This is the final report for
Comparing dialects using statistical measures of rhythm.
More information can be found at
http://kochanski.org/gpk/papers/2008/aesopus2/ .

Greg Kochanski and Anastassia Loukina (the project also involved extensive cooperation from Elinor Keane, Chilin Shih, and Burton Rosner.)

1. NON-TECHNICAL SUMMARY

Please provide below a project summary written in non-technical language. The summary may be used by ESRC to publicise your work and should explain the aims and findings of the project. *[Max 250 words]*

Does each language have its own performance style, or are they all the same, except that the words are changed? One gets the strong feeling that they are not all the same, and Linguists have talked about a different style or rhythm for each language, in analogy with musical rhythms. Our aim in this project was to find quantitative metrics that would capture this impression.

Our working assumption was that the rhythmic differences between languages are large enough so speakers of one (e.g. English) rarely produce the rhythms of another language (e.g. French). So, we collected data from five languages to test this idea. First, we investigated all 15 of the techniques that other researchers have published; they were based on the duration of speech sounds, in analogy with musical rhythm.

Somewhat surprisingly, we found that there people from one language often speak with the duration patterns of another: for instance, French people whose patterns of sound duration were typical of Greek. This will lead to a reassessment of some research that assumed that each language had a distinctive rhythm. Rather, our data suggests that each person has their own rhythmic style, and the language they speak influences that style. We also showed that the old idea of dividing languages into two clear rhythmic classes is too simple: there are more than two kinds of rhythm.

2. PROJECT OVERVIEW

a) Objectives

Please state the aims and objectives of your project as outlined in your proposal to the ESRC. *[Max 200 words]*

The main aim of this project was to evaluate current rhythm metrics on a large corpus of data to identify improved methods to make quantitative analysis of rhythmic differences between languages and dialects. We focused on three research questions:

1. To understand how rhythm measures (RMs) depend on changes of the text within a language (§3 of the Case for support).
2. To evaluate existing RMs and identify the combinations of RMs that produce the most accurate identification of a speaker's language (§4).
3. To measure the differences between dialects of English and compare them to variation among languages (§5).

To answer these questions, objectives were to:

- Collect a large corpus of texts from speakers of five languages,
- Use novel methods such as automatic segmentation algorithms and classifiers to compare the variation in rhythm measures,
- Check the accuracy of the automatic segmentation against manual labels (§6) and check the inter-labeler consistency and its effect on the rhythm measures. (Two subsidiary experiments.)
- Release the transcriptions, audio data, and relevant software under an open-source license to enable other researchers to replicate our results or use our techniques.

b) Project Changes

Please describe any changes made to the original aims and objectives, and confirm that these were agreed with the ESRC. Please also detail any changes to the grant holder's institutional affiliation, project staffing or funding. *[Max 200 words]*

Dr. Keane (co-Investigator) joined the project 7 months later than the plan, and has taken maternity leave from May 2010. As a result, we have been granted a one-month no-cost extension to the project.

c) Methodology

Please describe the methodology that you employed in the project. Please also note any ethical issues that arose during the course of the work, the effects of this and any action taken. *[Max. 500 words]*

1. Data collection.

This project involved the analysis of corresponding read texts in five different languages. We identified 11 phonological properties that could influence the perception of rhythm and defined these cross-linguistically. We then selected in each of our languages 31 paragraphs with unusually high or low values of these properties, 4 Aesop's fables, a version of 'Cinderella' matched for length across languages and short poems matched for meter across languages. We also selected 2000 short sentences to train our automatic alignment systems. We have obtained the necessary ethics approval and during spring-summer 2009, recorded 63 speakers (148 hours of recordings).

2. Segmentation
We developed several segmentation algorithms of various complexity to segment the data into vocalic and intervocalic algorithms. Some of these algorithms used the HTK toolkit. We used several criteria to compare the outcome of these algorithms.

3. Analysis

For RQ1 ('Text dependence') we ran a series of linear regressions to compare the effect of phonological properties on the rhythm measures. Beyond what we originally planned, we studied the text-dependence in all five languages, not just Southern British English and Russian as initially proposed.

For RQ2 ('Evaluation of the rhythm measures and optimized rhythm measure), we conducted an exhaustive set of analyses, comparing combinations of automatic segmentation algorithms and 15 published rhythm measures. Our approach has been to measure how reliably a machine learning system (a classifier) can correctly predict the language, based on one or more rhythm measures. We built a classifier for each combination of rhythm measures and different pairs of languages, (more than 50,000 different classifiers in all) giving us a very strong set of results.
In our search for optimal measures of rhythm, we also investigated prosodic properties beyond
duration and searched for rhythmic predictability over a range of about 4 syllables. We measured how predictable the acoustic properties of speech are, based on the previous few segments.
To do this, we used a linear regression of the acoustic properties of segments against the acoustic properties of preceding segments. We computed ten acoustic properties for each segment based on duration, loudness, frication and speech rate. We then computed many linear regressions to measure how predictable these properties are, over various ranges.

For RQ3 ('Dialects of English'), we applied our automatic segmentation and machine classification algorithms to the IViE corpus (Grabe 2001). The corpus used in this study included speakers from Belfast, Cambridge, Newcastle, Leeds, Dublin, London residents of Jamaican decent (monolingual English speakers) and bilingual Bradford residents of Punjabi decent (speaking English). We used the reading and re-telling of the "Cinderella" story (12 speakers for each dialect). This was automatically segmented into vocalic and consonantal intervals; we then computed 15 published rhythm measures and used classifier techniques to evaluate variability within and differences between dialects.

Subsidiary experiments: We have also obtained manual segmentation from 9 trained phoneticians and used measures of inter-rater variability to compare manual and automatic segmentation and evaluate consistency between phoneticians.

d) Project Findings

Please summarise the findings of the project, referring where appropriate to outputs recorded on *ESRC Society Today.* Any future research plans should also be identified. *[Max 500 words]*

RQ1 "Text dependence":

We found that languages have different phonotactics based on canonical transcriptions, but phonological properties do a surprisingly poor job of predicting differences in rhythm measures. Surprisingly, we found that the bulk of the differences were speaker-to-speaker differences. These results were reported in conference papers by Keane et al (2010) and Loukina et al (2010a); a journal paper is in progress.

RQ2a "Evaluation of rhythm measures":

We found that the language-to-language differences in rhythm are no larger than the person-to-person differences. For example, one can find French speakers who use Greek rhythms, and similarly for most of our language pairs. Together with the last, this suggests that phonology may overstate the differences between languages, possibly because each language is represented in a different way; this may affect the interpretation of many cross-linguistic studies in linguistics.

While most pairs of our languages could be separated fairly well with a classifier based on just one RM, each different pair typically needed a different RM. Combinations of three RMs were needed to maximise the identification rate for all five languages at once. In addition to these statistical results, our data show that the languages group differently, depending upon which RMs are used to classify them. Thus, languages differ in several different ways and there are no absolute rhythm classes. These findings (supported by a multidimensional scaling analysis) show that linguistic rhythm is multidimensional. These results were reported in a conference paper Loukina et al. (2009) and submitted to JASA as a journal paper. We plan to propose a perceptual study to validate these results; they substantially improve our understanding of rhythmic typology and methodology.

RQ2b "Improved measures":
Going beyond existing rhythm measures that are (primarily) focussed on segment duration, we developed a novel technique, defining the strength of rhythm as the predictability of acoustic properties. Testing this, we confirmed that it allowed us to effectively separate poetry from prose in all five languages. Notably, patterns of segment duration are among the least predictable of the properties we investigated. These results were reported in conference papers Kochanski et al (2010a) and Kochanski et al (2010b). This work may lead to an entirely new class of rhythm measures.

RQ3 "Rhythm of English dialects":

We found that English dialects have overlapping rhythms. In agreement with findings from the main corpus, we found that different groupings of dialects could be observed depending on the choice of rhythm measure(s). Based on RQ2b, we also developed new loudness-based measures. These separated dialects better than chance and revealed yet more different groupings. These results have been reported in a conference paper Loukina and Kochanski (2010).

SE1-2: Our work has highlighted the subjectivity of manual labelling. We have shown that acoustic properties of segments do not always match their expected phonological or even phonetic category. These differences are language-specific and suggest that rhythm measures based on manual labelling are strongly affected by the conventional phonological interpretation of sounds in that language.

These results are in the journal paper.

e) Contributions to wider ESRC initiatives (eg Research Programmes or Networks)

If your project was part of a wider ESRC initiative, please describe your contributions to the initiative's objectives and activities and note any effect on your project resulting from participation. *[Max. 200 words]*

Not applicable

3. EARLY AND ANTICIPATED IMPACTS

a) Summary of Impacts to date

Please summarise any impacts of the project to date, referring where appropriate to associated outputs recorded on *ESRC Society Today*. This should include both scientific impacts (relevant to the academic community) and economic and societal impacts (relevant to broader society). The impact can be relevant to any organisation, community or individual. *[Max. 400 words]*

Results

The project has important methodological consequences for the field of rhythm metrics:

- The corpora that were used in many previous studies do not appear to be large enough for reliable results.
- Conclusions about the similarity of two varieties based solely on rhythm measures seem to depend on which measures are used.
- Contrary to previous beliefs, there is only a weak dependence of rhythm measures on phonological properties.
- We have suggested new methods for measuring rhythm, which include acoustic properties other than duration.

We have presented these results in 4 conference papers, 2 posters and 1 journal paper. These outputs have been uploaded to the ESRC website, Oxford University Research Archive and Project website to ensure wider dissemination. By 24 November 2010 these had already collected five citations.

Methodology

We have developed new methods and tools. These include bespoke data collection software and classifier techniques to compare variability in values of rhythm measures. The sofware has been released open-source to enable further research.

Data

We have collected and released our main data set. It is called "The Oxford Aesop

Corpus", and consists of 3241 paragraph-sized texts which include read paragraphs, short poems and spontaneous speech and 30 583 short sentences suitable for training an ASR system. The documentation, audio files and transcriptions are all freely available from the project website. We have already used this data for other projects in the Phonetics lab.

General public

We have produced a project webpage: www.phon.ox.ac.uk/speech_rhythm .

The page reports the main findings of the project and contains links to all publications associated with the project. We have also produced a special page on best practice in using the rhythm measures aimed for students who would like to use the methodology in their research. These pages have been viewed 257 times to date.

Kochanski has also written 32 non-technical essays in his blog "Science and language". These essays are clustered around the project's work; the blog gets over 25,000 non-robot accesses per year.

Project staff

The project has contributed to further professional development of the project staff.

Loukina has greatly expanded her IT skills and learnt scripting in Python and obtained an experience of working in a distributed team. Kochanski has developed an extensive knowledge of the design of speech alignment systems and the use of the HTK toolkit. We expect this will be valuable for future work.

b) Anticipated/Potential Future Impacts

Please outline any anticipated or potential impacts (scientific or economic and societal) that you believe your project might have in future. *[Max. 200 words]*

By making use of the world's investment in the Internet and search engines, the web lets people efficiently find our results. Consequently, anyone with internet access can generate impact, possibly in ways we cannot imagine.

The societal and economic impacts we *can* imagine are related to language learning. Rhythm is an important aspect of speech, and in modern Europe, language learning is very important for the UK. Economically, we benefit from people who learn English, and we need to learn other languages to live and work elsewhere. Our work should lead to improvements in second language learning via computerized language learning systems (though we do not expect any immediate, direct impact). These systems will eventually improve mobility across Europe.

More immediately, the corpus we released will be a useful resource for linguists, language teachers and speech technologists. Our other corpora are downloaded regularly, and we expect this one will also be used.

We have also begun a project with the University of Provence (France) to apply methods from this project to evaluate of prosody in the speech of second language learners (British learners of French and French learners of English). This Provence project is a step towards practical applications.

You will be asked to complete an ESRC Impact Report 12 months after the end date of your award. The Impact Report will ask for details of any impacts that have arisen since the completion of the End of Award Report.