlated? Put another way, how text-dependent are the RMs? If the correlations are close, the RMs **are** largely predictable from the particular phonological properties of the texts to which they're applied – they're highly text-dependent. If the correlations are **not** close, the RMs are not simply reflecting aspects of phonological structure – text-dependence is low. That may be because of differences in individual speaking style, so…

•secondly, we also looked at variability between speakers. Since the phonological properties were computed from the text and didn't change from speaker to speaker, comparing RMs across speakers should provide a clear measure of their speaker-dependence.

Another reason why correlations between phonological properties and RMs may not be close is that languages differ in their phonetic implementation of phonological properties, so

•thirdly we were also aiming to establish how far the RMs are language-dependent.

It'll be the first issue that I'm focussing on: for more on how well RMs separate languages and speakers, go to Anastassia's poster this afternoon!

So that's what we hoped to achieve. Now for a bit more detail:

Corpus: speakers

UNIVERSITY OF OXFORD

- 62 speakers:
  - 24 British English (Southern England)
  - 10 Russian (Moscow/St.-Petersburg)
  - 10 Taiwanese Mandarin (Taipei)
  - 9 Modern Greek (Athens)
  - 9 French (Paris)
- 20-28 years old
- <4 years outside their home country

4

Our corpus of audio data was specially recorded for the purpose and contains speech from 62 speakers, split between the 5 languages you see there: British English, Russian, Mandarin, Greek and French. The speakers fell within a fairly narrow age range (20-28-years-old) and were recorded in the Phonetics Laboratory in Oxford. All of the non-English speakers had lived outside their home country for less than 4 years.

Corpus: texts

UNIVERSITY OF OXFORD

- 40 short texts for each language:
  - paragraphs from '*Harry Potter and the Chamber of Secrets*'
  - Aesop's fables
  - children's poetry

= 2730 recorded paragraphs

- sentences from '*Harry Potter and the Chamber of Secrets*' for each language

= 22,899 recorded sentences

5

The corpus contains **read** speech: part paragraphs, part sentences. Most of the paragraphs were selected from '*Harry Potter and the Chamber of Secrets'*, which conveniently has translations in all of our languages (not to mention 62 others!). There were also a few Aesop's fables and some children's poetry. I'll not be making any further reference to the poetry now: to hear more about that, do go to Greg's poster this afternoon.

So that gives a total of 2730 paragraphs. The sentences also came from *Harry Potter* and were chosen to be easily readable. There were getting on for 23,000 sentences in total, so we're talking about a corpus of considerable size, certainly compared with the data on which many recent rhythm studies have been based.

Segmentation

UNIVERSITY OF OXFORD

- automatic, using the HTK toolkit
- cross-linguistic
- divided speech into consonant-like, vowel-like and silent intervals

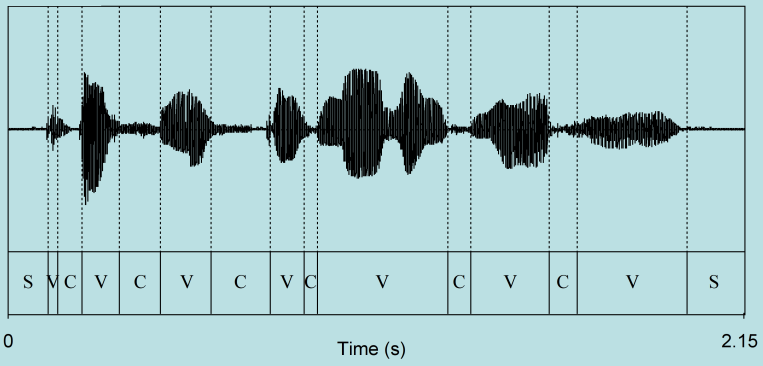
6

Such studies have generally relied on manual segmentation of data into vocalic and consonantal intervals: clearly that wasn't feasible with a corpus this big. So we segmented automatically using the HTK toolkit, identically for all 5 of our languages and divided speech into intervals of 3 types – consonant-like, vowel-like and silence.

Segmentation

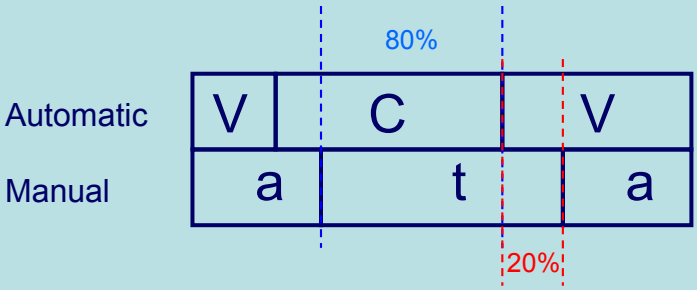
UNIVERSITY OF OXFORD

‘The basilisk was moving toward Harry.’



Here’s an example – it’s one of our English sentences with the automatic segmentation shown underneath.

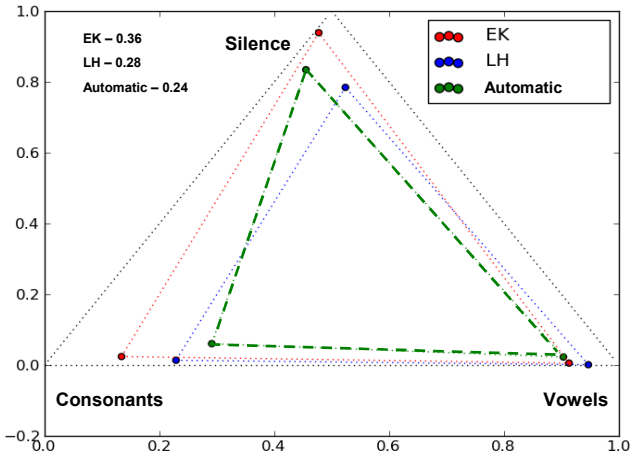
Comparing segmentations



8

For comparison we also produced manual segmentations of a subset of our data, all by trained phoneticians. As you might expect, there are instances where the segmentations diverge slightly. Compare, for example, a VCV stretch like this in a manual segmentation with a hypothetical automatic segmentation. The two overlap to a significant extent, but the segment boundaries don't coincide perfectly. To get a grip on how good the match is, we computed for every manually segmented phone the percentage that is classified as a consonant vs. the percentage classified as a vowel or a silence. So, for our [t], 80% is classified as a C by the automatic segmentation and 20% as a V.

Comparison between segmentations



9

Taking these figures allowed us to produce plots like this one, which shows the match between our automatic segmentation (the green triangle) and three different manual segmentations.

Two of these are shown in red and blue: the other, the black outline, serves as the point of reference, and the distance between its corners and those of the other triangles shows how far they diverge from it. The bottom left corner shows the comparison for consonants, and it's here that you find the biggest difference for the automatic segmentation; the bottom right corner is for vowels (there's not too much difference here) and the top shows silences (notice that here the automatic segmentation actually fares better than one of the manual ones).

Rhythm measures

UNIVERSITY OF OXFORD

- statistical indices based on temporal properties (Ramus et al. 1999; Grabe & Low 2002)
  - acoustically defined
- typically rely on manual segmentation into vocalic and consonantal intervals

10

We then used the results of this segmentation to compute various different RMs. By ‘rhythm measure’ I mean a statistical index, such as those proposed on the one hand by Franck Ramus and colleagues and on the other by Grabe and Low.

Their work has subsequently inspired a number of different variants, and we took a fairly inclusive approach, calculating 15 of those that have been proposed. All are based purely on temporal properties of the speech and are acoustically defined, relying on segmentation into vocalic and consonantal intervals.

Examples

UNIVERSITY OF OXFORD

- global measures, e.g:
  - %V – percentage of vocalic intervals
  - ΔC – standard deviation of consonantal intervals
  - Vdur/Cdur – ratio of vowel to consonant duration
- PVI-like measures, e.g:
  - CrPVI – raw consonantal pairwise variability index
  - PVI-CV – PVI of consonant+vowel group

They fall roughly into two categories, depending on the domain over which variability is calculated:

•firstly global measures. Here are some examples: %V – the percentage of vocalic intervals in speech, ΔC – the standard deviation of consonantal intervals, Vdur/Cdur – the ratio of vowel to consonant durations

•the second category I've called PVI-like measures, PVI being the pairwise variability index developed by Low and colleagues. This is based on differences in duration between successive vocalic (or consonantal) intervals, or even C+V groups, as in the PVI-CV.

Phonological properties

UNIVERSITY OF OXFORD

- computed from roughly phonemic transcriptions of the texts
- cross-linguistically defined
- plausibly have a direct effect on rhythm measures

12

So, if you remember, our aim was to see how far these RMs are predictable from phonological properties. So how did we go about computing the phonological properties?

•they were computed from transcriptions of the texts for each language, for both paragraphs and sentences, and these were roughly phonemic.

•in choosing properties, we needed something that would make sense in all of our languages, and so could be defined cross-linguistically.

•we were also looking for properties that might be expected to affect the rhythm measures directly.

Examples

UNIVERSITY OF OXFORD

- **ccluster** – mean number of consonants between vowels
- **voiced** – fraction of voiced segments in total number of segments
- **sonority** – mean sonority index based on scale:
  [stops, fricatives, affricates] – 1
  [nasals, liquids] – 2
  [glides, vowels] – 3
- **pvi-variants**, e.g. ccluster_pvi – mean-square difference between numbers of consonants on adjacent inter-vowel gaps

13

Here are a few examples from the eleven that we computed.

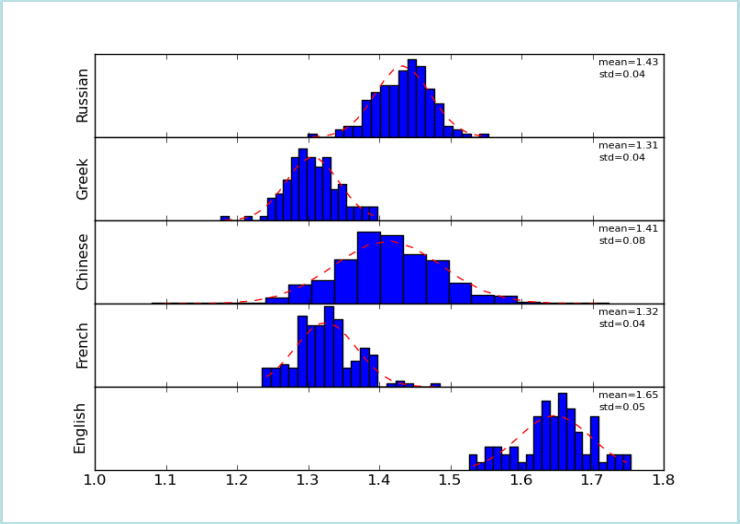
•**ccluster**, as we've called it, is the one I gave at the start. It measures the complexity of consonant clusters by straightforwardly averaging over the number of consonants occurring between vowels, and might well be expected to correlate with the RM ΔC.

•**voiced** looks at the fraction of phonologically voiced vs. unvoiced segments. We'd expect that to bear some relation to %V, the percentage of vocalic intervals in speech.

•**sonority** rates the average sonority of the text by assigning sonority values according to this scale and then calculating the mean. Finally, for several of the properties we computed **pvi-variants**, to see how closely, for instance, applying the pairwise variability index to ccluster would compare with CrPVI.

Comparing phonological properties across our 5 languages, we found evidence of the kinds of phonotactic differences you might expect. Take ccluster, for example:

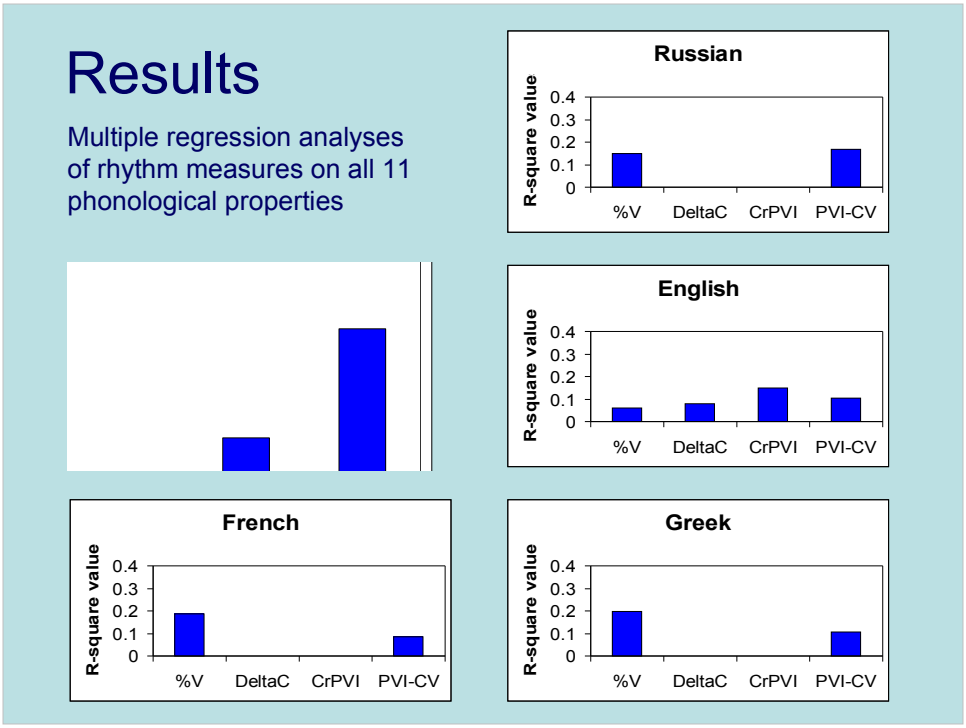
Ccluster for ‘Harry Potter’ paragraphs



14

Here you see histograms for values of ccluster computed for paragraphs in the *Harry Potter* book in each of our 5 languages. The high complexity of consonant clusters in English is immediately obvious, so that’s reassuring confirmation that we were capturing expected phonotactic differences.

We used this information on the distribution of values to make sure our paragraphs represented the full range of phonological variation in each language. For each property we included pairs of paragraphs that fell at either end of the distribution. So here, for instance, we selected one of the top five in each language to represent a maximum value for ccluster and one of the bottom five to be a minimum value.

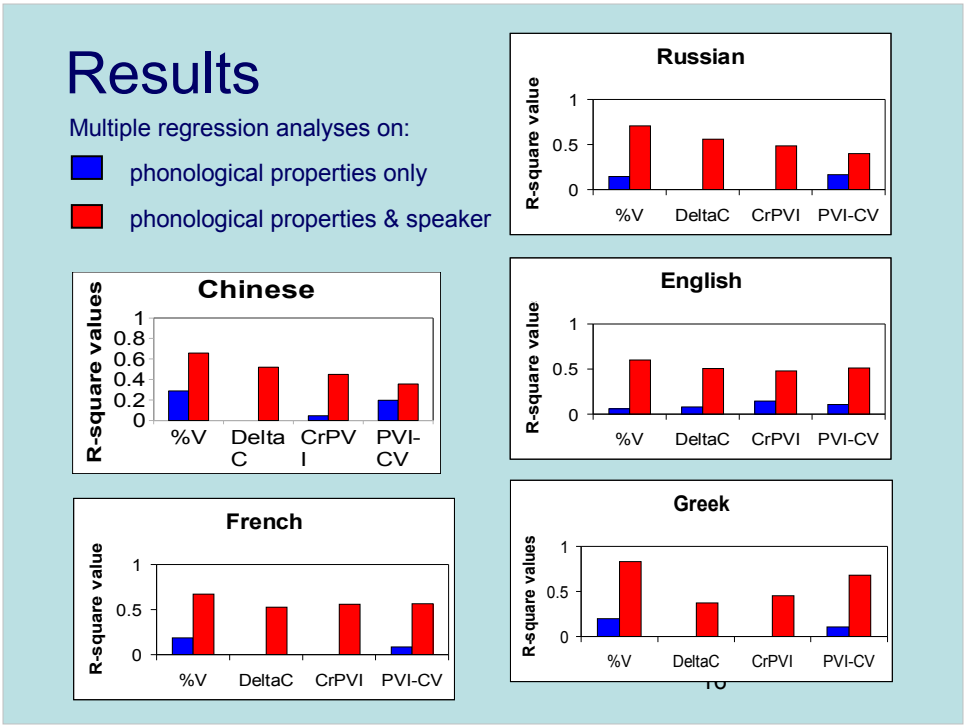


If you recall, we were keen to see how far RMs can be predicted from the phonological properties of texts, so we calculated multiple regressions for each individual RM on all of our eleven phonological properties. What you see here are the results for four of the RMs (shown on the x-axis) in each of our languages. The height of the bars show the R-square values for each, i.e. how far the phonological properties all together allow us to predict the value of the RM.

You'll notice firstly that none of the bars is very high. %V comes out best overall but even in Chinese, its highest value, it still doesn't quite reach 0.3.

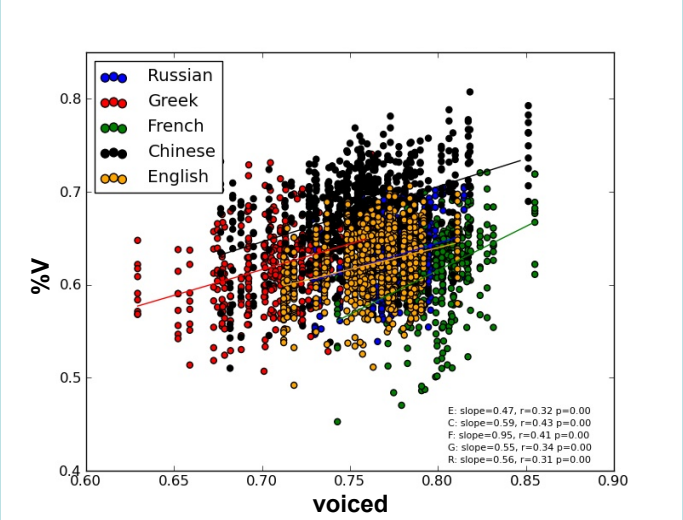
Secondly, there are a fair few gaps in the charts, and in such cases **none** of our phonological properties made a significant contribution to predicting that RM. Notice that ΔC fares particularly badly – so not even Ccluster, which we thought might be quite closely correlated with ΔC, turns out to be a significant predictive factor.

Compare that with what happens when speaker identity is taken into consideration.



Here the blue bars are exactly as on the previous slide, showing what you can predict from the phonological properties; the red bars shows what happens when you consider both phonological properties **and** speaker identity. Clearly differences between speakers are accounting for much more of the variation in the RMs.

Results



17

This shows you a bit more detail for the phonological property that best predicted a rhythm measure – the correlation between the phonological property 'voiced' and the RM %V. The correlations are highly significant in each language, but their strength is pretty low.

You can see that there's a fair bit of overlap between languages, and certainly lots of variation within each language, given how widely dispersed the dots are.

Conclusions

UNIVERSITY OF OXFORD

How far can rhythm measures (RMs) be predicted from the phonological properties of texts?

Text-dependence: low

Speaker-dependence: high

18

Returning, then, to our original question – how far can RMs be predicted from the phonological properties of texts? Not very well seems to be the basic answer – text-dependence is low and that seems to be largely because it's massively overshadowed by differences between speakers – speaker-dependence is high. That obviously has implications for the use of these RMs in separating languages – for more on that see Anastassia's poster.

Acknowledgements

Our project *'Comparing Dialects and Languages using Statistical Measures of Rhythm'* is supported by the ESRC via research grant RES-062-23-1323.

The support for Dr Shih was provided by NSF IIS-0623805 and NSF IIS-0534133.

Thank you!