Fingerprints of rhythm in natural languages

Antonio Galves

Universidade de São Paulo

Some recent results of the Tycho Brahe Project

www.ime.usp.br/~tycho

1 The rhythmic classes conjecture

It has been conjectured in the linguistic literature that languages are divided in two or maybe three rhythmic classes (Lloyd 1940, Pike 1945, Abercrombie 1967, ...).

- *Morse code* or *stress-timed* languages: English, European Portuguese, Dutch, Finish, Polish, ...
- *Machine-gun* or *Syllable-timed* languages: Brazilian Portuguese, Catalan, French, Italian, Spanish,...
- *Mora-timed* languages: Japanese, Fidji, ...

But for nearly half a century

- There was no real definition of the rhythmic properties characterizing a class.
- And no empirical correlates of the rhythmic properties in the speech signal was known.

2 Correlates of linguistic rhythm in the speech signal

Ramus, Nespor and Mehler (1999) gave for the first time evidence that simple statistics of the speech signal could discriminate between different rhythmic classes.

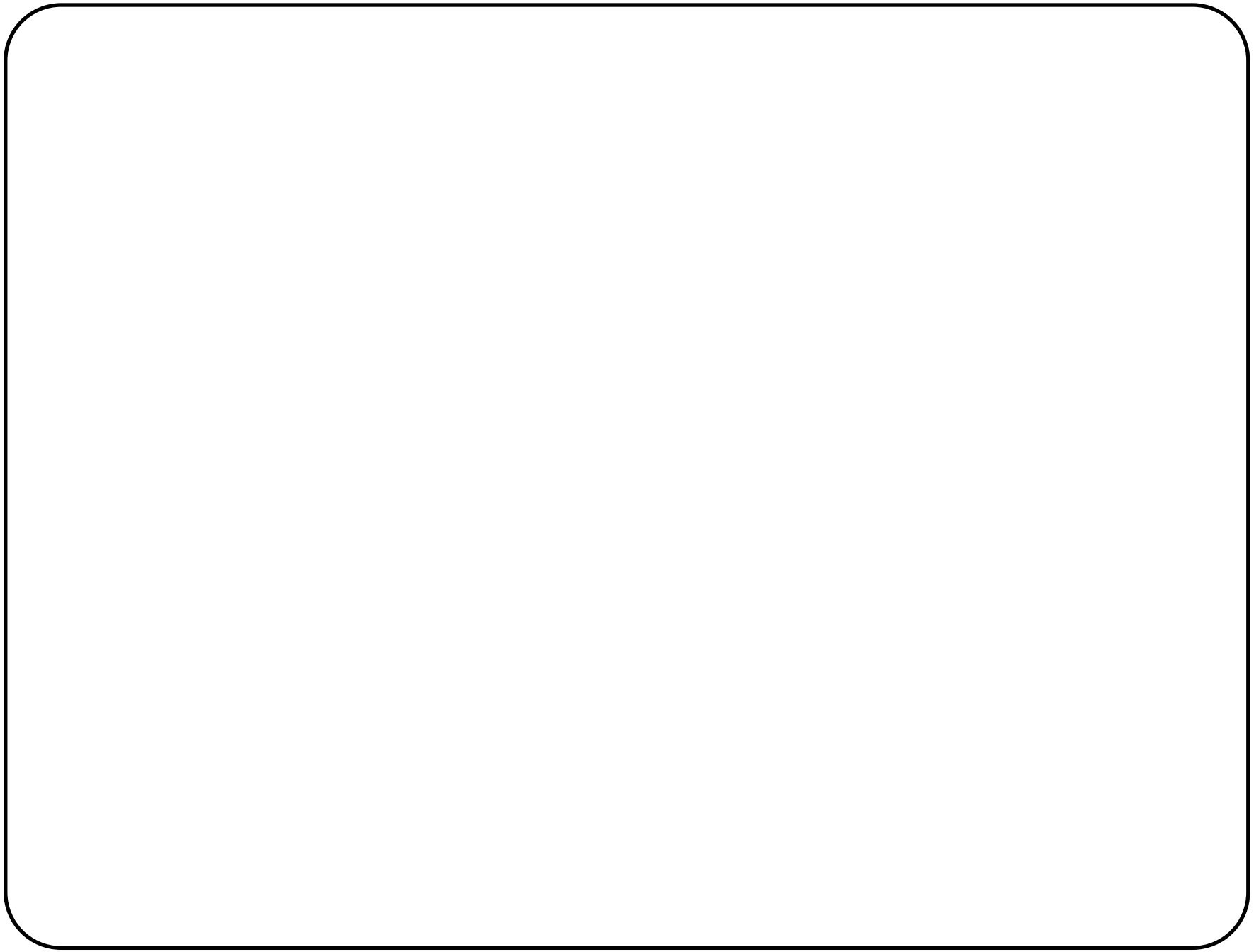
3 RNM's approach

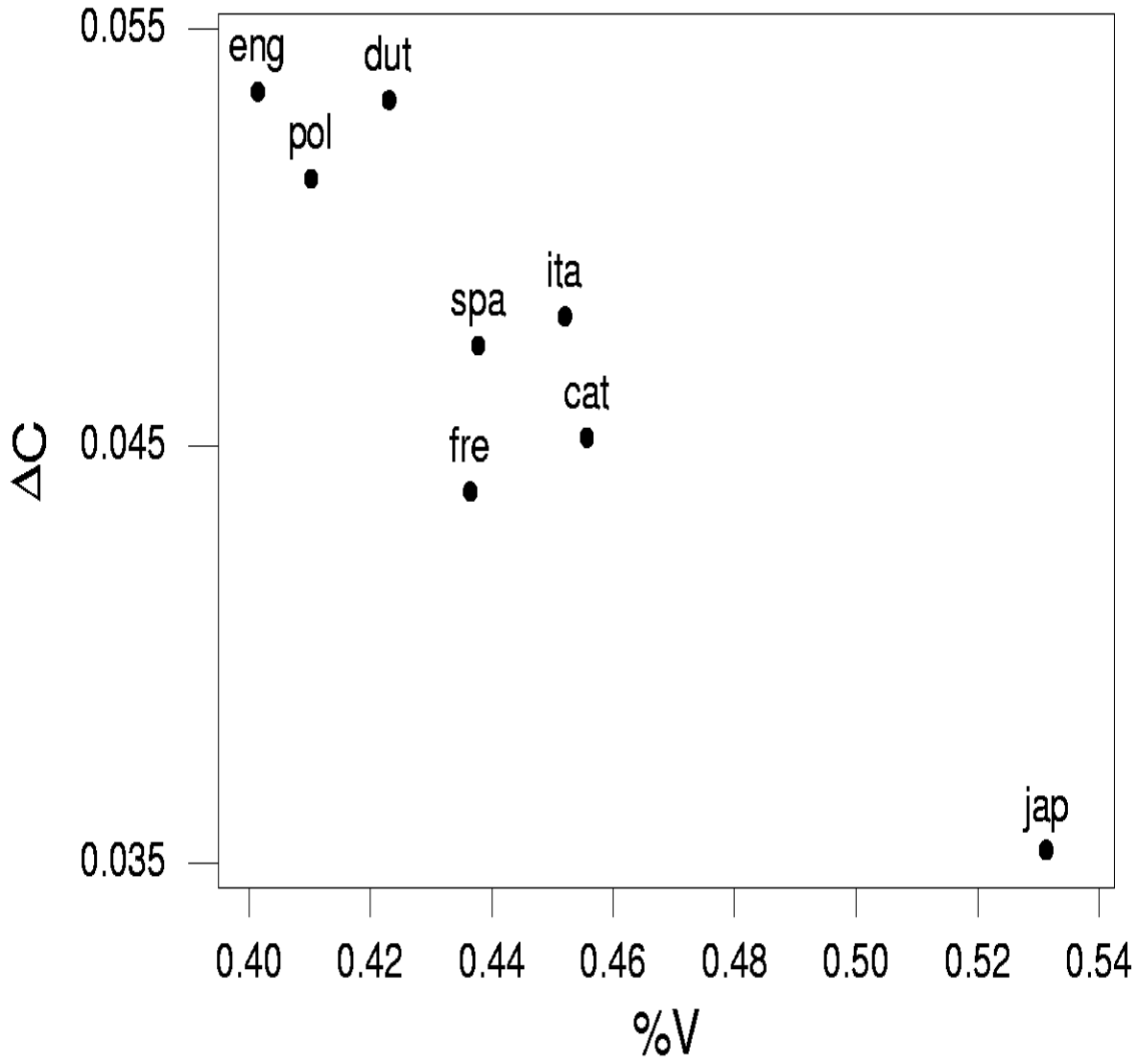
- RNM analyzed the acoustic signal of 20 sentences produced by 4 speakers of each of the following languages: English, Polish, Dutch, Catalan, Spanish, Italian, French and Japanese.
- The chosen sentences were segmented into vocalic and consonantal intervals.
- For each language, the empirical standard deviation of the durations of the consonantal intervals (ΔC) and the proportion of time spent in vocalic intervals ($\%V$) were computed.

4 Intuition behind the choice of these statistical parameters

- Languages conjectured to be syllable timed (like *Italian*) spend a greater proportion of time in vocalic intervals than languages conjectured to be stress-timed (like *English* or *Dutch*). This justifies the choice of $\%V$.
- Languages conjectured to be stress-timed display a greater variety and complexity of consonantal intervals than languages conjectured to be syllable-timed. This justifies the choice of δC.

It turns out that this was a good choice ...




0.055
0.045
0.035
ΔC
eng
dut
pol
ita
spa
cat
fre
jap
0.40
0.42
0.44
0.46
0.48
0.50
0.52
0.54
%V

5 Problems with RNM's approach

It is based on a hand label-ling of the speech signal.

- This is a time-consuming task.
- Moreover this hand label-ling is often based on decisions which are difficult to reproduce in a homogeneous way

This makes it difficult to reproduce RNM's approach on large samples.

6 A new approach to the problem

- Newborn babies are able to discriminate rhythmic classes with a signal filtered at 400Hz (Mehler et al. 1996).
- At this level, it is hard to distinguish nasals from vowels and glides from consonants.
- This strongly suggests that the discrimination of rhythmic classes by babies relies not on fine-grained distinctions between vowels and consonants, but on a coarse-grained perception of sonority in opposition to obstruency.

7 Sonority as a basis for rhythmic class discrimination

- Galves, Garcia, Duarte and Galves (2002) suggests that it is possible to discriminate rhythmic classes of language, using a rough measure of the speech sonority.
- This measure is defined directly from the spectrogram of the signal, with no need of previous hand label-ling of the data
- Applied to the same linguistics samples considered in RNM, it produces the three conjectured clusters.

8 Defining the speech sonority

- We define a function which maps local windows of the acoustic signal on the interval $[0, 1]$.
- This function is close to 1 for spans displaying regular patterns, characteristic of sonorant portions of the signal.
- In contrast, regions in which the acoustic signal present a chaotic behavior, for instance regions corresponding to stop consonants, will correspond to intervals in which S_t will assume values close to 0, with important variations

- Denote by $c_t(f)$ the power spectral density at time t and frequency f.
- Time is discretized in steps of 2 milliseconds. The values of the spectrogram are estimated using a 25 milliseconds Gaussian window.
- We only consider frequencies from 80 Hz to 800 Hz, by steps of 20 Hz.

The normalized power spectral density is defined by

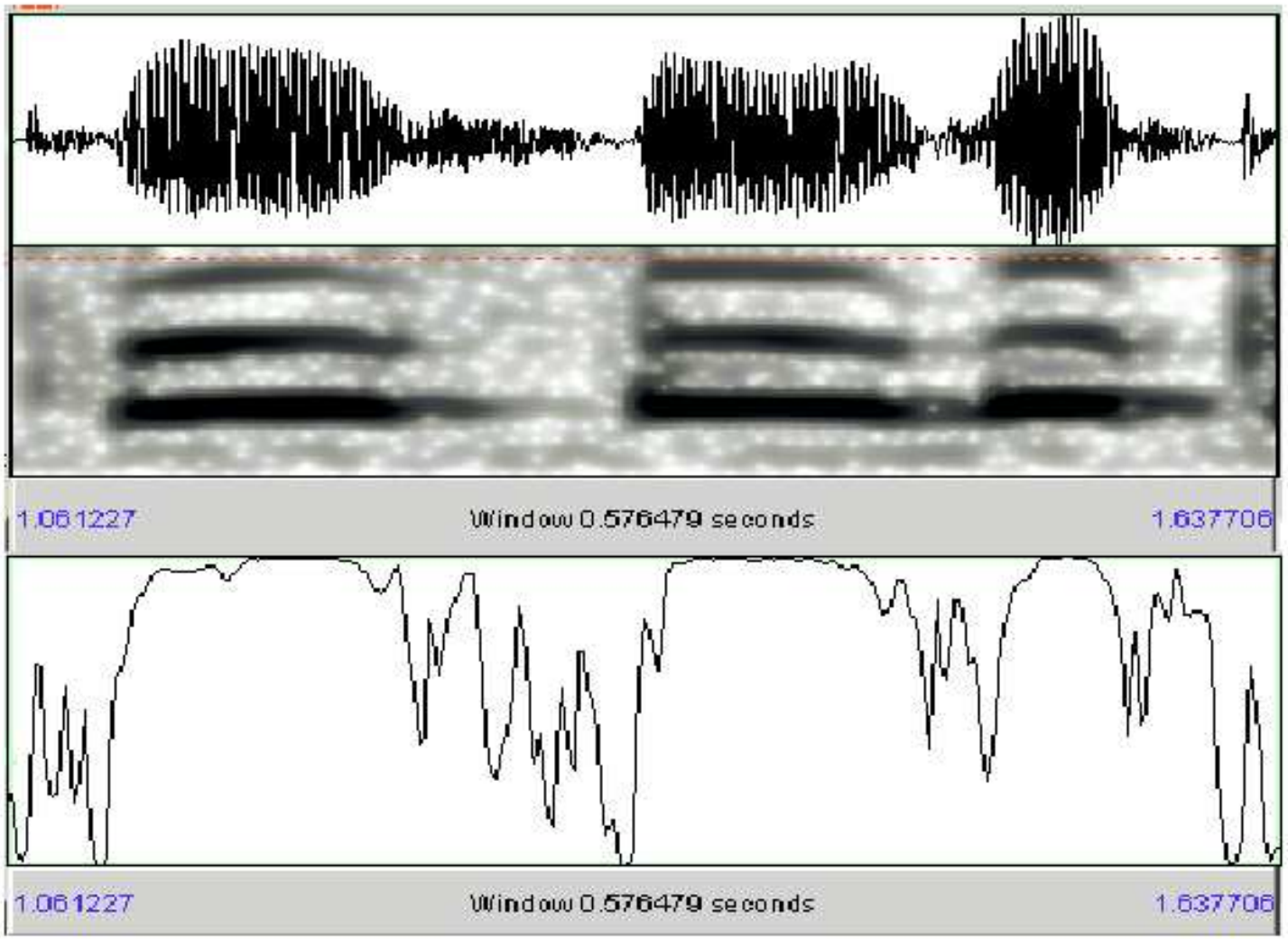
$$p_t(f) = \frac{c_t(f)}{\sum_{f'} c_t(f')} \ .$$

This defines a sequence of probability measures $\{p_t : t = 1, \ldots, \}$.

9 Definition of the sonority

$$S_t = e^{-\eta \sum_{i=1}^{3} h(p_t \mid p_{t-i})},$$

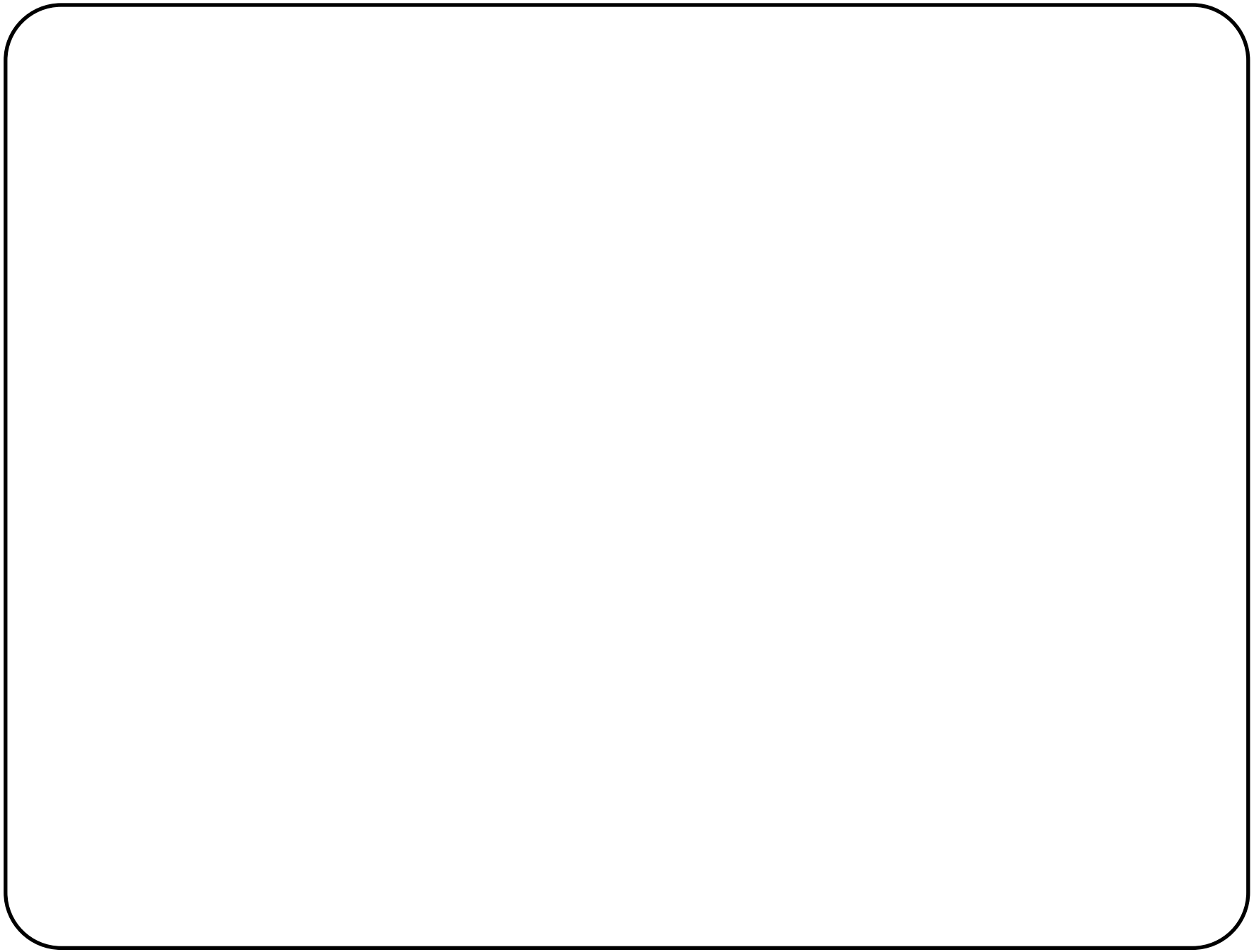
where h is the relative entropy and η is a free parameter taking positive real values.

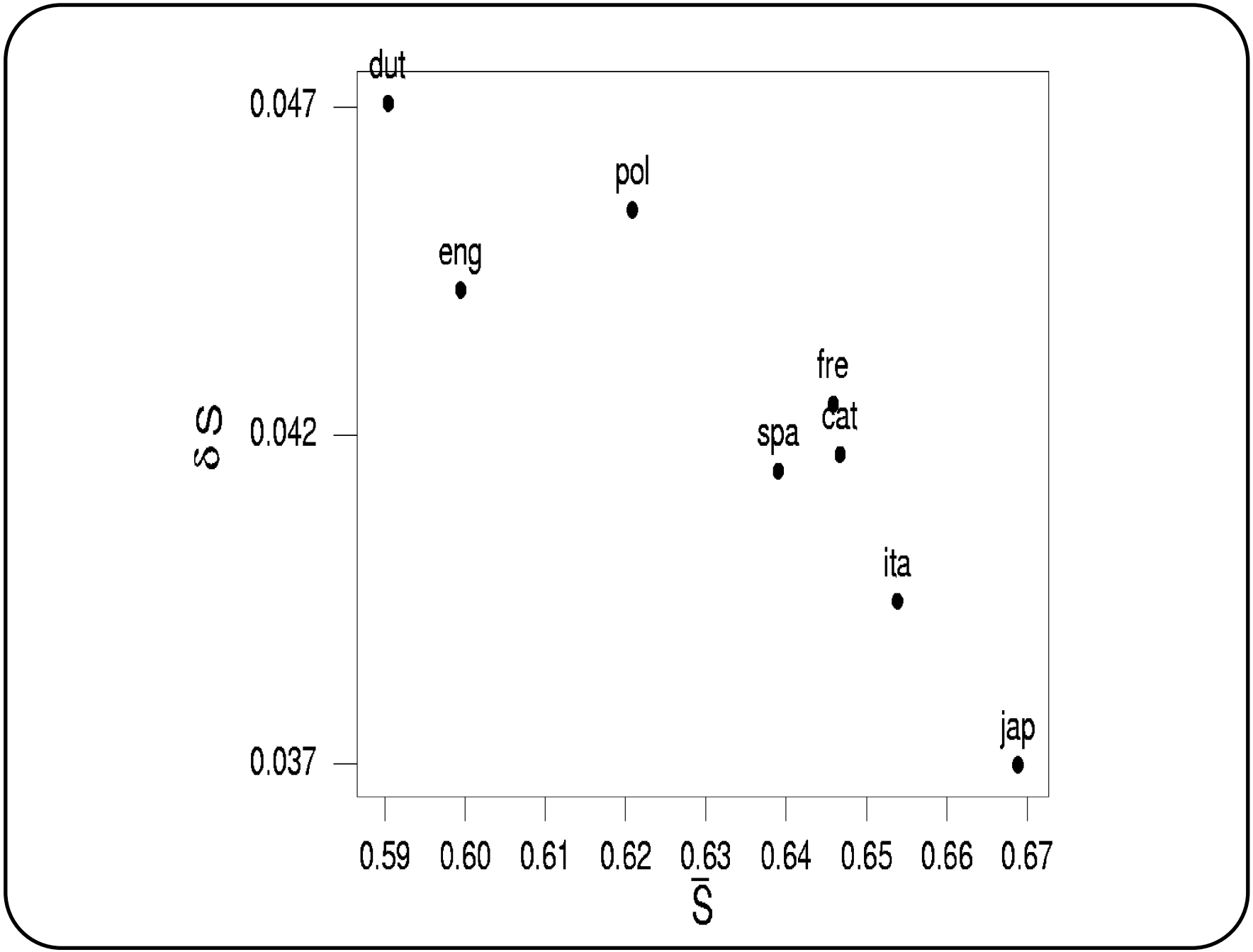

1.061227
Window 0.576479 seconds
1.637706
1.061227
Window 0.576479 seconds
1.637706

10 Reproducing RNM using the sonority

Define

- $\bar{S} = \frac{1}{T}\sum_{t=1}^{T} s(t)$. (*This will play the role of* $\%V$.)
- $\delta S = \frac{1}{T}\sum_{t=1}^{T} |s(t) - s(t-1)|$. (*This will play the role of* ΔC.)




dut
pol
eng
fre
cat
spa
ita
jap
0.047
0.042
0.037
s_S
0.59
0.60
0.61
0.62
0.63
0.64
0.65
0.66
0.67
$\bar{S}$

11 A sound statistical basis to the clustering

- The pictures produced with both approaches suggest that the existence of three classes.
- Is this a real statistical fact ?
- The projected Kolmogorov-Smirnov test presented in Cuesta-Albertos, Fraiman and Ransford (2004) makes it possible to compare the laws of the stochastic processes producing the time evolutions of the sonority for the different sentences and languages.

12 The projected Kolmogorov-Smirnov test

- First choose a *direction* $W = (W(t))_{t\in[0,T]}$ at random. and then project the sonority trajectories in this direction.
- Then calculate the Kolmogorov-Smirnov statistic $D_W(\mathcal{S}_l, \mathcal{S}_{l'})$ for the two projected samples
- Reject the null hypothesis that the two samples belong to the same population if $D_W(\mathcal{S}_l, \mathcal{S}_{l'})$ is large enough. Otherwise accept it.

13 Stable projected Kolmogorov-Smirnov test

Instead of taking only one random direction we will take many of them. For each pair of languages $l \neq l'$ we proceed as follows.

- Choose 100 independent direction $W_i : i = 1, ..., 100$.
- Test if the samples corresponding to l and l' belong to the same population using the projected KS test using direction W_i.
- Build up the auxiliary random variable $Z_i(l, l')$ which takes the value 1 if the projected test in direction W_i rejects the null hypothesis, and takes the value 0 otherwise.
- Reject the null hypothesis if the average statistic

$$\bar{Z}(l, l') = \frac{1}{N} \sum_{i=1}^{N} Z_i(l, l') \geq c_\alpha \, .$$

14 Results using the sonority

language	pol	ital	fren	span	dut	eng	cat
jap	0.04	**0.43**	**0.09**	**0.08**	**0.77**	**0.74**	0.01
pol		0.03	0.0	0.0	**0.60**	**0.21**	0.03
ital			0.03	0.02	**0.14**	0.0	0.05
fren				0.0	**0.50**	**0.19**	0.06
span					**0.41**	**0.08**	0.03
dut						0.0	**0.74**
eng							**0.58**
$c^{\star}_{0.05}(l)$	0.13	0.13	0.12	0.11	0.14	0.12	0.12
$c^{\star}_{0.1}(l)$	0.05	0.05	0.06	0.05	0.05	0.04	0.05

15 Results with groups of languages

We performed a new test by grouping the sonority sample paths in three groups.

- In the first group we put together the 60 sonority sample paths of the conjectured syllable-timed languages, French, Italian and Spanish.
- The second group contains the 40 paths of the conjectured stress-timed languages, Dutch and English.
- Finally the 20 sonority paths of Japanese, which is conjectured to be a mora-timed language, remain in a third group.

Results

category	mora–timed	stress–timed	$c^{\star}_{0.05}(i)$
syllable–timed	0.32	0.70	0.24
stress–timed	0.82		0.09
$c^{\star}_{0.05}(j)$	0.06	0.04	

Table 1: *Values of* $\bar{Z}(i,j)$ *and bootstrap critical values for the three groups, with* $N = 100$, $B = 1000$ *and* $\eta = 0.05$

The test found significant all the differences between groups. This reinforces the linguistic conjecture of existence of three different rhythmic classes.

16 A model for the sonority

- A cross-linguistic exploratory analysis of the data shows that the sonority is quite regular in high level regions and displays strong variations below a certain level.
- This suggests the modeling of the sonority time evolutions for different languages by a family of tied quantized chains.

17 Tied quantized chains

- The chains are tied together by the assumption that the distribution of the sonority, conditioned on the fact that it belongs to a given region, is *universal*, i.e. language independent.
- In particular the partition in regions of sonority is assumed to be language independent.
- In this model the specific features characterizing each language are expressed by the symbolic chain indicating in which region of sonority the process is at each time step.

18 Basic assumption

- There exist a positive integer N and an increasing sequence of cut-points $c_0 = 0 < c_1 < \ldots < c_N < c_{N+1} = 1$
- There exist $N+1$ probability measures π_j, $j = 0, \ldots, N$, such that the support of π_j is contained in the interval $I_j = [c_j, c_{j+1}[$
- such that at any time step t and for any $l \in \mathcal{L}$ we have

$$\mathbb{P}\left\{S_t^l \in \cdot | S_t^l \in I_j\right\} = \pi_j(\cdot) \, .$$

19 A consistent estimation of the universal cut-points

Cassandro *et al.*(2005) introduced a consistent algorithm to estimate the cut-points.

An analysis of the sonority data reveals four cut-points estimated as $c_1 = 0.19$, $c_2 = 0.46$, $c_3 = 0.67$ and $c_4 = 0.93$.

Cassandro, Collet, Duarte, Galves and Garcia (2005) proposes a model in which they claim that all the relevant linguistic information concerning the sonority should be retrieved from a symbolic stochastic chain taking values on a finite alphabet.

The statistical analysis support this claim!

20 Statistical analysis of the quantized sonority

We apply the projected Kolmogorov-Smirnov test to the quantized chains, obtained from the sonority sample paths.

The quantization was made using the four universal cut-points $c_1 = 0.19$, $c_2 = 0.46$, $c_3 = 0.67$ and $c_4 = 0.93$.

language	pol	ital	fren	span	dut	eng	cat
jap	0.04	**0.34**	**0.07**	**0.08**	**0.79**	**0.73**	0.02
pol		0.07	0	0	**0.73**	**0.26**	0.04
ital			0.03	0.02	**0.12**	0	0.02
fren				0	**0.48**	**0.09**	0.03
span					**0.37**	0.05	0.02
dut						0.01	**0.67**
eng							**0.5**
$c^{\star}_{o.05}(l)$	0.13	0.18	0.14	0.13	0.11	0.11	0.14
$c^{\star}_{0.1}(l)$	0.05	0.08	0.07	0.06	0.06	0.04	0.06

21 Final comments

If we consider only six languages (Dutch, English, French, Italian, Japanese and Spanish)the test based on the sonority paths suggest the existence of three clusters.

- The first one contains French, Italian and Spanish (syllable-timed).
- The second one contains Dutch and English (stress-timed) .
- Japanese appears isolated as a third cluster (mora-timed).

This clustering is compatible with the linguistic conjecture.

22 The case for Catalan and Polish

The ambiguous position of Catalan and Polish data is coherent with the fact stressed by several linguists that these languages present features common to both *stress-timed* and syllable-timed languages.

From this point of view, the failure of the statistical analysis could be considered a positive fact for the model ...

23 Further directions of research

- We still don't know what is a rhythmic feature!
- We still don't have a model for the rhythmic classes
- But the symbolic chains behind the sonority open new perspectives of research.
- An example of this is the discrimination between Brazilian and European Portuguese, modeling stress contours obtained by codifying written texts using Probabilistic Trees.

24 The Tycho Brahe team

The results presented have been obtained by

Marzio Cassandro, Pierre Collet, Denise Duarte, Ricardo Fraiman, Charlotte Galves, Jesus Garcia, Marcela Svarc

www.ime.usp.br/t̃ycho