Long Range Prosody Prediction and Rhythm

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

University of Oxford

*¤EALC/Linguistics, University of Illinois,
Urbana-Champaign, USA*

This talk is about long range prosody prediction in rhythm. Work was done mostly at Oxford by myself and Anastasia Loukina, Elinor Keane, Chilin Shih from the University of Illinois, and Burton Rosner. Another title for the talk could be 'Doing Rhythm with Fewer Assumptions'.

Long Range Prosody Prediction and Rhythm
(sometimes known as)
Doing Rhythm with Fewer Assumptions

Greg Kochanski, Anastassia Loukina, Elinor Keane, Chilin Shih¤ and Burton Rosner
(and thanks to the ICSLP 2009 Rhythm Special Session for interesting questions)

University of Oxford
¤EALC/Linguistics, University of Illinois, Urbana-Champaign, USA

Detecting Rhythm

Let's build a system for measuring rhythm.

Q1. What do we train it on?

--or--

Q2. What is rhythm?

Now, if we're talking about the detecting or measuring rhythm, we're talking about building a system, piece of software that does the job. And that gives us two choices. If we're building a data driven system, we can talk about what do we train the system on; or if we're directly writing code that measures rhythm, we have to ask "What is rhythm?"

Detecting Rhythm

Let's build a system for measuring rhythm.

Q1. What do we train it on?

- Low-pass filtered speech

- Duration

Well, what can we train it on? Most people train their systems on low pass filtered speech.

Q1. What do we train it on?

- Low-pass filtered speech

  - Was proposed to simulate the acoustic environment of the human fetus

  - ...except that the low-pass filters commonly used eliminate high frequencies much more effectively than actual sheep

An ewe as a low-pass filter

Sound environment of the fetal sheep.

Gerhardt KJ, Abrams RM, Oliver CC.

Department of Communication Processes and Disorders, University of Florida, Gainesville 32611.

The internal sound pressure levels within the intact amnion of pregnant ewes surgically implanted with a hydrophone was determined …. Measurements were made of sound pressures outside and inside the ewe, and sound attenuation through maternal tissues and fluids was calculated. Sound pressures generated by low frequencies (less than 0.25 kHz) were 2 to 5 dB greater inside than outside the ewe. Above 0.25 kHz, sound attenuation increased at a rate of 6 dB per octave. For 4.0 kHz, sound attenuation averaged 20 dB. … Thus the fetus is developing in an environment that is rich with internal and external sounds.

Now, low pass filtered speech was originally proposed to simulate the accoustic environment that the human fetus hears, except that someone had actually done the experiment with fetal sheep. They put a microphone in a pregnant ewe and measured what the fetus could hear. Then the crucial bit is here in red. Above 250 Hertz, the sound attenuation goes as 6 decibels per octave. So it's a one-pole filter: it's not a steep filter. It's one that won't cut off the high frequencies too strongly. And, one can hear and understand speech very well through a one-pole filter like that.

That means the fetus is developing in an environment where it can hear essentially all of the distinctions it could hear outside the womb. But what we typically train these systems on, is low pass filtered speech with a fourth order or steeper cut off. Such a filter is designed to delexicalise the speech to make it impossible to determine the segmental information. So, normally we train these systems on an impoverished environment compared to what the fetus gets.

Detecting Rhythm

Let's build a system for measuring rhythm.

Q1. What do we train it on?

Low-pass filtered speech:

Typically 300-400 Hz cutoff

4th order or steeper

Listen.

Let me give you an example of what this sounds like. Here's the 'Stars and Stripes Forever' performed by the US Marine Corps Band.
[plays 01:59 to 02:08]
This is your basic march. It's a dull thumpy thing or so it seems in the low pass filter.
[plays 02:15 to 02:27]
And couldn't you feel the sheer emotional impact of one of the best known American marches? Perhaps some of you with good imaginations and sharp ears could.
[plays 02:35 to 02:46]
As we can hear, there's a certain amount of rhythm and emotional impact in the upper registers too. You miss that with a low pass filter.
[plays 02:53 to 03:03]
That was the thumpy section we heard at first and you can hear that the beats are not uniform, that they are an alternation between two clusters of instruments so you get a different metrical pattern by looking at the whole signal than you would by looking at the low frequency part. So rhythm, as defined by a low pass filter, is a much simpler, less interesting thing than rhythm is in reality. And yet we build our systems that way.

Detecting Rhythm

Q1. What do we train it on?

- Duration differences between languages.
  - **Logic**: phonology and phonotactics are basic properties of each language,
  - and the language specifies stress and prominence,
  - and all these affect the duration of sounds,
  - **Therefore**, rhythm should be a property of a language
  - **Except** that each step above is an approximation
  - So it doesn't work all that well

Well, the other kind of thing that people train their systems on are duration measurements, and they attempt to make them descriminate between languages. The logic of this is pretty straightforward. Phonology and phonotactics are basic properties of each language, and the language also specifies stress and prominence. And all of these phonological properties affect the duration of sounds. Therefore, rhythm should be a property of language. Except that each step above is an approximation. There are also other things that affect rhythm. So overall, it doesn't work all that well.

Detecting Rhythm

- A system for measuring rhythm, trained to discriminate between 5 languages (Loukina et al Interspeech 2009).

- 62 subjects, 2253 paragraphs.

- Automatic segmentation.

- More than 1000 2-way and 3-way combinations of published rhythm measures tried.

Let me talk about some results we got this summer in Loukina et al. presented at InterSpeech 2009. It's a large study of rhythm measures. We used 62 subjects and 22,053 paragraphs. We did automatic segmentation because there was too much data to do manually, and we tried more than a thousand two-way and three-way combinations of published rhythm measures.

Text-driven and individual variation is large
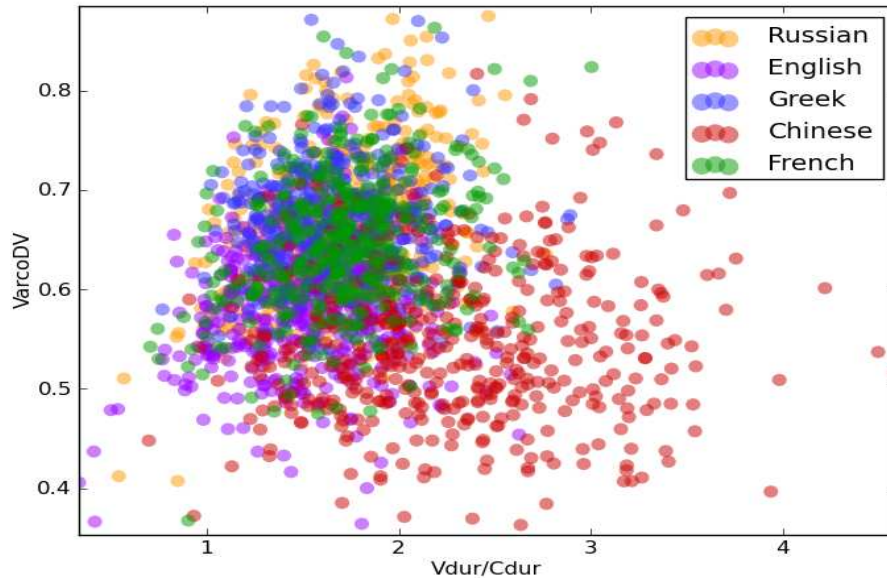
The figure here shows you pretty much what's going on. And what's going on is that there's a lot of variation from paragraph to paragraph and from speaker to speaker. The red dots, for instance, are Chinese and the blue dots are Greek. And what you see is that by and large you can separate Chinese from most other languages. But if you look carefully you'll find some Chinese paragraphs are right next to Greek paragraphs. In other words, a single paragraph's worth of data is not sufficeint to tell the difference from one language to the next using publish rhythm measures.

That's because each person speaks in their own way, so that you have some speakers of Russian who sound very English, some speakers in Greek who sound very French. And also because each paragraph has it's own properties. Some English paragraphs will have rhythms close to Russian, and some English paragraph will have a rhythm close to Chinese. Put those together and you will see that an individual paragraph is not sufficient to tell the difference between languages. And yet many papers use data from just a single paragraph, a single speaker, to draw conclusions about languages.

End of Introduction

- Overall, the field is not entirely satisfactory.
- A lot of original assumptions have failed the test
    - ...but they are still there as part of our techniques.
    - ...and often form the core of our introductions.

- So we decided to start from scratch.

So, the other way we can build a system is to programme a definition of rhythm into our algorithm. But where do we get a definition of rhythm? Expert opinion provides some, but we don't really want expert opinions. For one thing, experts read papers about what the answers should be, and experts defer to each other, so their opinions are not independent. Also, of course, experts don't provide answers for all languages before the data appears. For example, to my knowledge, no one has discussed the rhythm of Croatian and no one really has an opinion on whether it's syllable-timed, stress-timed or mora-timed.

Well, we could ask naïve native speakers. They have many advantages. But how do you ask them? How do you tell them what rhythm is so you can ask them what rhythm is? And can they do the task? It's not an easy task to ignore certain parts of speech and to pay attention to others.

Overall, the field is not entirely satisfactory. A lot of the original assumptions have failed the test, but they are still there as part of our techniques, and they often form the core of our introductions.

Starting from Scratch

- Many cultures have a rhythmic form of poetry
  - Nursery rhymes, for instance.
- No need to cite experts
  - Don't *want* to get expert opinion:
    - Experts read papers about what the answer *should* be
    - Experts defer to each other so opinions are not independent.
  - We *want* naïve native speaker opinion:
    - most will agree poetry is rhythmic

Because of this, we decided to start from scratch with as few assumptions as possible. Starting from scratch, we came upon poetry. Many cultures, perhaps even most cultures, have some rhythmic form of poetry. Nursery rhymes, for instance. And when they do have nursery rhymes, for instance, everyone agrees that they're rhythmic, everyone agrees that they are poetry and everyone agrees that they have a metre. Children like predictable poetry. So we will use poetry.

Rhythm is predictability

- "Be kind to your web-footed friends
- For a duck may be some body's uncle
- Be kind to your friends in the swamp
- Where the weather is cold and damp
- Now you may think this is the end
- Well it is."

Here's an example of something that has a meter and is predictable. And I'm going to use this to show how important predictability is to our brain. This is, again, from the 'Stars and Stipes Forever', otherwise known as, 'Be kind to your web footed friends'. It's a comedy routine from the 1920's:

[plays 07:44 to 08:06]

Lyrics: Be kind to your web footed friends
            For a duck may be somebody's uncle.
            Be kind to your friends in the swamp
            Where the weather is cold and damp.
            Now, you may think this is the end.
            Well, it is.

That's quite a jolt there. And it just stops. Your brain is in the middle of a bunch of rhythm and metre, and its keeping track of what's going on, and it's heard it all before and it's expecting to hear it all again, and then suddenly it doesn't. So it's a rhythmically predictable passage. You can figure out what's coming next and you're right, all the way up to the end, when the song violates all the predictions, and suddenly ceases to be rhythmical.

Rhythm is predictability

- "Web footed friends" is rhythmically predictable
  - Stress patterns
  - Phrases
  - Meter
- The brain knows exactly what's coming next, right up to the end, when the song violates all the predictions.

Starting from Scratch

- Find rhythmic poetry.
- Get native speakers to read it
- The resulting audio is typically rhythmic.
- Do the same with ordinary text (prose).
- Since/if rhythm = predictability,
  - We should be able to a better job of predicting what comes next in the poetry.

So, our general strategy from starting from scratch will be to find rhythmic poetry, and get native speakers to read it. The resulting audio is typically rhythmic. Them we do the same thing with ordinary prose, as a control set. And, if rhythm is predictability, we should be able to do a better job of predicting what comes next in the poetry.

Four poems, each of 8-12 lines

- Southern British English
  - Iambic or Trochaic Tetrameter
  - 24 speakers
- Parisian French
  - 8 syllables per line
  - 9 speakers, 10 now
- Standard Modern Greek
  - Iambic or Trochaic Tetrameter
  - 9 speakers, 10 now
- Standard Russian (Moscow & St. Petersburg)
  - Iambic or Trochaic Tetrameter
  - 10 speakers
- Tiawanese Mandarin
  - Nursery Rhymes in trochees & 1 modern poem in iambic meter
  - 10 speakers

For our experiment, we used four poems each of eight to twelve lines in southern British English, Parisian French, standard Modern Greek, standard Russian, and Taiwanese Mandarin. Speakers weren't really told how to read them; they weren't told to read them with a strong rhythm, but they typically did anyway. The prose paragraphs (the control set) are translations of Harry Potter, chosen for another experiment. Each is about a minute long, 2328 utterances in total. They are recorded in the same sessions as the poetry; the same language and the same speakers.

Prose Paragraphs – the control set

- Translations of Harry Potter
- Chosen for another experiment
- About 60s long
  - Except Mandarin paragraphs: about 30s long.
- 2328 total utterances
- Same speakers
- Recorded in same session as poetry
- Same five languages

Data Preparation

- Re-reads allowed
  - Speaker: at will
  - Experimenter: for major stumbles
  - Experimenter: for major divergences from the text
- 20% of paragraphs were re-read
- Word-level transcription manually fixed
- Phone level transcription
  - Initially automated
  - Manual repairs as needed

Data Preparation: Re-reads were allowed until people were happy with the paragraphs. 20% of the paragraphs were repeated. Word level transcriptions were manually repaired and phone level transcriptions were derived from an automated system and then manually repaired as needed.

Data Analysis

- Segment speech into C, V, Silence regions
- Compute acoustic properties of each segment
- Build predictors
- Evaluate each predictor

In the data analysis we segmented speech into consonantal, silence and vocalic regions. Then, we compute accoustic properties for each segment. We build a linear regression model to predict each property, and then we evaluate each of these predictors.

Segmentation

- Strictly language-independent design
- Automatically segmented
  - Using custom recognizer
  - Trained on our subjects
    - 36 manually segmented minutes of speech
  - Recognizes three phones: C, V, Silence
  - Standard MFCC front-end

The segmentation is a strictly language independent design. In this experiment, we want the segmentation to be language independent because we're trying to compare languages, and we want to be able to separate differences in the segmentation from differences in the languages themselves. We used an automatic segmenter because of the volume of data. It's a speech recognitions system, effectively. It is trained on our subjects, on 36 manual segmented minutes of speech, and has a standard MFCC front end. And it is designed to recognize only three phones: **C**onsonantal, **V**ocalic and **S**ilence. Those three phones are common to all languages, at least all languages that we're studying.

Compute Acoustic Properties

- Ten acoustic properties are computed for each segment
  (descriptions are simplified here)

  - log(Duration)

  - Loudness – related properties:
    - loudness * Duration
    - Is most of the loudness early or late in the segment?
    - Standard deviation of the loudness across the segment.

  - Partial voicing: frication noise vs. harmonics
    - Averaged over the whole segment
    - Near the center of the segment
    - Standard deviation of this across the segment

  - How close is the spectrum to the average speech sound?

  - Two properties based on the rate of spectral change:
    - Mean interval between substantial spectral changes
    - Log(interval between major spectral changes / duration)

Then, in each segment, we compute ten accoustic properties. One is the duration of a segment; three of them are loudness related properties, including how loud is it, is the loudness early or late; standard deviation of the loudness across the segment. We do the same with a partial voicing property, which is essentially, how much frication noise is there amongst the voicing. For another property, we ask how close the spectrum is to the average speech sound. The average speech sound for English is somewhat schwa-like. So you can think of this, for vowels, at any rate, as the vowel quality. And then there are two properties based on rate of spectral change. Spectral change means if you look at the spectrum now and a little bit later, has there been a major change in the spectrum in between? These are sort of related to speech rate.

Build Predictors

- Pick what to predict:
  - Will we predict for a C, V, or S segment?
  - Is the target segment phrase initial, medial, or final?
  - Which of those ten properties shall we predict?
  - (there are 70 sensible combinations of the above)
- How much historical context to use in the prediction?

So, we build predictors then we pick where we want to predict. For instance, are we predicting for a consonantal, a vocalic or a silent segment? Is the target segment phrase-initial, phrase-medial or -final? And which of these ten properties should we predict, and then, how much historical context should we allow for the prediction? And then, of course, which language? And is it prose or poetry? Then we make all these choices every possible way, and we end up with a lot of predictors.

How much context?

In winter I get up at night
And dress by yellow candle-light.
In summer quite the other way,
I have to go to bed by day.
I have to go to bed and see
The birds still hopping on the tree,
Or hear the grown-up people's feet
Still going past me in the street.
And does it not seem hard to you,
When all the sky is clear and blue,
And I should like so much to play,
To have to go to bed by day?

Now, how much context do we include? This is poetry, and poetry can have a complicated rhyming scheme. Here's an example that we used for English:

In winter I get up at night
And dress by yellow candle light.
In summer, quite the other way,
I have to go to bed by day.
I have to go to bed and see
The bird still hopping in the tree.

If you have the preceeding line, it will tell you a lot about the line you're looking at.

I have to go to bed and see

Tells you a lot about the bird still hopping on the tree. So that means we'd like to include the entire last line as our context: that'd be eight syllables. In our data, eight syllables is about 16 segments. Because we define **C**onsonantal to be an entire string of consonants, almost all syllables from our point of view here are **CV**. So if we had 16 segments that will let you predict directly from the previous line. Even with less than that, you can get a lot of information. Two syllables of context (about four phones or four segments) will make the stress determination much more reliable. You have seen a stressed syllable and an unstressed syllable and you can predict what comes next. Even one syllable (that is about two phones) will tell you something. So we'd like a lot of context. But for practical reasons we stop eight segments.

What is a predictor?

- Here, it is a linear equation
  - that predicts the desired acoustic property
  - in terms of all 10 acoustic properties
  - for the preceding N syllables

$$D_t = c_0 + c_1 \cdot D_{t-1} + c_2 \cdot L_{t-1} + \ldots + c_{11} \cdot D_{t-2} + c_{12} \cdot L_{t-2} + \ldots$$

What are the predictors? Well, here it's a linear equation that predicts the desired accoustic property in terms of all ten of the accoustic properties of the preceeding syllables. So, if we're predicting, say, the duration at time, *t*, the predictor starts with a constant term. And that constant actually depends on the subject. And we're using information on the *dt-1*, i.e the preceeding value of duration. And the preceeding values of other properties too, like the loudness of the preceeding syllable. And then we have *c11 · dt-2* if we're going back two segments and looking at the duration. With two segments of context, you're looking at *t-2*, and so forth.

These predictors can be fairly long: they could have up to 80 or so terms, and each term has a co-efficient. Each predictor yields a value for a single property. And, we train it in one of these classes of syllables, like phrase-final, consonants in English.

How do you evaluate a predictor?

- Standard statistical tools.
  - Pearson's r^2 tells you how well it predicts.
  - r^2 is the fraction of the total variance that is predicted.

Actual

$R^2=1$ 100% predictable
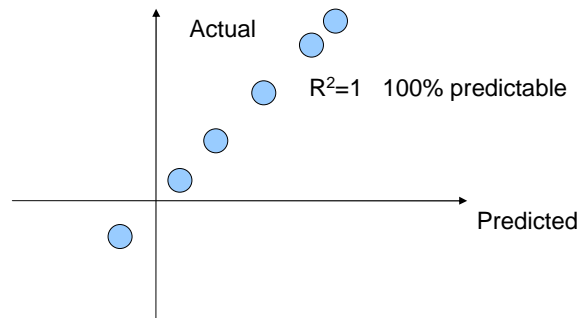
Predicted

How do you evaluate a predictor? Use standard statistical tools: Pearson's *R2* tells you how well it predicts. *R2* is the fraction of total variance that is correctly predicted. If, for example, we have predicted accurately, a plot of predicted *vs.* actual values will form a nice straight line. Here's a different example where prediction is useless and R2 is zero; the predicted value isn't related to the actual value.

How do you evaluate a predictor?

- Standard statistical tools.
  - Pearson's r^2 tells you how well it predicts.
  - r^2 is the fraction of the total variance that is predicted.

$R^2=0$ unpredictable
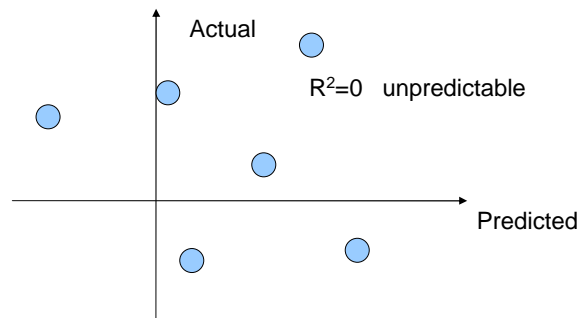
Then what?

- There were 320 predictors
- We ran each one 10 times
  - Each with a new bootstrap sample of data
  - This allows us to do see if some values of r^2 are just lucky accidents
    - (Generally, they are not.)
- Then, we made histograms of r^2.

Then what? We have 320 predictors for poetry or prose in each language. We ran each one ten times, each with a new bootstrap sample of the data. That's a statistical technique that allows us to see if some values of R2 are just lucky accidents or not. And then we made histograms of R2: there are lot of linear regressions here.

Results

Prose

mean=0.18
std=0.12
n=8680

Poetry

mean=0.46
std=0.27
n=9470

Pearson's r^2
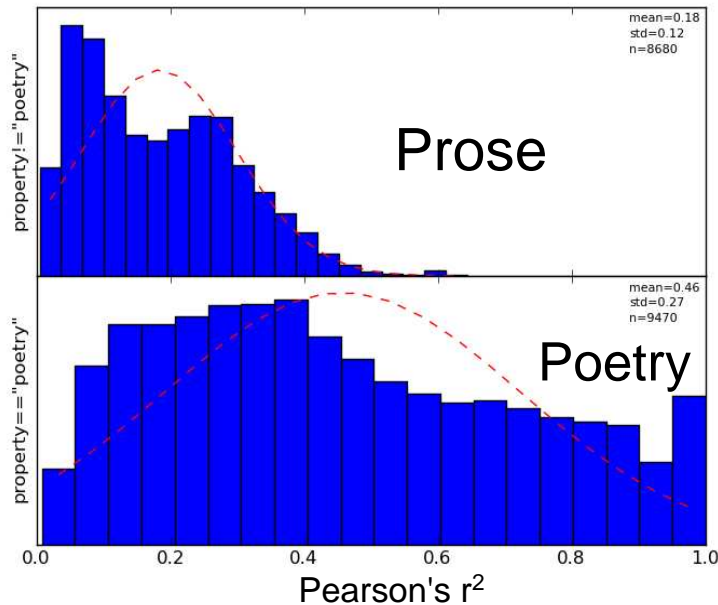
Can we tell poetry from prose with this technique?

Yes!

At least half (i.e. about 160) of the predictors give statistically significant discrimination. Actually almost all of them.

Effect sizes are huge.

So, can we tell poetry from prose with this technique? Definitely, yes. In this plot here we have Pearson's R2 across the bottom, going from unpredictable on the left to fully predictable on the right. Prose is on top, poetry is on the bottom. And you can see that most of the prose values are reasonably predictable. And we can see that most of these predictors, and remember each dot on this histogram is the result of an entirely different linear regression, each predicts a different thing with a different equation. Most of these equations are fairly good predictors, giving a Pearsons R2 of about 0.2 or 0.3, so they explain 20% or 30% of the variance. But for poetry we find (shockingly: almost never you see this in speech and language) you find some equations predicting almost 100% of the variance. Some aspects of poetry are very predictable. So almost all these predictors are statistically significantly better for poetry than prose, and the effect sizes are huge.

Poetry is generally twice as predictable as prose is.

Story 1:

English is the typical stress-timed language

Should have a strong alternation in properties in the iambic/trochaic feet

It ought to be easily predictable being an alternation.

French is the typical syllable-timed language

Should have little difference from syllable to syllable

Smaller differences will be hidden more by normal variation, and therefore will be generally unpredictable
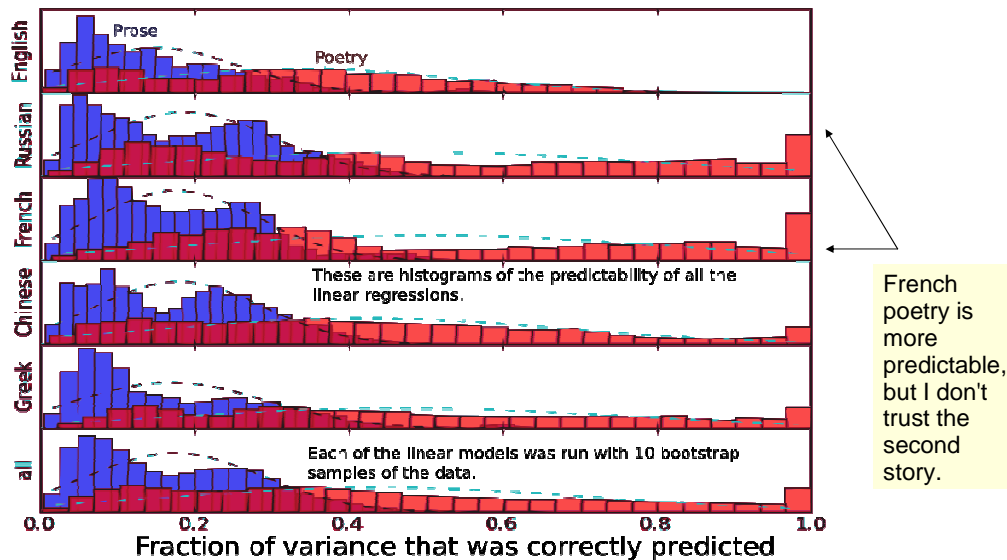
Therefore English poetry should be more predictable than French

Now, this is science here, and we are going to take the opportunity to make a prediction and test it. There aren't more than a few people in this room who actually know the answer to it, so I'm hoping they'll keep quiet and the rest of you guess. I'm going to tell you two stories about whether English or French poetry should be more predictable, and you get to choose: Story number one is that English is a typical stress timed language. Given lexical stress it should have a strong alternation in properties in the iambic or trochaic beat. [beats out time16:47] These are easily predictable, being an alternation. On the other hand, French is a typical syllable timed language. It should have relatively little difference from syllable to syllable. Smaller differences will hidden more by normal variation, and therefore properties of French syllables should be less predictable. So, story one says that English poetry should be more predictable than French.

- Story 2:
  - French has something like stress
    - It's determined by position in the phrase
      - Therefore it should be completely predictable
      - (although there aren't very many of them)
  - English has stress
    - But English poetry doesn't always follow a perfect meter
    - (although in this easy poetry, it normally does)
  - Therefore French poetry might be more predictable than English

Story two, of course, says the reverse. French does have something like stress, at least, it behaves accoustically like stress, but it's determined by position in the phrase, not the word. So it should be completely predictable. (Although, perhaps, not too strong of an affect because there aren't that many of these phrasal stresses.) English has stress, but English poetry doesn't always follow a perfect metre. So, it should be fairly predictable, but maybe not perfectly predictable. Therefore, in this story, French poetry might well be more predictable than English. Hands up for which you believe?
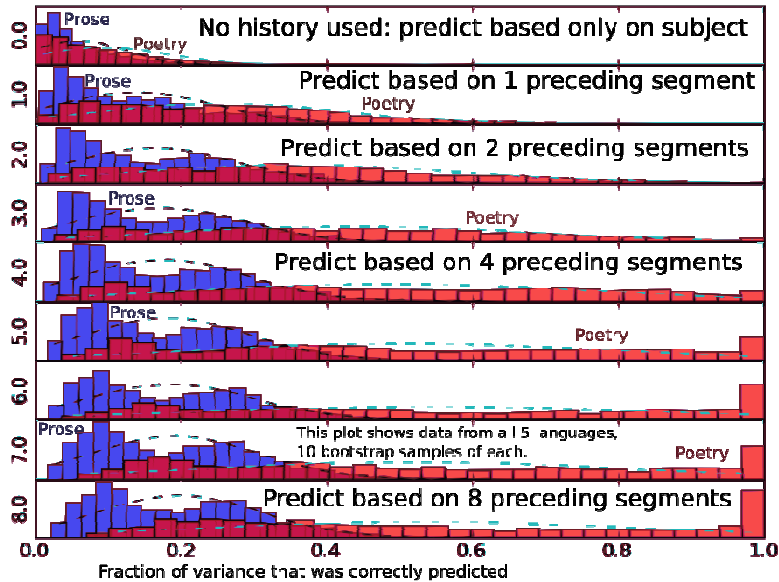
In all languages studied, acoustic properties of poetry are much easier to predict than those of prose.

English Russian French Chinese Greek all

Prose

Poetry

These are histograms of the predictability of all the linear regressions.

Each of the linear models was run with 10 bootstrap samples of the data.

French poetry is more predictable, but I don't trust the second story.

0.0 0.2 0.4 0.6 0.8 1.0

Fraction of variance that was correctly predicted

Well, I'd believe the first story, but it actually turns out that French poetry was rather more predictable than English. English is the top line here, and the horizontal axis is as before. Pearson's R^2 ranging from zero (means unpredictable) to one (meaning that at given predictor works very, very well). English poetry pops out at about 75% predictable and most of it is in the neighbourhood of 40% - 50%. But French poetry, oh the French, the French. They have quite a number of predictors that work, get near 100%. And the average predictability was probably, oh, 0.6 or 0.7, I'm guessing. So, do we know why that's true? No. But one of the beauties of an experiment is you can find out what the right answer is and you can test your theory. And you can see that the theories we have aren't really sufficient to make a prediction.

Now, you can see a few things from this block. One is that, in all the languages, (English at the top, Russian, French, Chinese, Greek) that poetry is more predictable than prose, by a substantial factor. Even in English. And you can see that there are differences from language to language. English poetry is probably the least predictable; Chinese poetry is the next, followed by Greek; and then French and Russian poetry are quite predictable. And you can see some differences in the predictability of prose too. Russian and French and Chinese have a second hump at around $R^2 = 0.25$, whereas English and Greek don't.

How much history matters?

The plot contains the following labels from top to bottom:
- **No history used: predict based only on subject** (Prose, Poetry)
- **Predict based on 1 preceding segment** (Prose, Poetry)
- **Predict based on 2 preceding segments**
- (Prose, Poetry)
- **Predict based on 4 preceding segments** (Prose)
- (Prose, Poetry)
- (Prose, This plot shows data from a 15 languages, 10 bootstrap samples of each. Poetry)
- **Predict based on 8 preceding segments**

X-axis: **Fraction of variance that was correctly predicted** (0.0, 0.2, 0.4, 0.6, 0.8, 1.0)
Y-axis: 0.0, 1.0, 2.0, 3.0, 4.0, 5.0, 6.0, 7.0, 8.0

More than 2 or 3 segments of history does not help predict prose. (About 1-word's worth of history matters.)

Poetry keeps getting better and better to eight segments and perhaps beyond. (Four syllables matter, maybe more.)

The other thing we can look at is how much history matters. Remember, some of these predictors use a long history (up to eight segments), and other predictors use, no history, just knowledge of the subject, to predict what's going on in the next segment. This plot here is the same as before on the bottom axis: Pearson's R^2 ranging from zero to one. In the plot, we split the predictors by the amount of history involved, ranging from zero history (just knowledge of the subject) on the top line, down to eight segment of history on the bottom line (about four syllables).

And what you see is that things get better. Not surprisingly. You use more history, you pick up more information about the past, and in poetry that lets you predict the future. In prose a little bit of history also helps you predict the future, but fairly soon, things stop getting better. You look at the blue, zero to one to two to three, predictions are better and better, but when you get up to four, they're really not much better. And five, and six and seven are about the same. So prose only can make use of a little bit of history, because prose does not have long-range patterns. Prose is intrinsically unpredictable at long distances. Poetry, on the other hand, the red here, the predictions just keep on getting better. Zero history, to one, to two, to three, to four history, continuously improving. Five history, we're starting to get some predictors coming up near $R^2 = 1$. Six and seven, and eight, more and more end up near $R^2 = 1$. So poetry intrinsically uses long-range information, because poetry has long-range metrical patterns.

A final quiz

- Which class of syllables should be more predictable in poetry?
  - Phrase initial
  - Phrase medial
  - Phrase final

  - (All predictions are based on the preceding 1-8 segments, and predictions are only made when preceding segments exist.)

Now, a final quiz. Which class of syllables should be more predictable in poetry? Phrase initial, phrase medial, or phrase final? Phrases are defined by a pause and, what it usually means is that phrase is a line in the poetry. Phrase-initial means predicting the first segment of the line, phrase-final means predicting the last segment of the line, and -medial is predicting anything in between. Anyone have a prediction?

Which part of a phrase is most predictable?
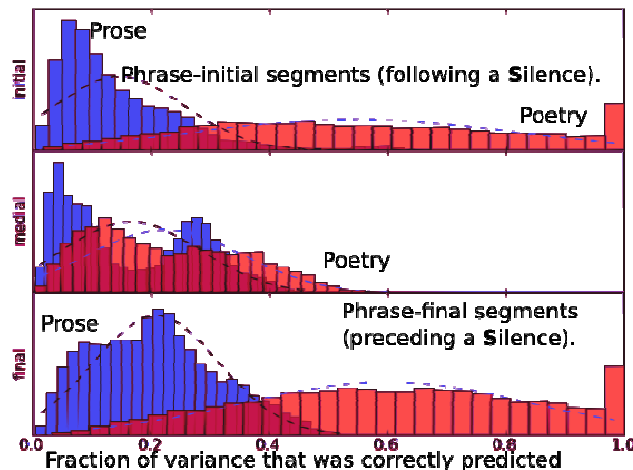
Both ends, in poetry.
Final is best for prose.

Alright, what you find, is those of you who picked phrase-final are the winners. Phrase-final syllables are much more predictable. Phrase-initial syllables, interestingly enough, are also fairly predictable, even though we're predicting across the pause. Even though we're predicting across the gap between lines. We're predicting the beginning of a line from the latter half of the preceeding line. And phrase-medial are least predictable: in poetry the middle is only a little bit easeir to predict than in prose. That's a very interesting result, and not one that I understand at this moment.

Conclusions

- It is a powerful technique.
- Unexpected results already.

So, in conclusion, I hope you agree that this is a powerful technique, and it's capable of producing unexpected results, already. Using it, we have the opportunity of learning something about rhythm that we didn't put into the technique.