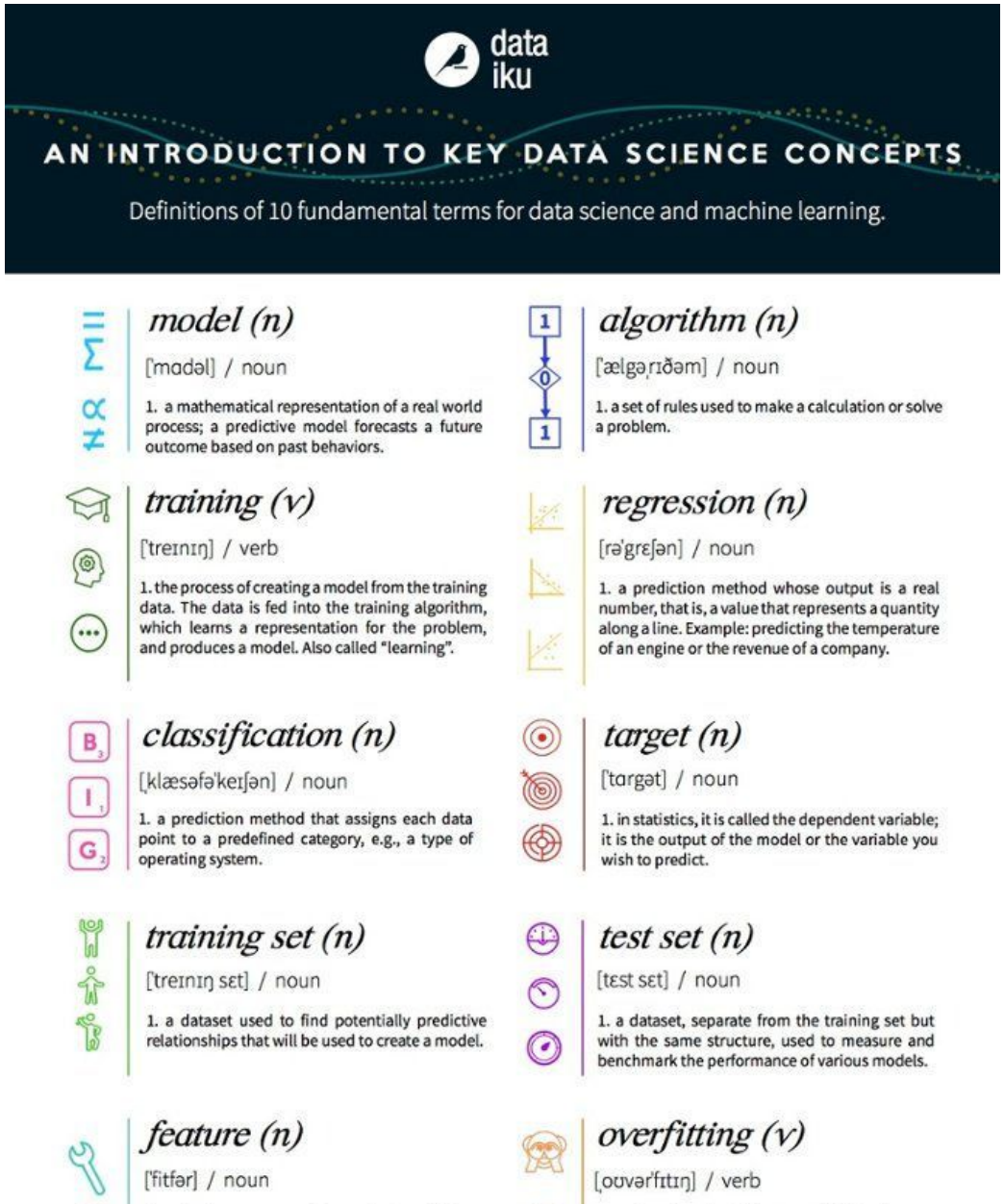
A set of functions to analyze sound time series developed in python

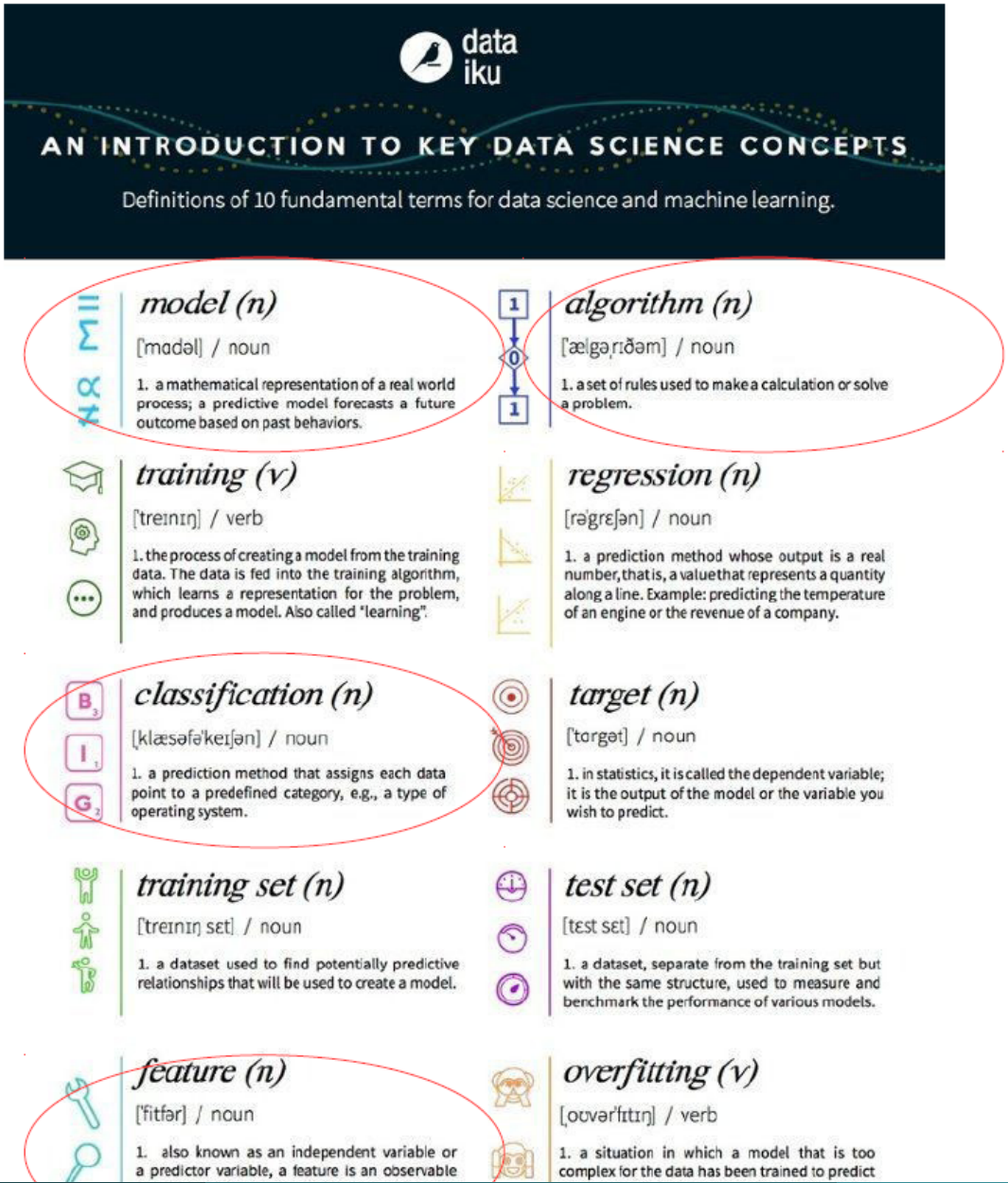
Cecilia Jarne

cecilia.jarne@unq.edu.ar

Common task on data Analysis



Common task on data Analysis



Time series

A time-series T is an ordered sequence of n real-valued variables:

$$T = (t_1, ..., t_n), t_i \in R$$

is often the result of the observation of an underlying process in the course of which values are collected from measurements made at uniformly spaced time instants and according to a given sampling rate.

Common task on time series data mining

- **Representation and indexing:**
  How to represent the time series data and the approach of transform to another domain for dimensionally reduction
- **Segmentation:**
  How to cut time series in meaningful parts.
- **Similarity measure:**
  Could be between different time series or between segments of the same time series.
- **Visualization and mining:**
  Including: pattern discovery, clustering, classification and other tasks.

WTPC workshop en técnicas de programación científica

The type of data: Bird songs

@Gabrielluislio

Four ideas to show:

- **Representation:** How to obtain envelop (adaptable to different kinds of data).
- **Dimensionally reduction:** How to obtain fundamental frequency of tonal sounds (Canary bird songs example).
- How to cut signal and obtain certain **features**.
- How to align segments (**Signal averaging**).

WTPC
workshop en técnicas de programación científica

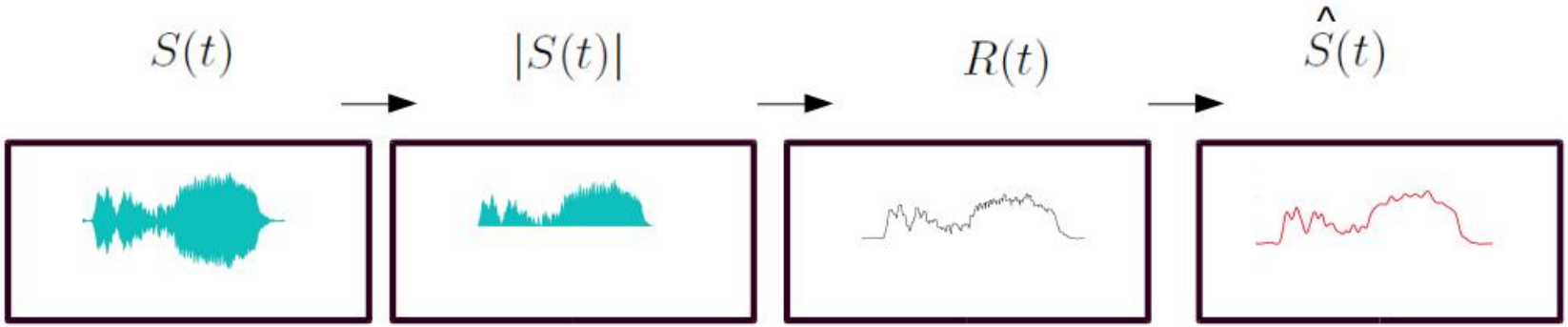
Representation: An heuristic way to obtain signal envelop

- First step: to take absolute value of signal $S(t)$ meaning $|S(t)|$
- Second step:to divide the $|S(t)|$ in k bunches of N samples:

$$|S(t)| = S_1(t) + ... + S_i(t) + ... + S_k(t)$$

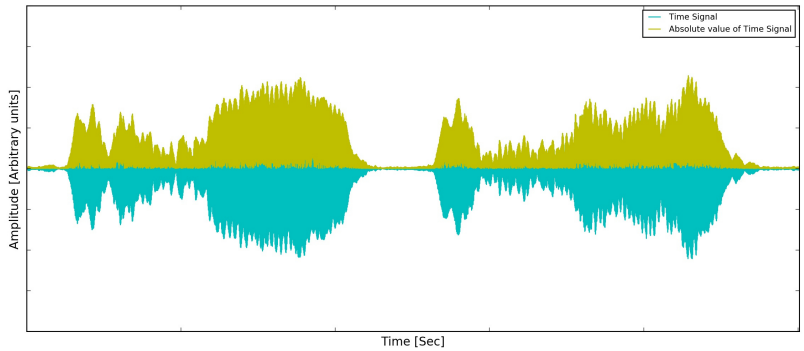
(Each $S_i(t) > 0$). The maximum value is: $Max(|S_i(t)|) = M_i$ corresponding to the j element value in each bunch.

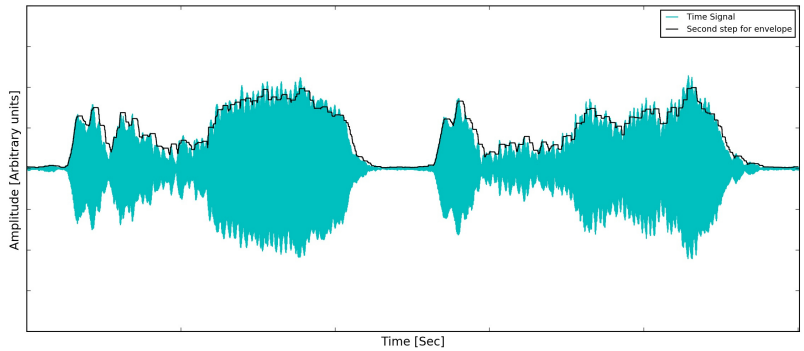
- The third step: a lowpass-filter applied to the resulted signal $R(t)$ to get rid of the remaining staircase ripple $\hat{S}(t)$.

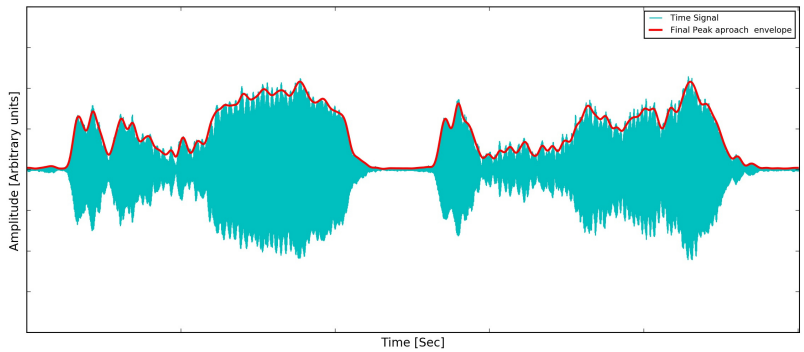


https://arxiv.org/abs/1703.06812

Representation: An heuristic way to obtain signal envelop

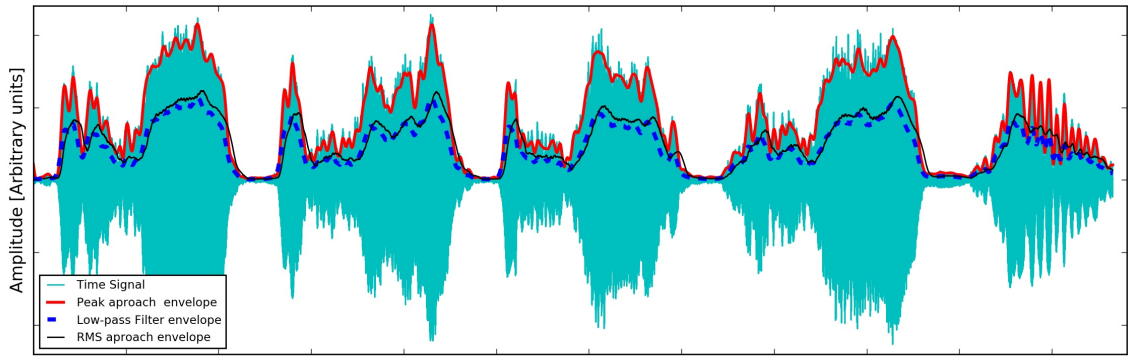


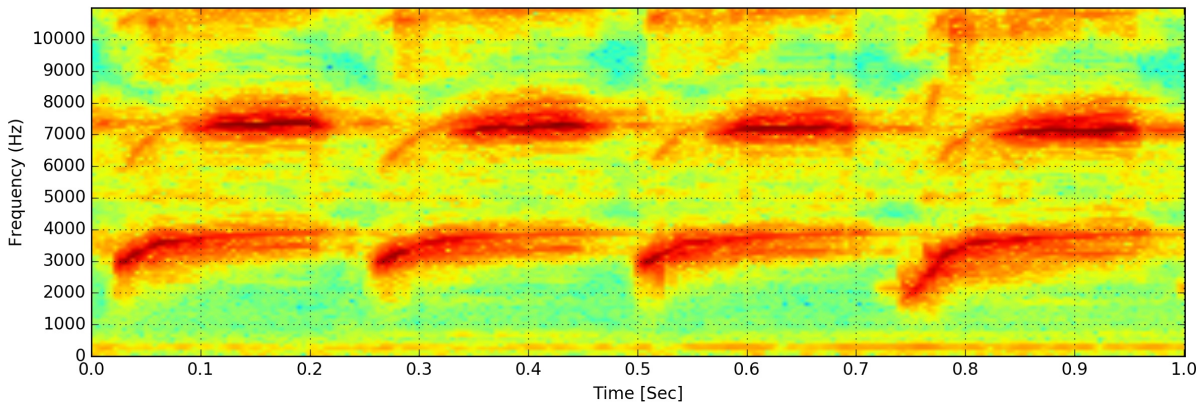




WTPC
Workshop en técnicas de programación científica

Representation: Comparison with other methods

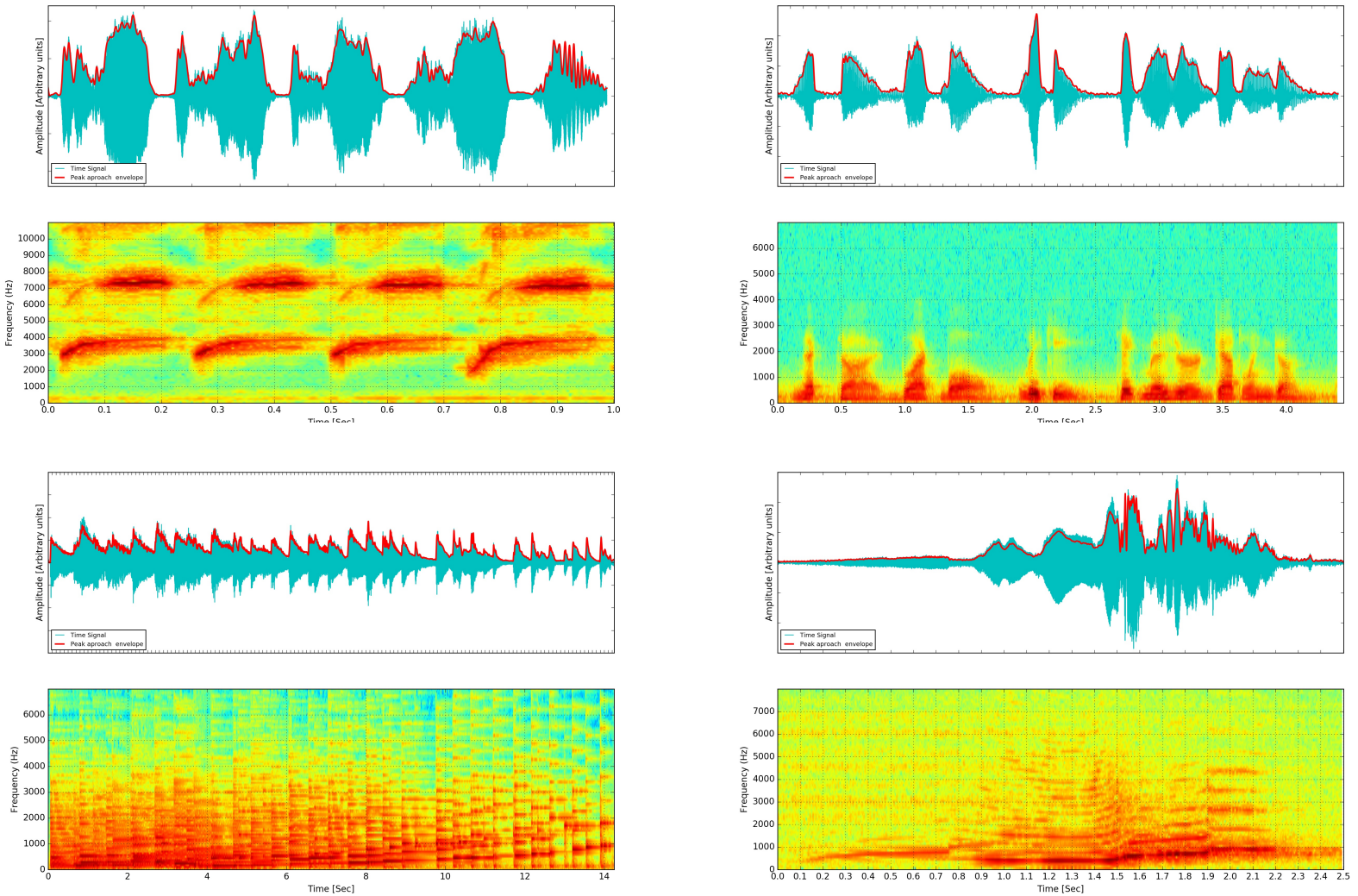




WTPC
Workshop en técnicas de programación científica

Representation: An heuristic way to obtain signal envelop

Different examples of application:

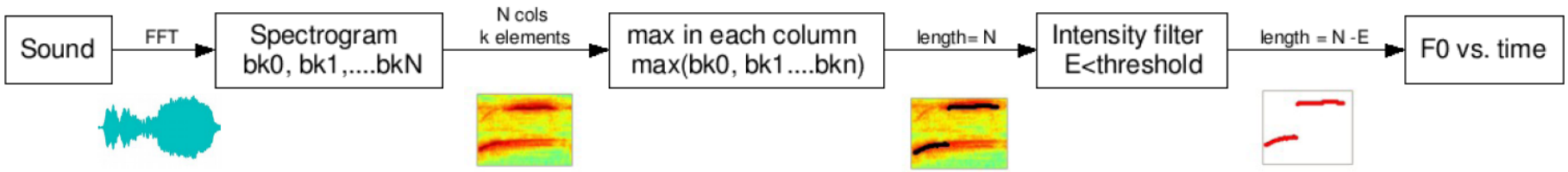


Code available in: `http://ceciliajarne.web.unq.edu.ar/investigacion/`

Dimensionally reduction: fundamental frequency (tonal)

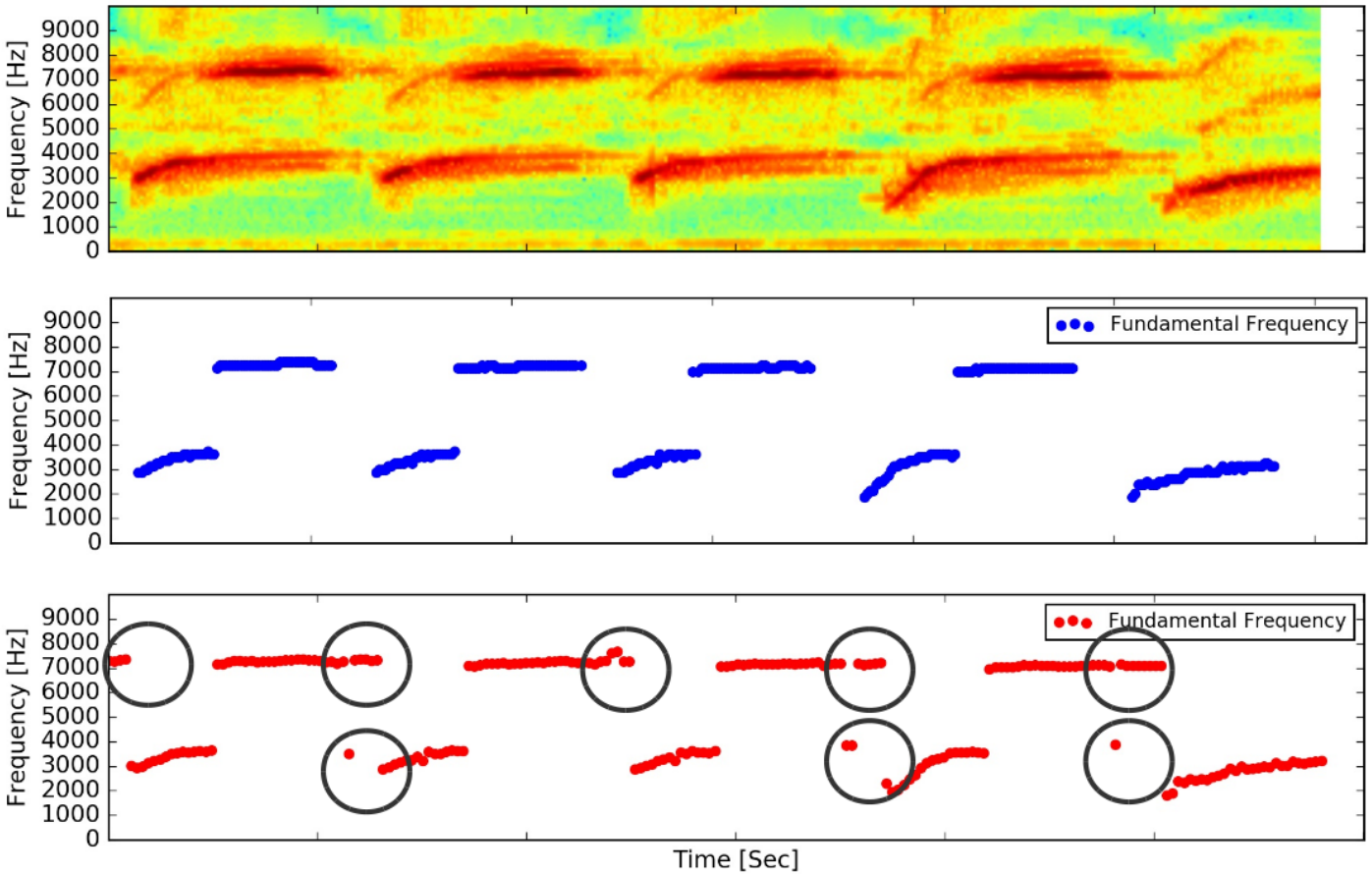
An algorithm to obtain fundamental frequency for tonal sounds:

- Spectrogram function is used over the consecutive signal segments.
- Only the value of maximum amplitude is saved as a new variable in each temporal bin.
- A second intensity filter is used to get rid of the noise and keep only the values representing the desired sound (consider only sounds that exceed a certain threshold).



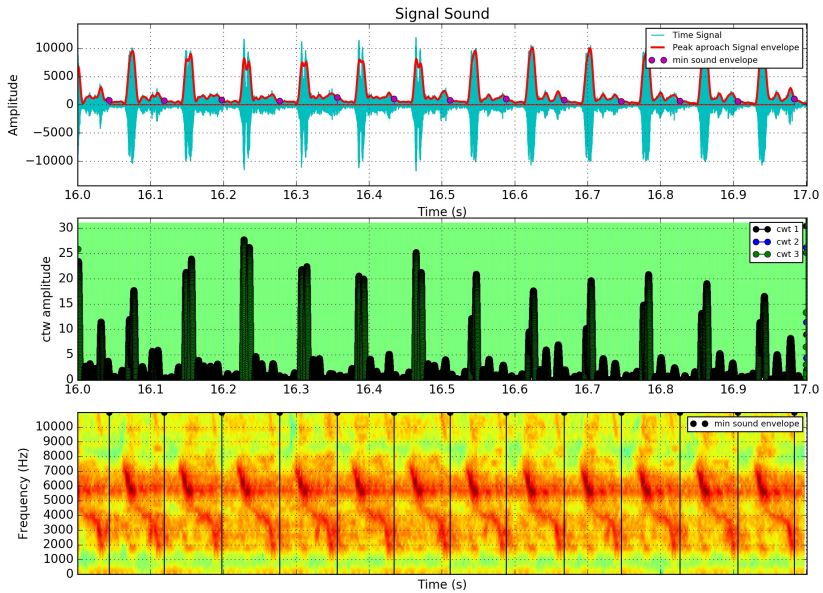
Code avialable in: urlhttp://ceciliajarne.web.unq.edu.ar/investigacion/

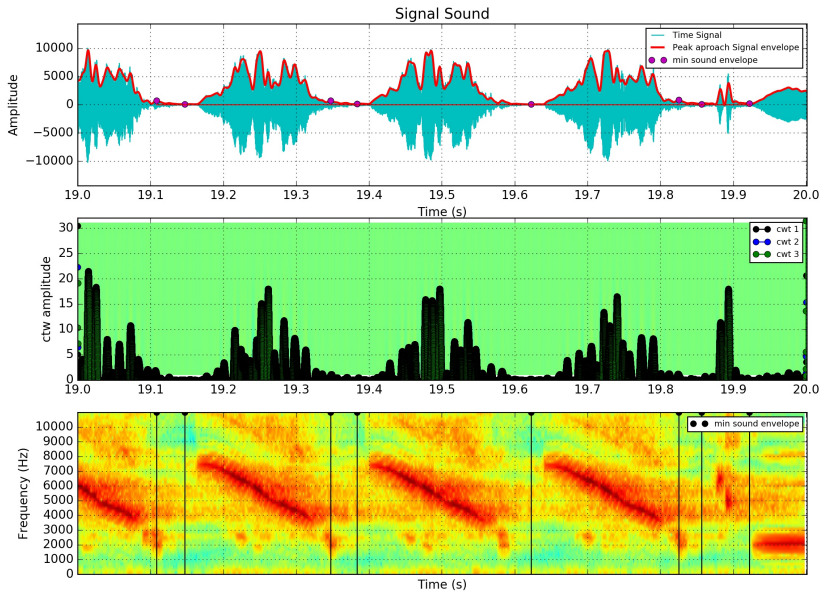
An algorithm to obtain fundamental frequency for tonal sounds:



Dimensionally reduction: Wavelet decomposition

Performs a continuous wavelet transform on data, using the wavelet function.

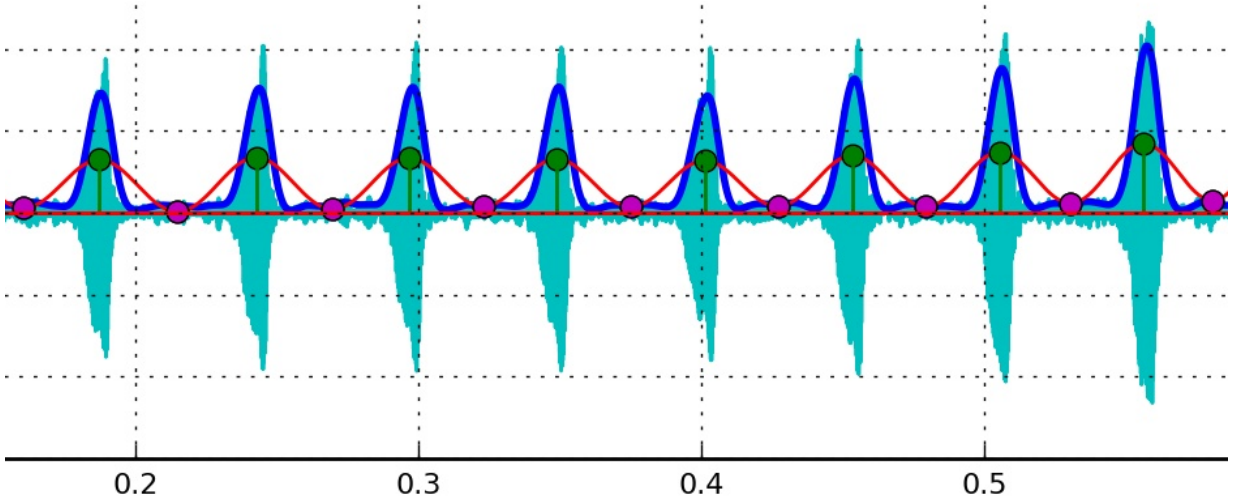





```
1 import scipy.signal as signal
2 ...
3 cwtmatr = signal.cwt(filtered_aver, signal.ricker, widths)
```

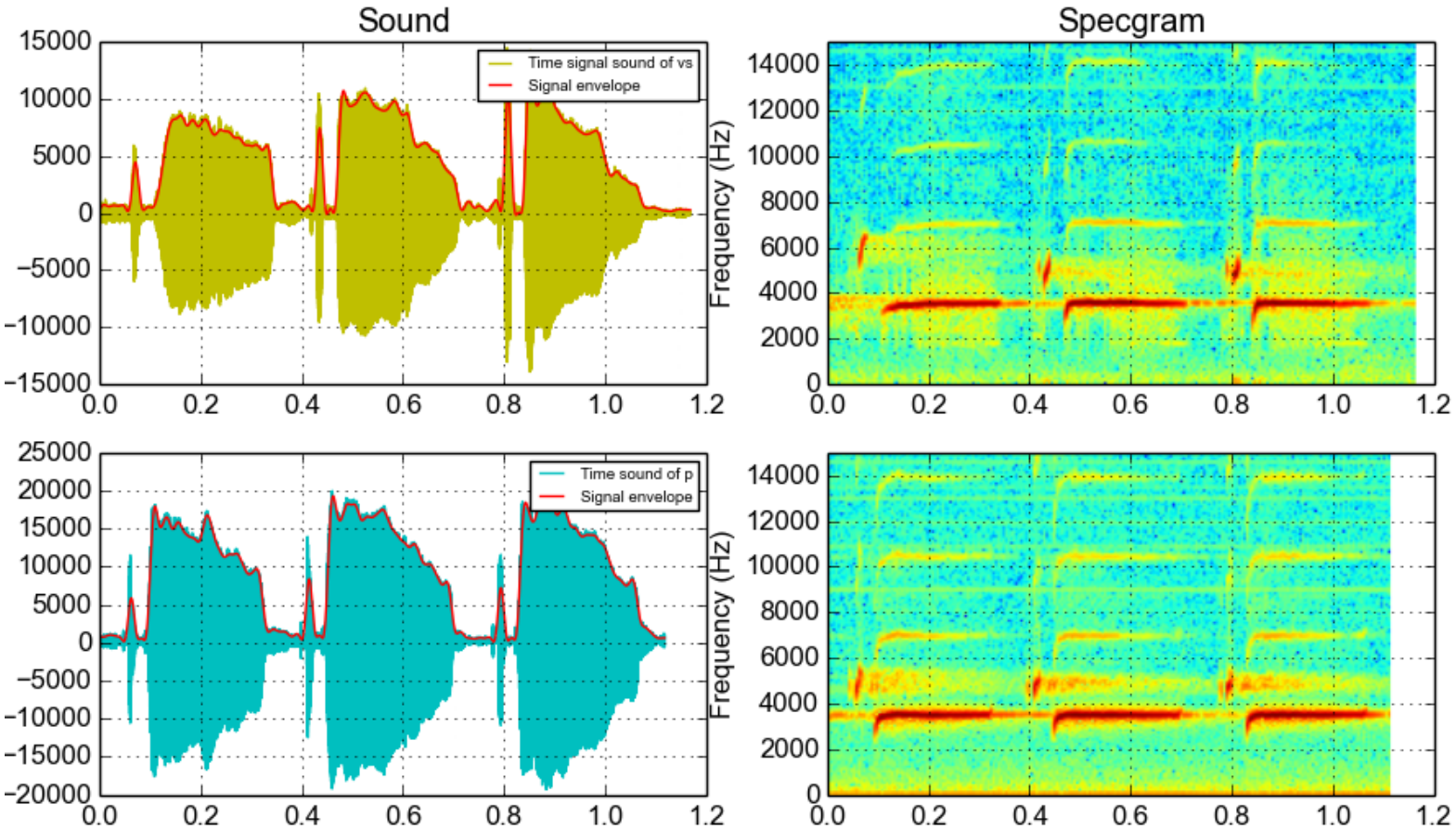
Syllable individualization: An idea for periodic section of time series

A very low pass filter applied and then finding the minima:



A way to align similar signals using correlation

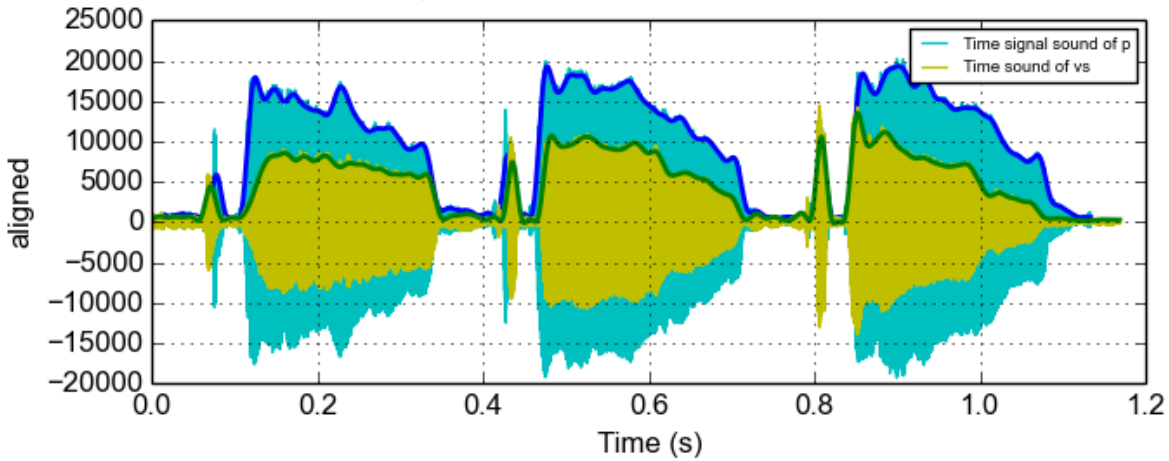
Two sound segments with similar frequency behavior.



A way to align similar signals using correlation

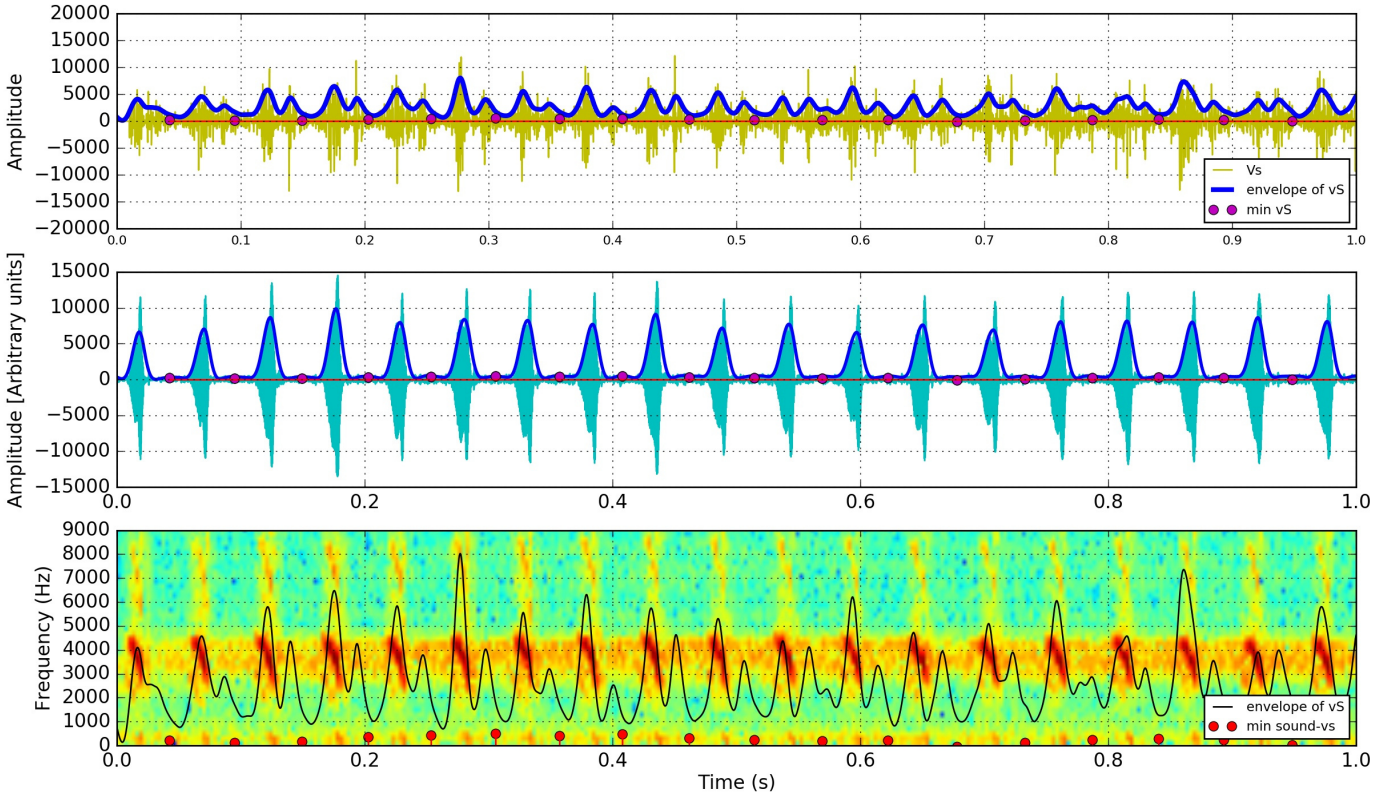
If we need to align the signals:

```
for target in [data]:
   print "target:",target
      if len(target)>0:
          dx        = np.mean(np.diff(time_pedacito))
          cros_cor = np.correlate(abs(signal),target,mode='full')
          pepe_cor = np.argmax(cros_cor)
          shift     = (pepe_correlaciona-len(target))*dx
```



Alignment for averaging:

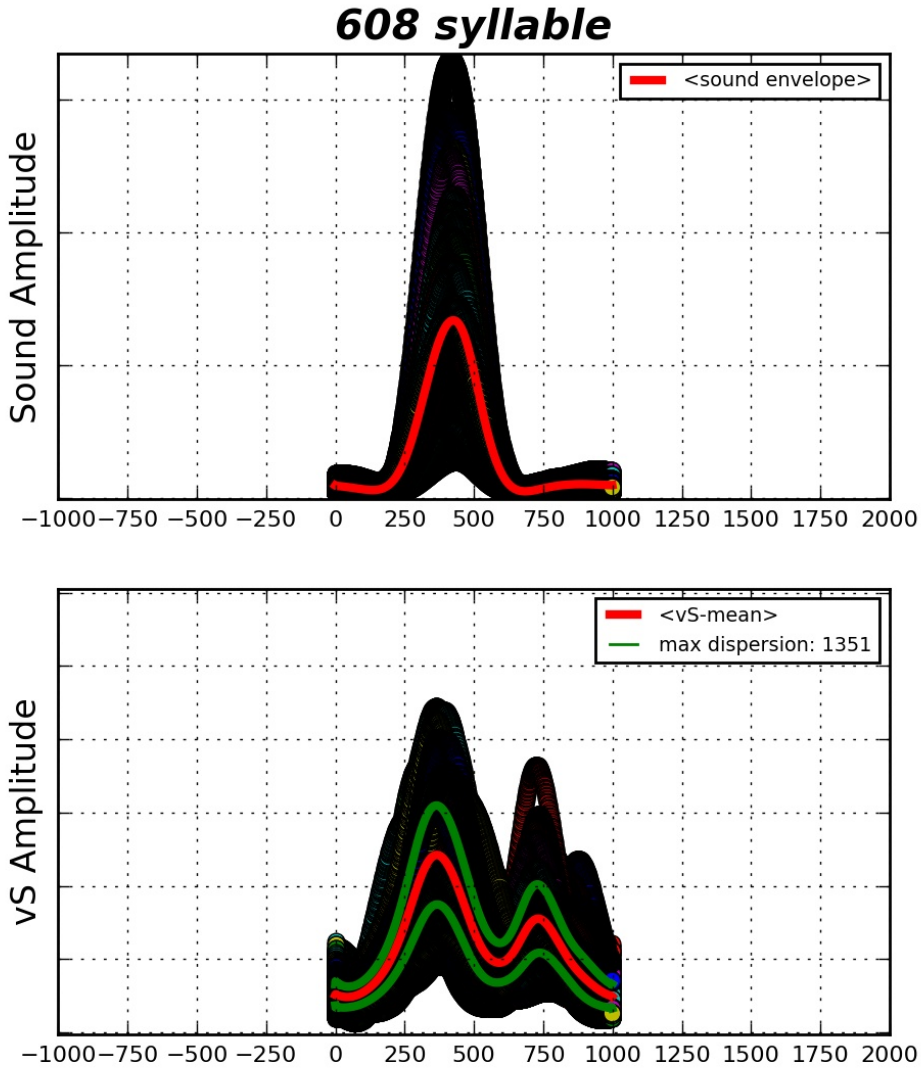
An example of multivariative time series



A way to align similar signals using correlation:procedure

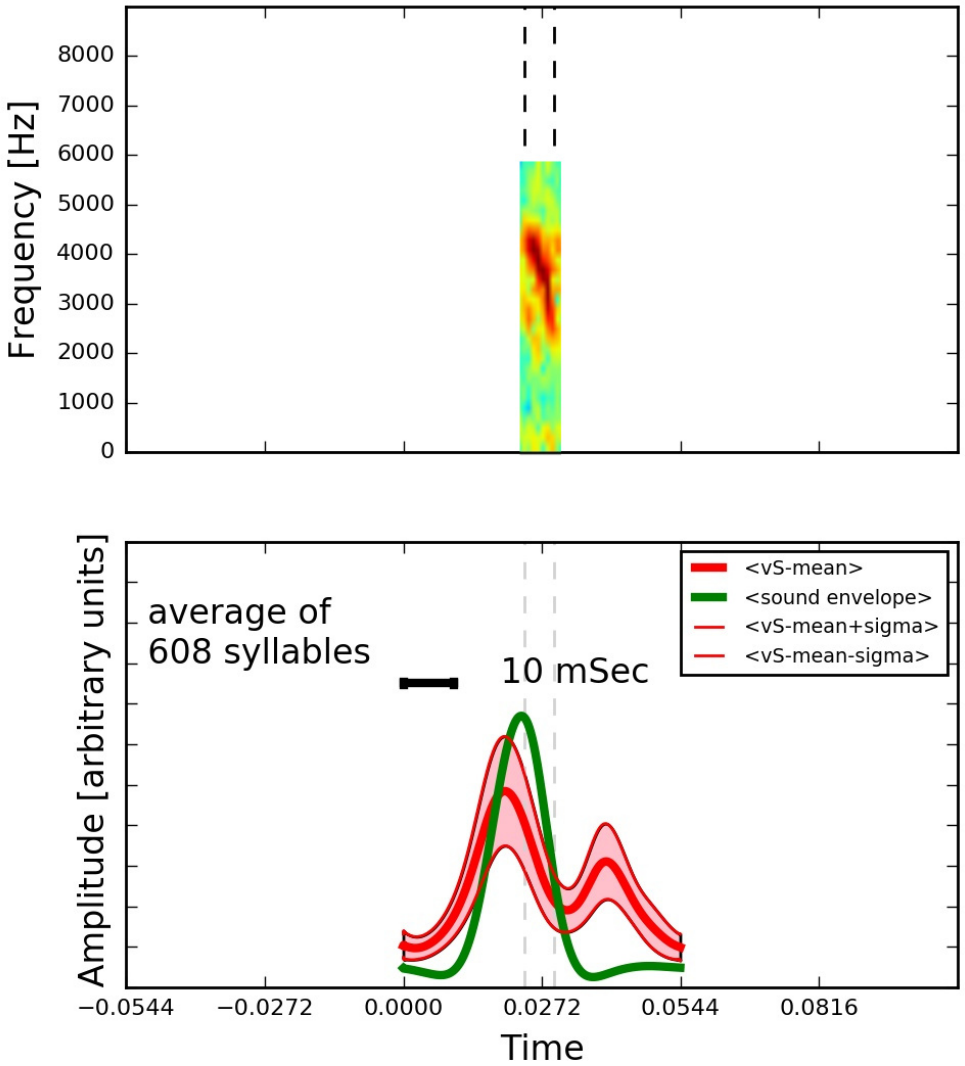
- Cut signal(each sound segment is automatically cut by using a burly envelope)
- True envelope is estimated.
- We rescale by re-sampling each segment to a fix number of points.
- To synchronize all syllable of multiple segments, we used fafeature of the sound envelope (In most cases the value maximum, or the minimum.)
- We take the average in each position.

A way to align similar signals using a signal feature



WTPC
Workshop en técnicas de programación científica

A way to align similar signals using correlation



Conclusions

- An heuristic and simple technique for envelope estimation is proposed and a open source code is provided.
- An method to obtain fundamental frequency with a simple method.
- Some ideas to segment and align time series are shown.
- A method to estimate an average pattern is proposed.

WTPC workshop en técnicas de programación científica

Our group: Thanks!!!

Pablo Alcain, Rodrigo Lugones, Graciela Molina y yo

Our Motivation

What is Different in the Scientific Software Development Process?

- Requirements often are not that well defined
- Floating-point math limitations and the chaotic nature of some solutions complicate validation
- An application may only be needed once
- Few scientists are programmers (or managers)
- Often projects are implemented by students (inexperienced in science and programming)
- Correctness of results is a primary concern, less so the quality of the implementation

WTPC workshop en técnicas de programación científica

Our work

2015/ABR. ICTP-SAIFR

2015/NOV. SMN
(Curso Intensivo y personalizado)

2016/MAR. WTPC16 - UBA

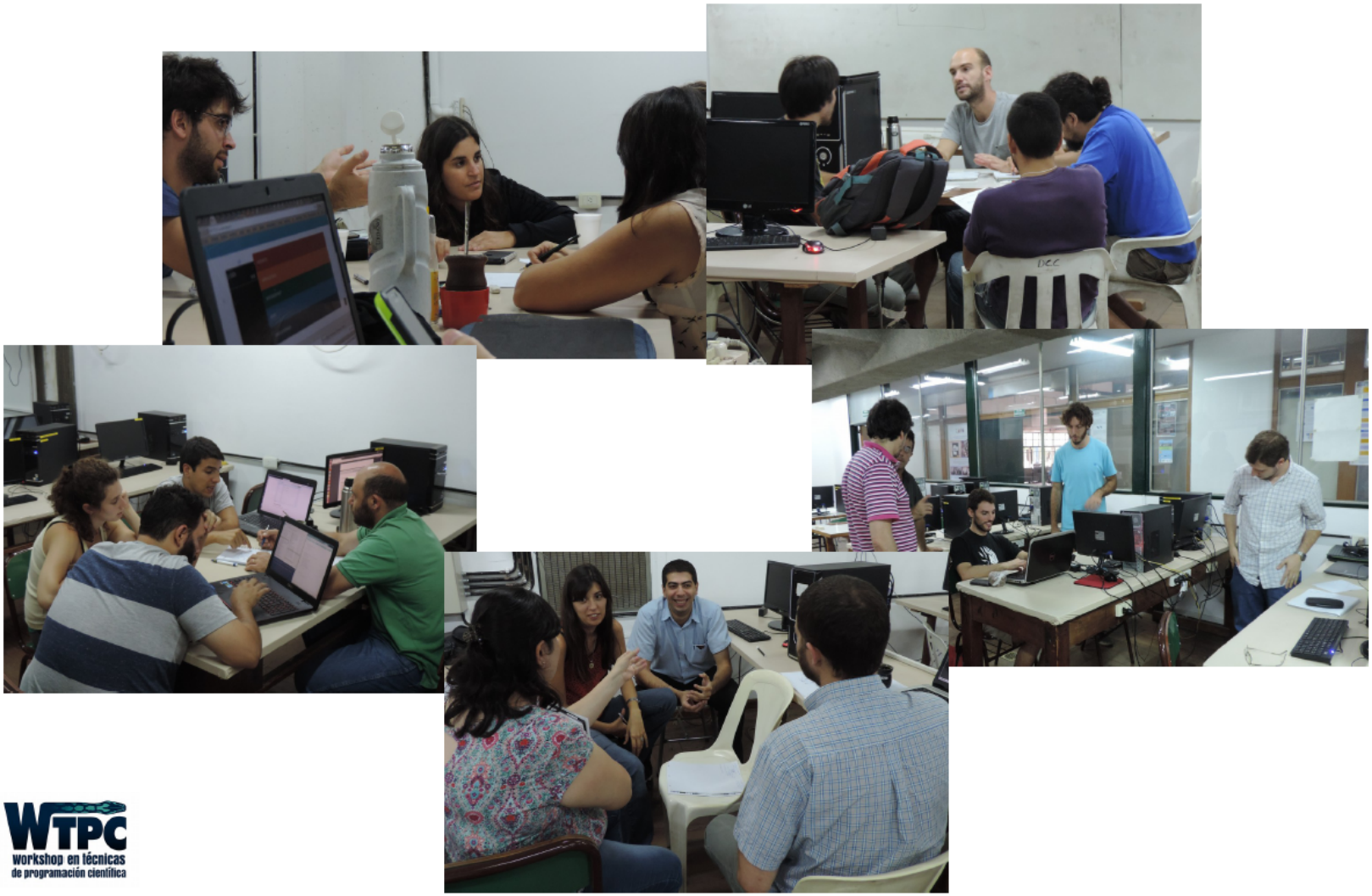
(~ 100 aspirantes)

2016/NOV.
CURSO POSTGRADO UNQ

2017/MAR. WTPC17 - UNT

(~ 60 aspirantes)

More about our works

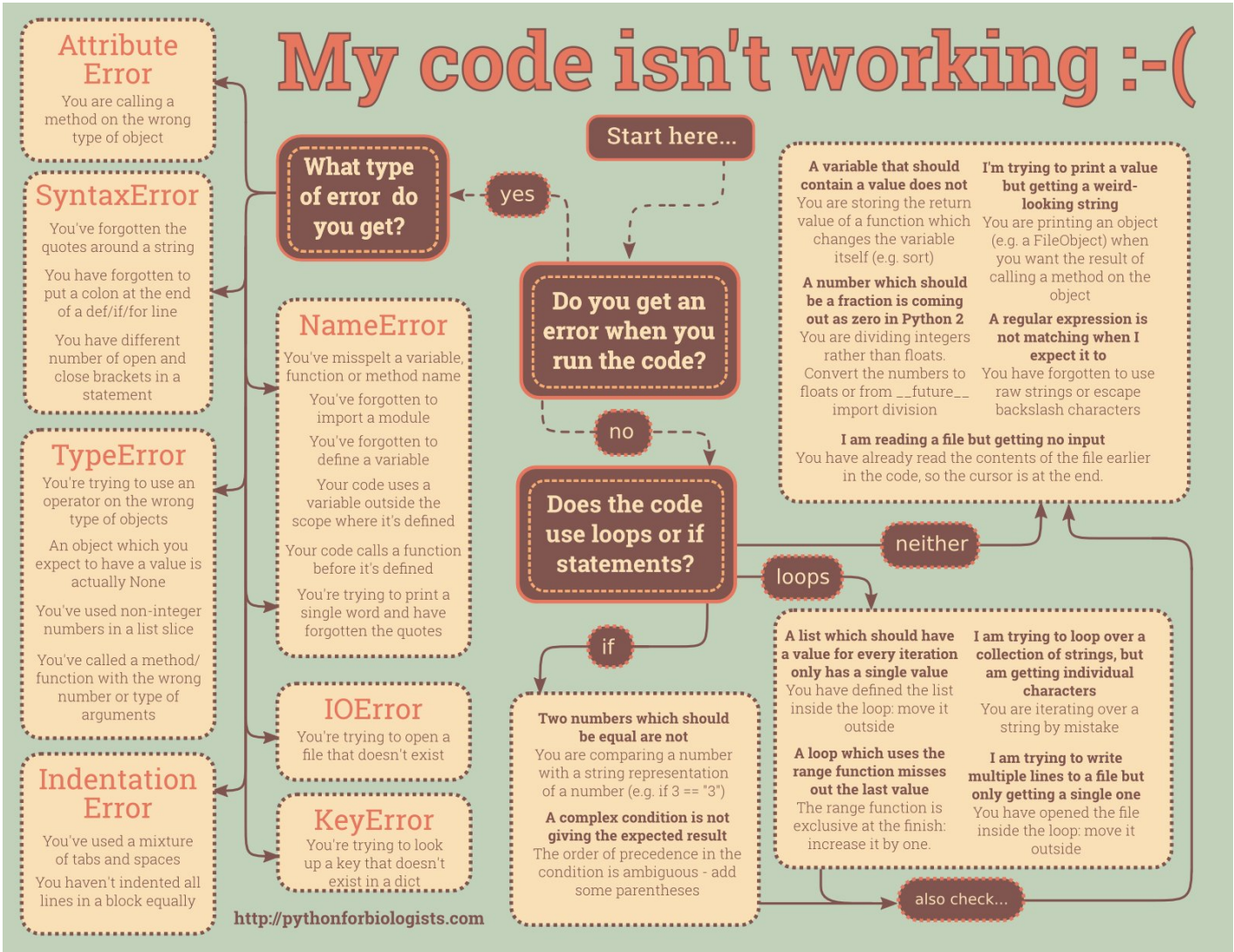


Invitación

Tercer workshop de tecnicas de programación científica:
https://wtpc.github.io/
https://twitter.com/workshopTPC

WTPC
workshop en técnicas de programación científica

Extra



WTPC
Workshop en técnicas de programación científica