# Segmentación de Señales de Audio usando Correlaciones de Largo Alcance

**Pedro Pury**

GTMC - Agosto 10, 2010

# Segmentación de Señales de Audio usando Correlaciones de Largo Alcance


**Guillermo J. López**
Facultad de Ingeniería
Instituto Universitario Aeronáutico


Pedro A. Pury
Facultad de Matemática, Astronomía y Física
Universidad Nacional de Córdoba

# DSP

Digital Signal Processesing: is often implemented using specialised microprocessors. These often process data using fixed-point arithmetic, although some versions are available which use floating point arithmetic. Also, multicore implementations of DSPs have started to emerge.
The main applications (real–time) of DSP are audio signal processing, (ADC, DAC), audio compression, digital image processing, video compression, speech processing, speech recognition, digital communications, etc.

# Objective

Increasing power for DSP $\longrightarrow$ need for algorithms.

Segmentation of audio signals: our final objective is to address the detection of voice and music from an audio source in streaming or "at flight".

# Auto–Correlation Function

- $\{x_i\,, i = 1, \ldots, N\}$ $N$ equidistant measurements

$$\langle x \rangle = \frac{1}{N} \sum_{i=1}^{N} x_i \qquad y_i = x_i - \langle x \rangle$$

$$C(s) = \frac{1}{N-s} \sum_{i=1}^{N-s} y_i \, y_{i+s}$$

- $$C(s) \sim \begin{cases} \exp(-s/\tau) & \text{short–range} \\ s^{-\gamma} \quad (0 < \gamma < 1) & \text{long–range} \end{cases}$$

# Detection of Long–range Correlations

- Direct calculation: ln $C(s)$ vs $s$
- Power spectrum density $S(f)$ of the original (non-integrated) signal is simply the Fourier transform of the autocorrelation function, $S(f) \sim 1/f^{\beta}$: $\gamma = 1 - \beta$
- Problems with $C(s)$ and $S(f)$
  - data with noise superimposed
  - data are affected by non-stationarities
  - trends in data induce artificial correlations

# Detrended Fluctuation Analysis (DFA)

- Introduced by Peng C-K, Buldyrev SV, Havlin S, Simons M, Stanley HE, and Goldberger AL., *Mosaic organization of DNA nucleotides*, Phys. Rev. E **49**, 1685 (1994).

# DFA steps

- Creation of profile: $Y(i) = \sum_{k=1}^{i} y_k$
- Partition of profile into $N_s = N/s$ non-overlapping segments of equal length $s$.
- Detrended profile for each segment $\nu = 1, \ldots, N_s$:

$$Y_s(i) = Y(i) - p_\nu(i)\,; \qquad i = (\nu - 1)s + j\,, \;\; j = 1, \ldots, s$$

- $p_\nu(i)$ least-squares polynomial fit of data with order $n$ (DFA$n$)
- DFA1: linear fit
  DFA2: quadratic fit
  $\vdots$

# Second–Order Fitting



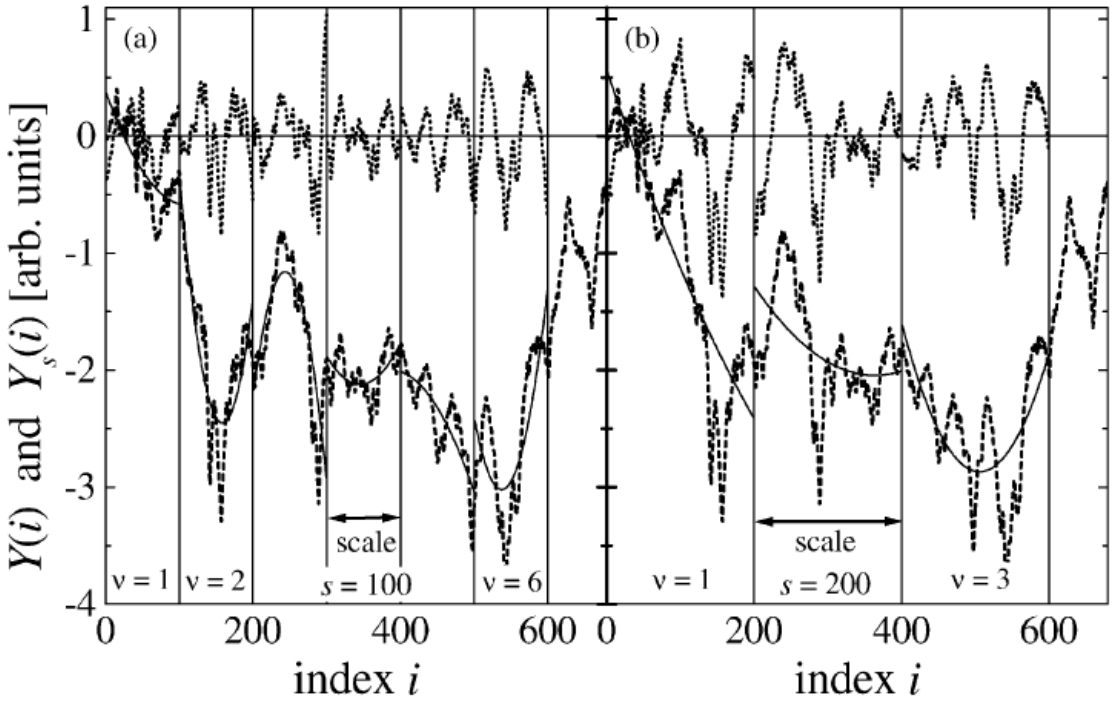



Fig. 1. Illustration of the detrending procedure in the detrended fluctuation analysis. For two segment lengths (time scales) $s = 100$ (a) and 200 (b), the profiles $Y(i)$ (dashed lines; defined in Eq. (3)), least squares quadratic fits to the profiles (solid lines), and the detrended profiles $Y_s(i)$ (dotted lines) are shown versus the index $i$.

# DFA steps

- Variance

$$F^2(\nu) = \frac{1}{s}\sum_{j=1}^{s} Y_s^2[(\nu-1)s+j]$$

- DFA fluctuation function

$$F^{(n)}(s) = \sqrt{\frac{1}{N_s}\sum_{\nu=1}^{N_s} F^2(\nu)}$$

# DFA scaling

- $F^{(n)}(s) \sim s^{\alpha}$, for $s >> 1$
- $1/2 < \alpha < 1$, persistent long–correlations,
  - $\gamma = 2(1-\alpha)$,
  - $\beta = 2\,\alpha - 1$.
- $\alpha = 1/2$, white noise.
- $0 < \alpha < 1/2$, anti–persistency.
- $\alpha > 1$, correlations exist but cease to be of a power-law form.
- $\alpha = 1{,}5$, brown noise (integration of white noise).

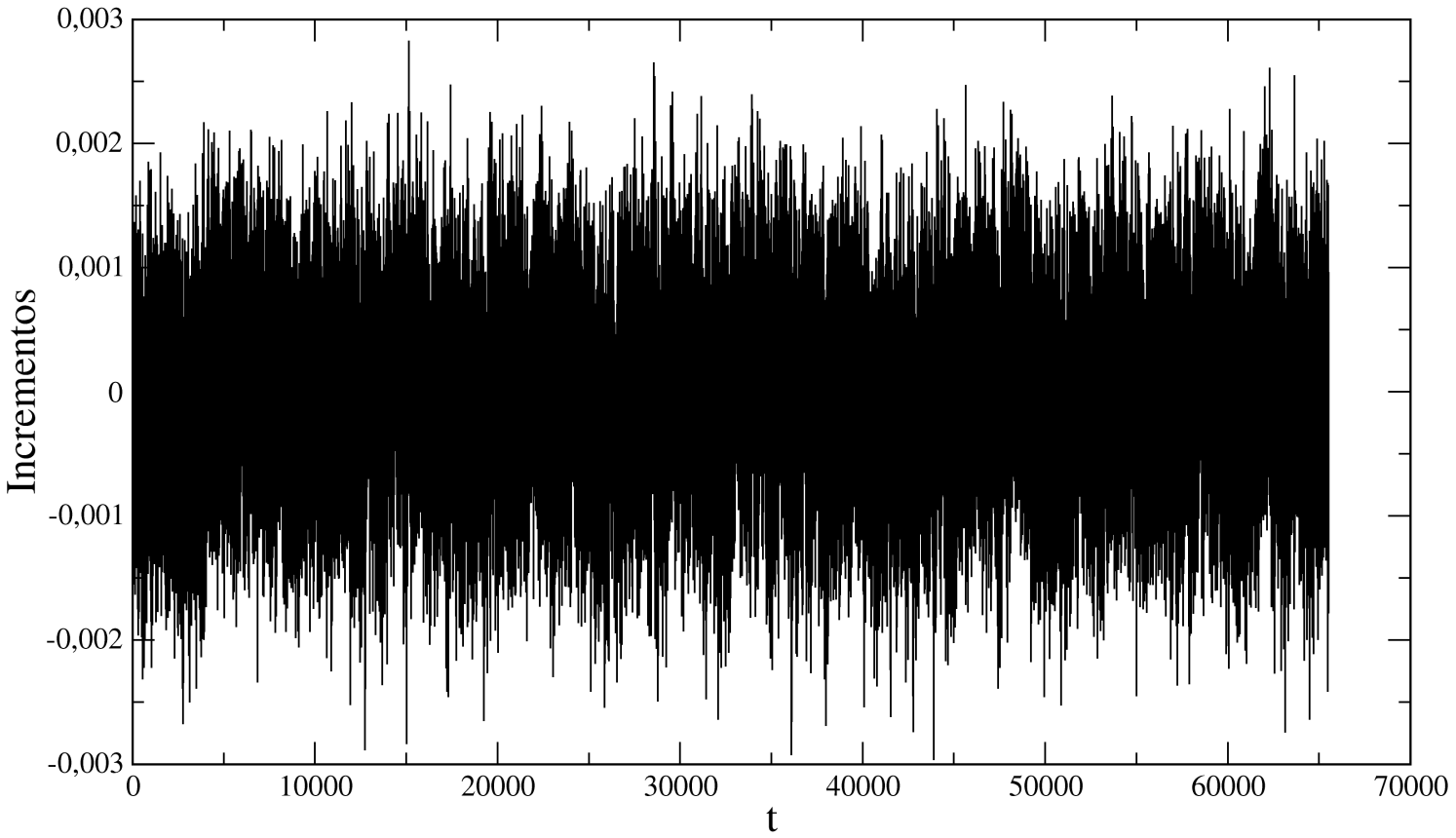
# Generation of Long Correlations

- Fractional Brownian Motion (FBM)
  (Mandelbrot and Wallis, 1969)
- Successive Random Addition (Richard Voss, 1985)

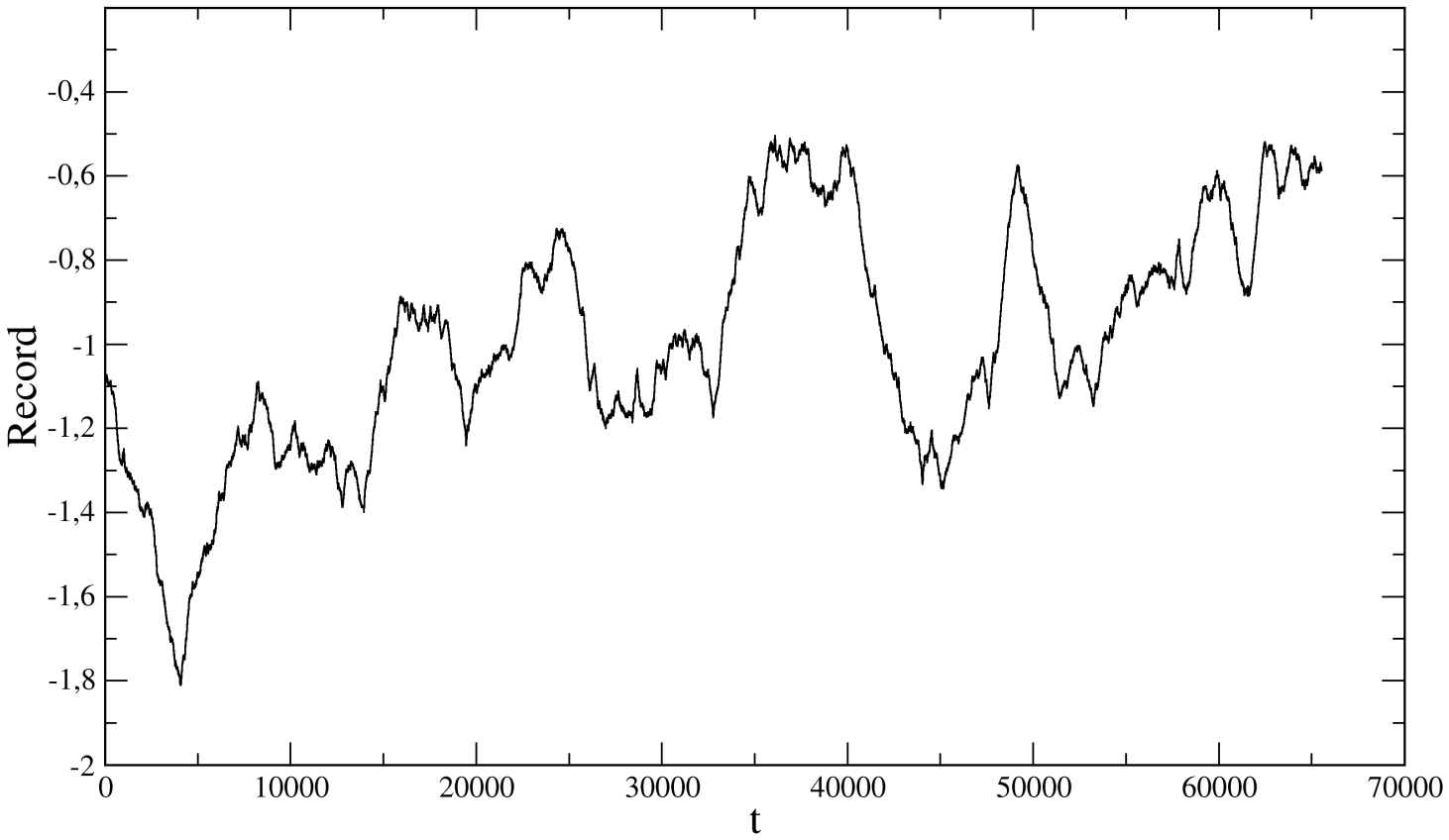
# Voss' Algorithm

- Step 0: $x_j = 0\,,\ \forall j = 0, \ldots, 2^N$
- Step 1: $x_j = \texttt{GAUSS}(0, \sigma_1 = 1)\,,\ j = 0, 2^{N-1}, 2^N$ and midpoints by interpolation
- Step 2: $x_j = x_j + \texttt{GAUSS}(0, \sigma_2 = 2^{-\alpha}\,\sigma_1)\,,\ j = (0, 1, 2, 3, 4) \times 2^{N-2}$ and midpoints by interpolation
- Step n: $x_j = x_j + \texttt{GAUSS}(0, \sigma_n = 2^{-\alpha}\,\sigma_{n-1})\,,\ j = (0, \ldots, 2^n) \times 2^{N-n}$ and midpoints by interpolation
- Step N: $x_j = x_j + \texttt{GAUSS}(0, \sigma_N = 2^{-\alpha}\,\sigma_{N-1})\,,\ j = 0, \ldots, 2^N$

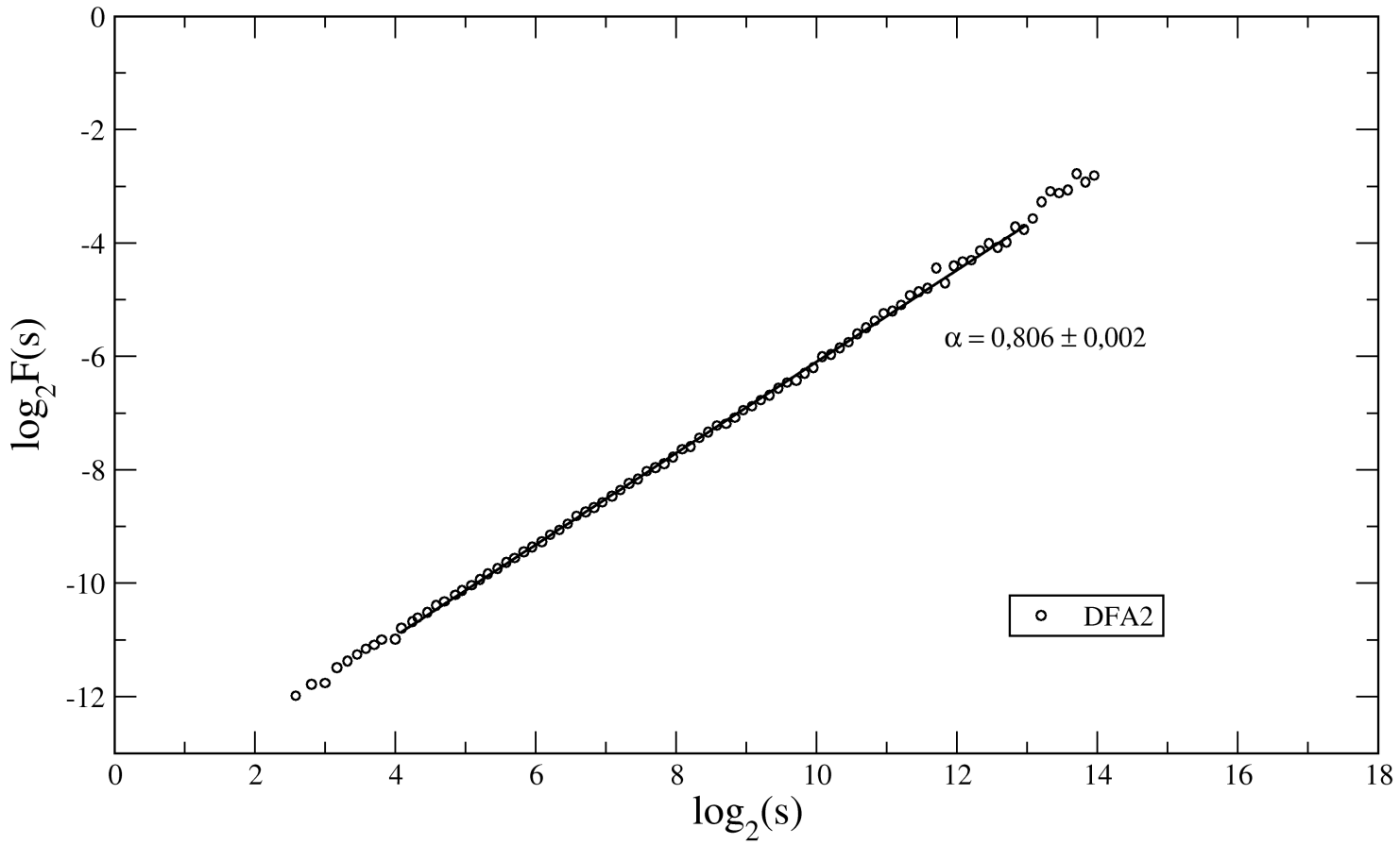
# Increments $N = 16, \alpha = 0{,}8$



# Record $N = 16, \alpha = 0{,}8$



# DFA analysis



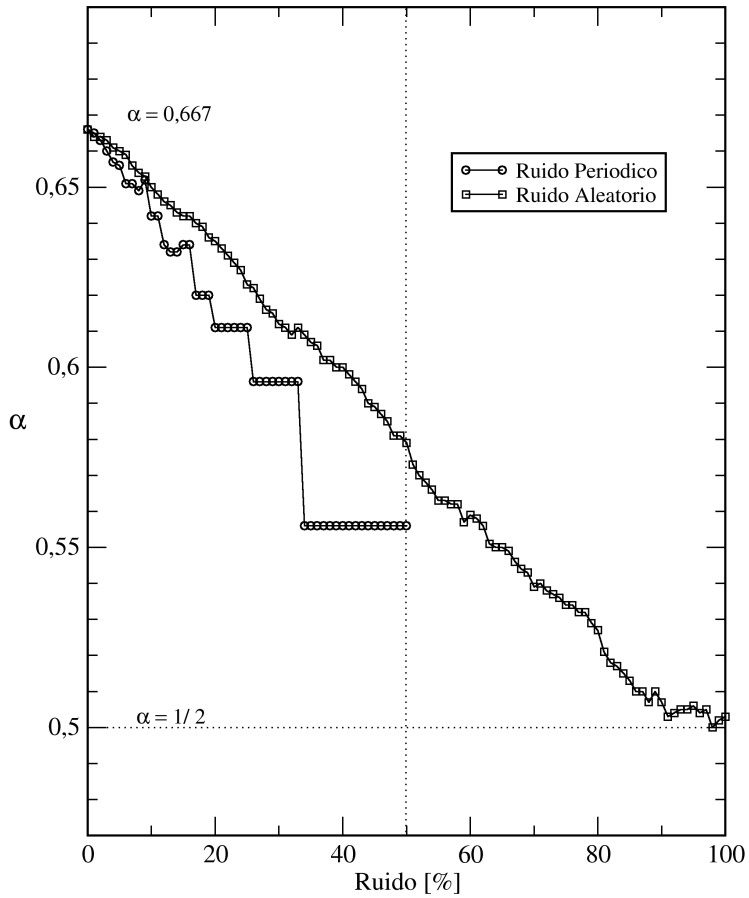
# Fractal Landscapes

MojoWorld (www.pandromeda.com)

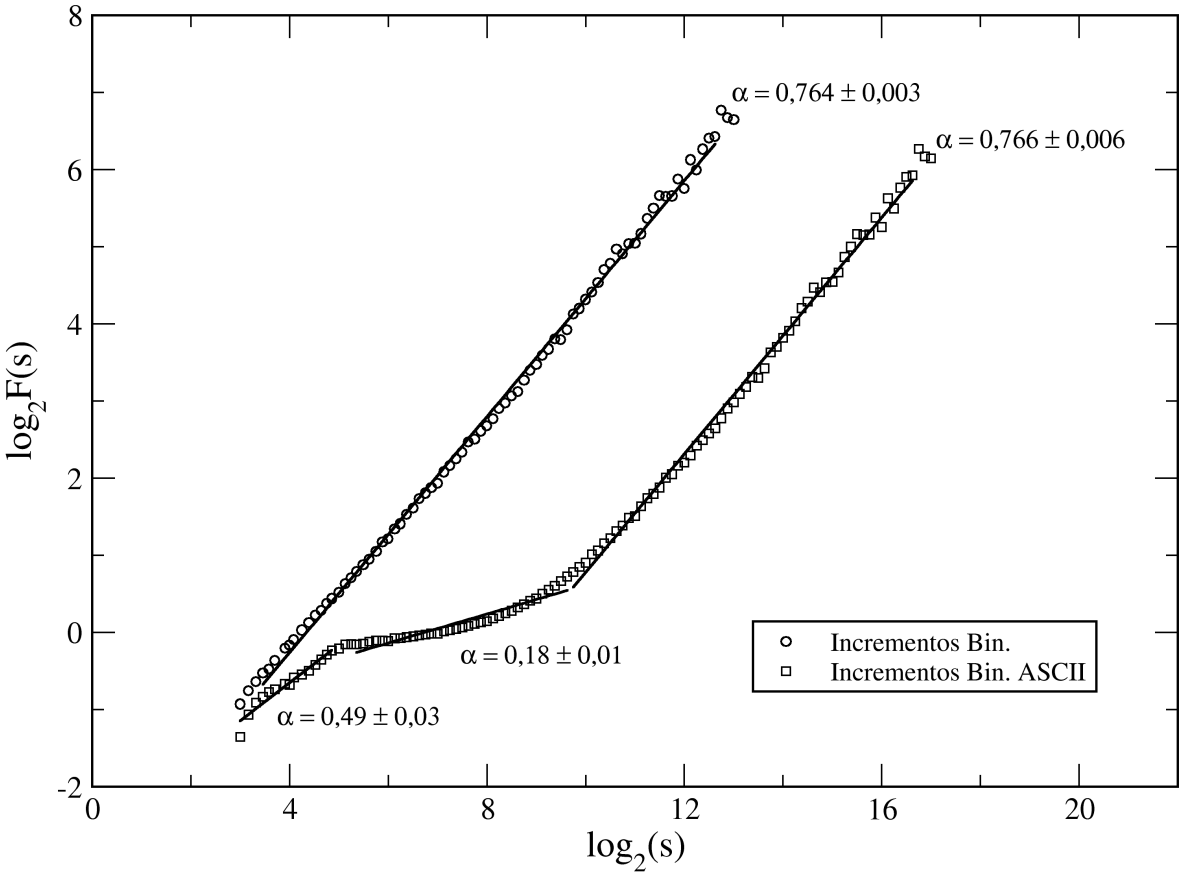
# DFA3, $\alpha = 0{,}7$

| N | Voss Generator | | Digital Generator | | Permutation | |
|---|---|---|---|---|---|---|
| | $\alpha$ | error | $\alpha$ | error | $\alpha$ | error |
| 15 | 0.686 | 0.005 | 0.648 | 0.005 | 0.518 | 0.005 |
| 14 | .684 | .005 | 652 | 005 | .508 | .006 |
| 13 | .684 | .006 | 669 | 005 | .501 | .008 |
| 12 | .685 | .008 | 648 | 007 | .560 | .007 |
| 11 | .687 | .010 | 628 | 009 | .534 | .010 |
| 10 | .696 | .013 | 642 | 014 | .582 | .013 |
| 9 | .706 | .019 | 661 | 015 | .584 | .016 |
| 8 | .707 | .022 | 700 | 023 | .505 | .019 |
| 7 | .722 | .035 | 606 | 027 | .490 | .028 |
| 6 | .748 | .042 | 625 | 040 | .669 | .034 |

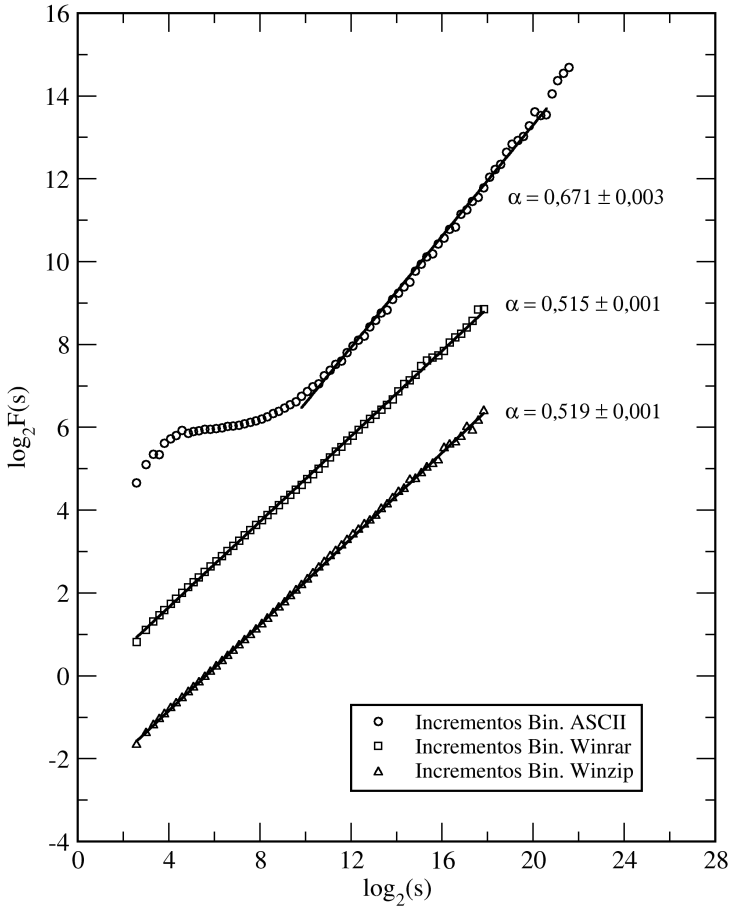
# DFA3, $\alpha = 0{,}7$



# DFA2, $\alpha = 0{,}8$



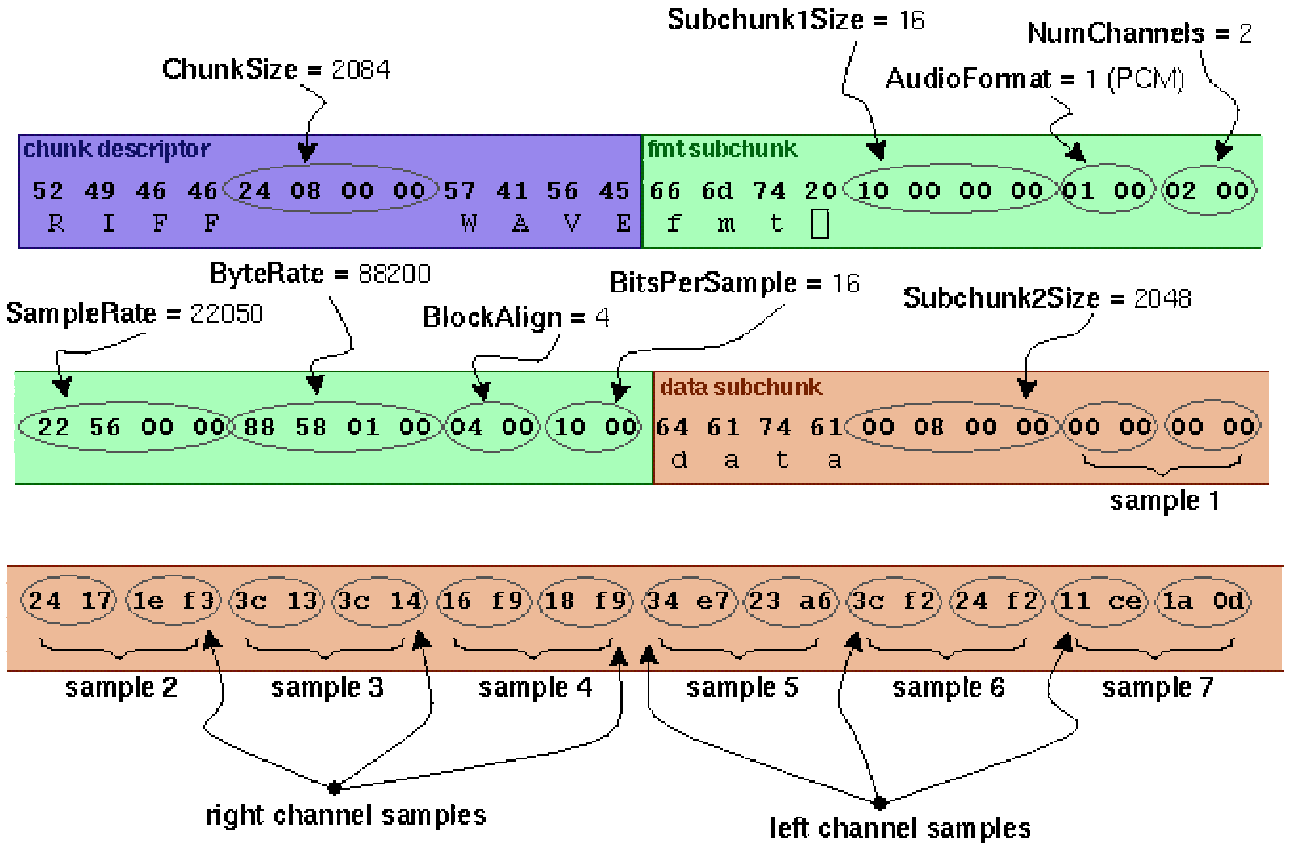
# DFA2, $\alpha = 0{,}7$



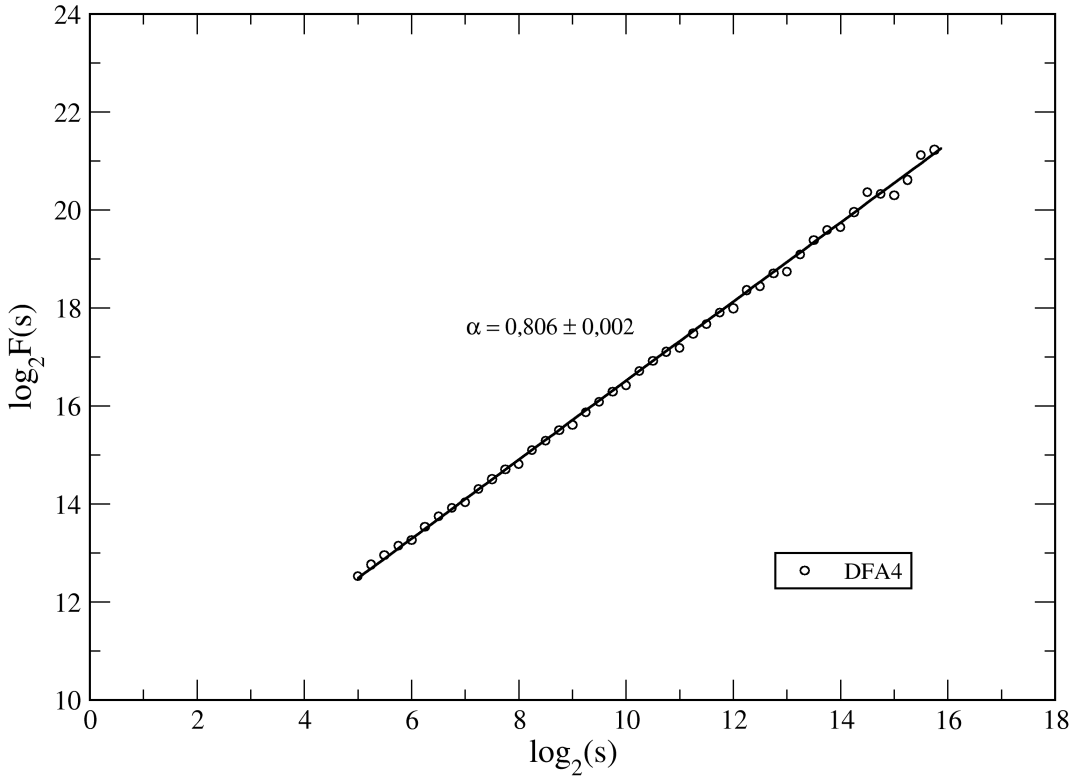
# Wave Format

- The WAVE file format is a subset of Microsoft's RIFF specification for the storage of multimedia files.
- A RIFF file starts out with a file header followed by a sequence of data chunks.
- The default byte ordering assumed for WAVE data files is little-endian.
- Common WAV format contains uncompressed audio in the linear pulse code modulation (LPCM) format.
- The standard audio file format for CDs, for example, is LPCM-encoded, containing two channels of 44,100 samples per second, 16 bits per sample. That is, $2^{16} = 65536$ levels for digitalization.

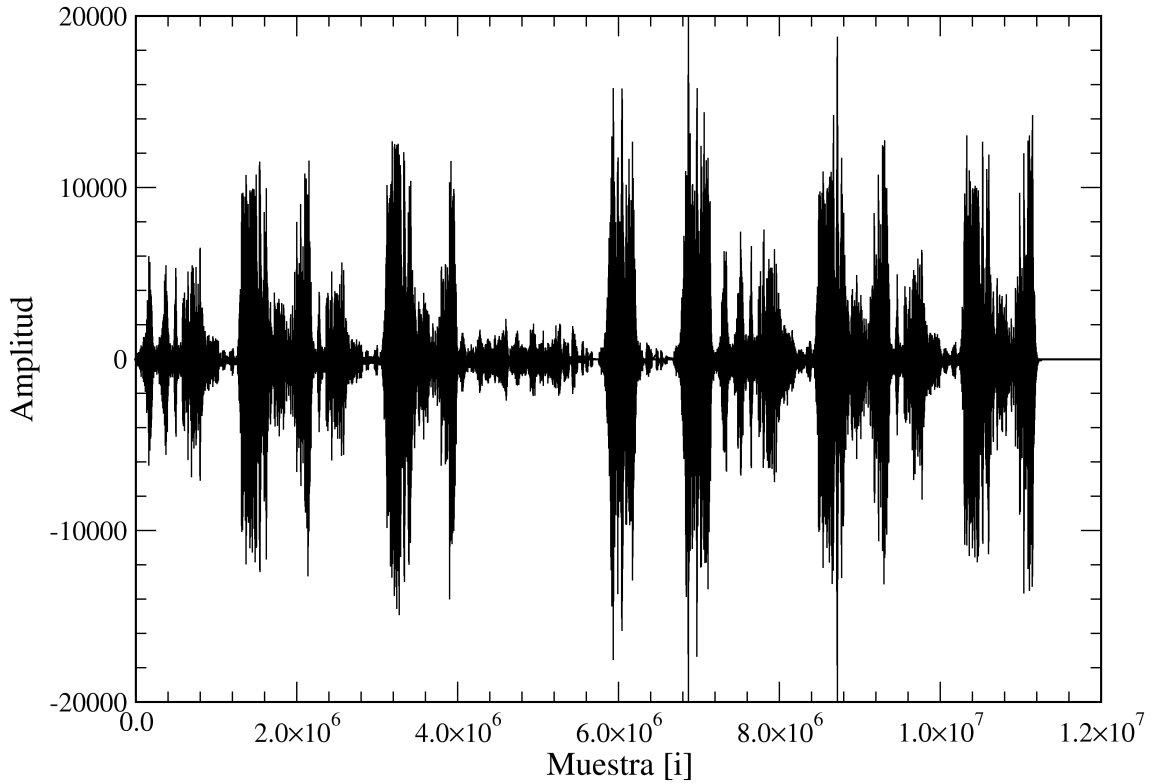
# Wave Soundfile



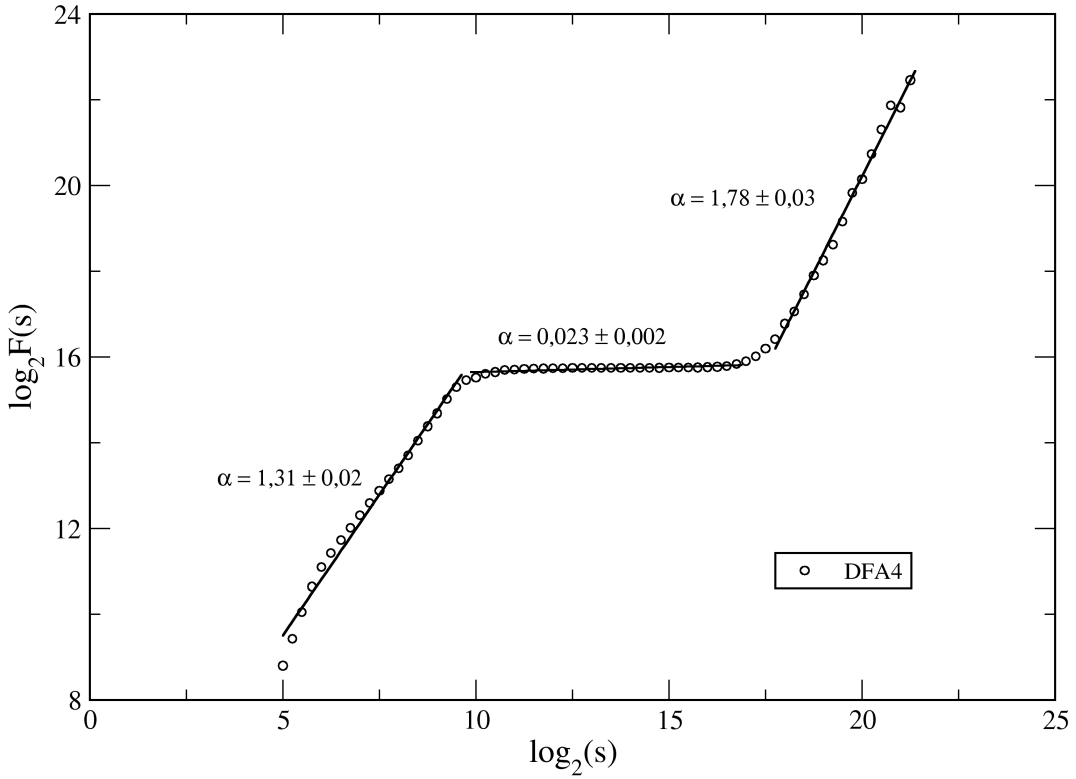
# Voss in Wav, 6s



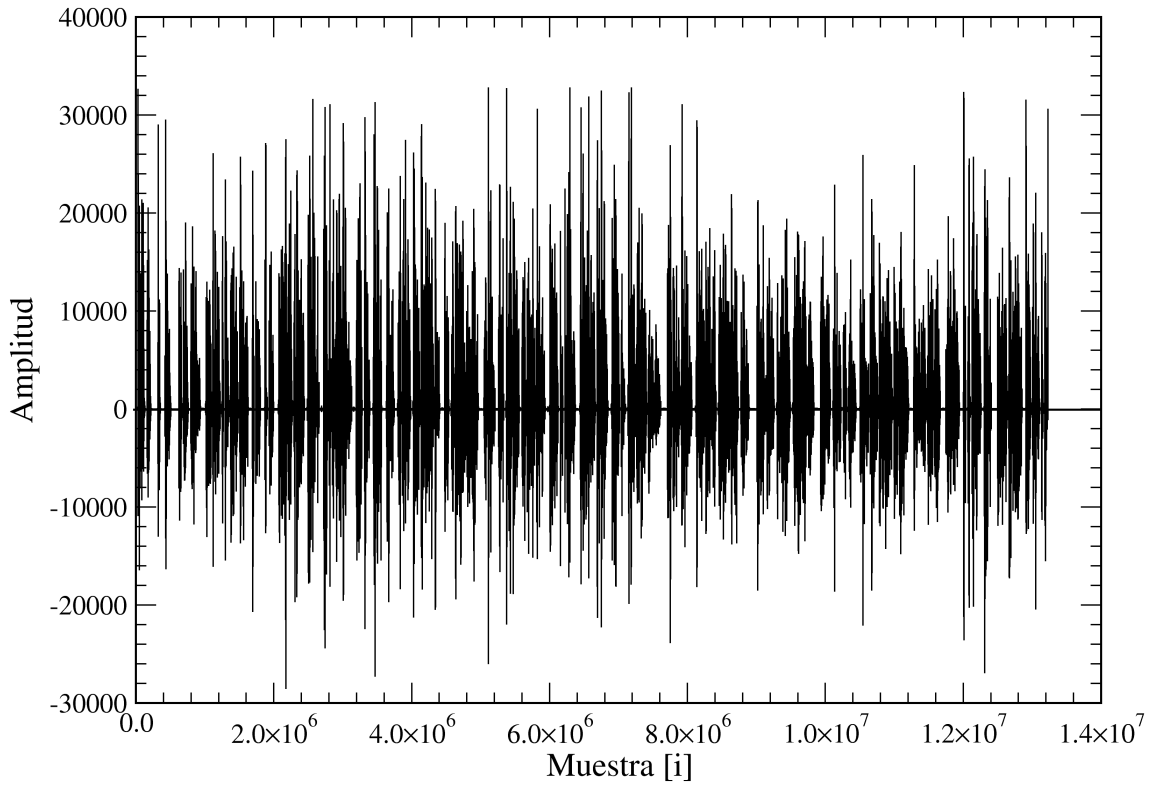
# Beethoven, 1st Symphony, 3rd mov



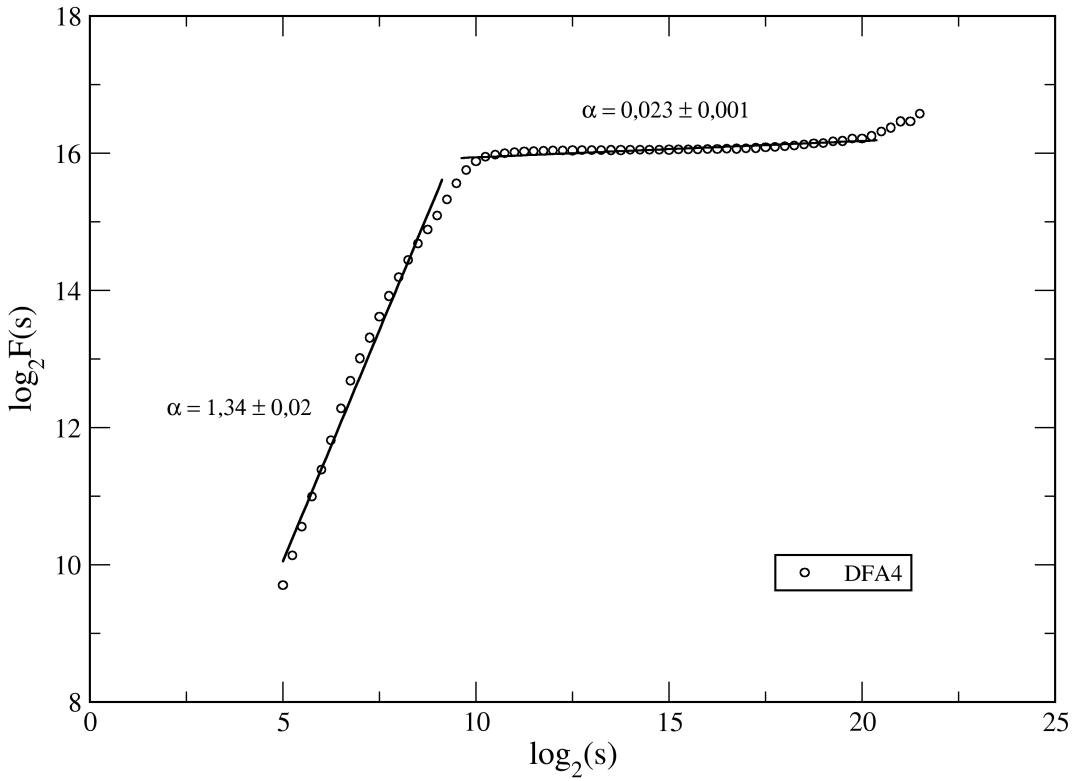
# Beethoven, 1st Symphony, 3rd mov



# Harry Potter and The Deathly Hallows



# Harry Potter and The Deathly Hallows

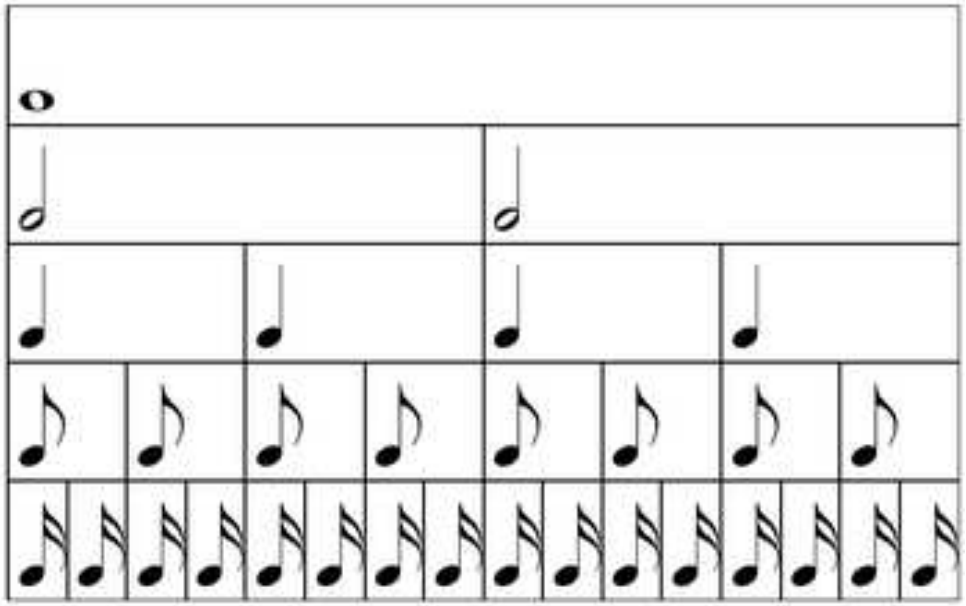


# Tempo

- Tempo (Bpm or quarter per minute)
- Tempo – Rates:

| | | | |
|---|---|---|---|
| Lento | – | (very slow) | (40–60 bpm) |
| Adagio | – | (slow and stately) | (66–76 bpm) |
| Andante | – | (at a walking pace) | (76–108 bpm) |
| Allegro | – | (fast and bright) | (120–168 bpm) |
| Presto | – | (very fast) | (168–200 bpm) |

- Note value



- Presto: $\sim 180 \times 4$ sixteenth notes per minute. Duration $1/12 \approx 0{,}08\,\mathrm{s}$.

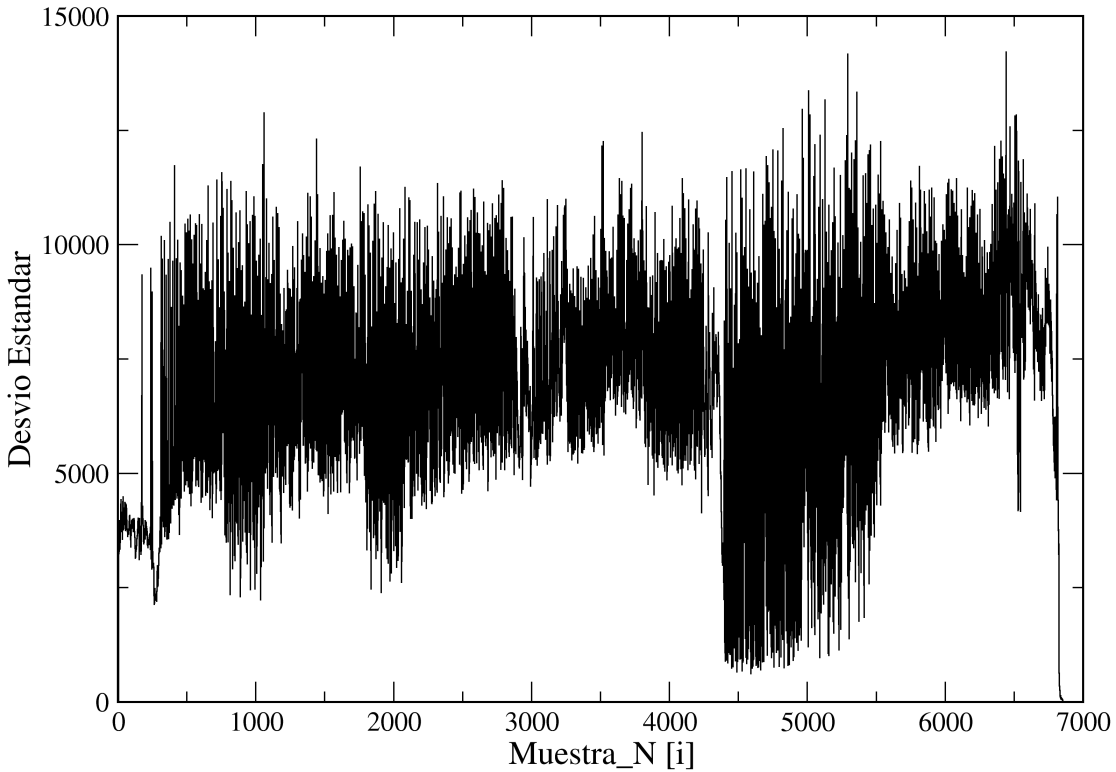
# IOI

- IOI (inter-onset interval)
  Time between the beginnings or attack-points of successive events or notes.
- IOI values larger than approx. 1500 ms are hardly usable.
- IOI 100-126 ms limit whence we start to lose control over the isochronality of pulsations.
- Minimum duration $1/10 \approx 0{,}10\,\mathrm{s}$.
- Sample duration $\sim 2^{-5}\,\mathrm{s}$.

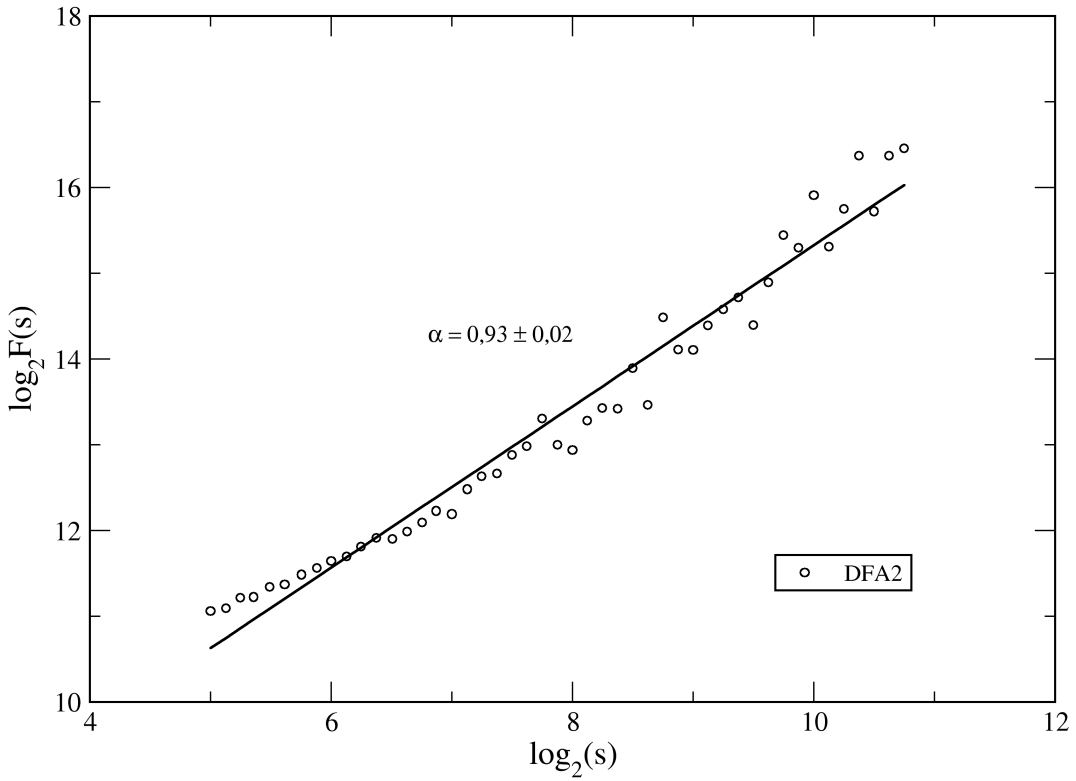
# Variance

- Unit increment: variance of 2048 samples.
- At 44100 sample/s approx. 21,5 increments/s.
- Increment duration $\approx 0{,}0464$.
- Jenning *et.al*, Physica A **336**, 585 (2004).

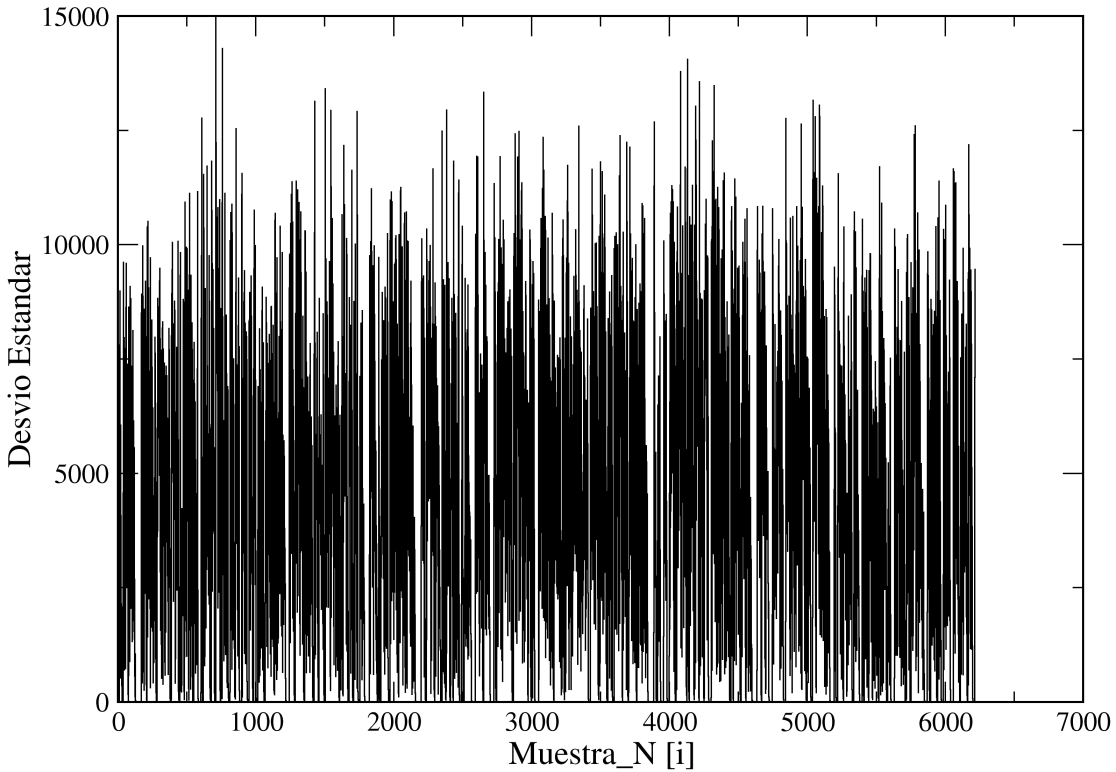
# AC–DC, Shoot to Thrill



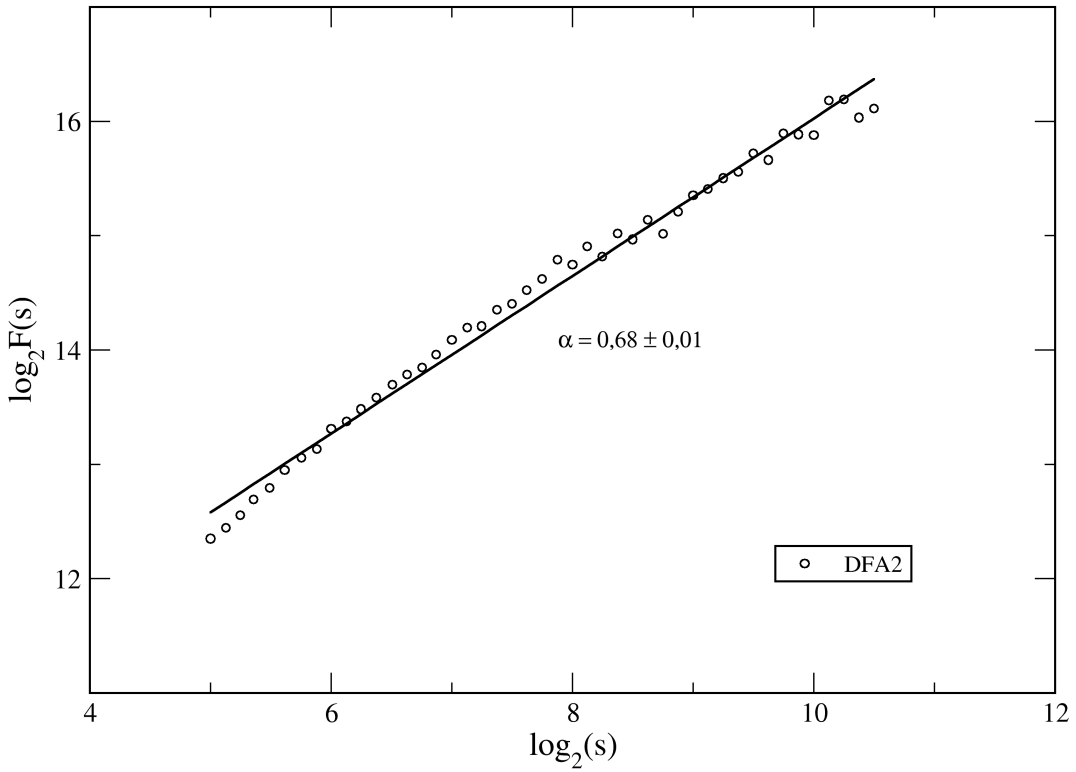
# AC–DC, Shoot to Thrill



# Don Quijote de la Mancha, 2nd. chap.



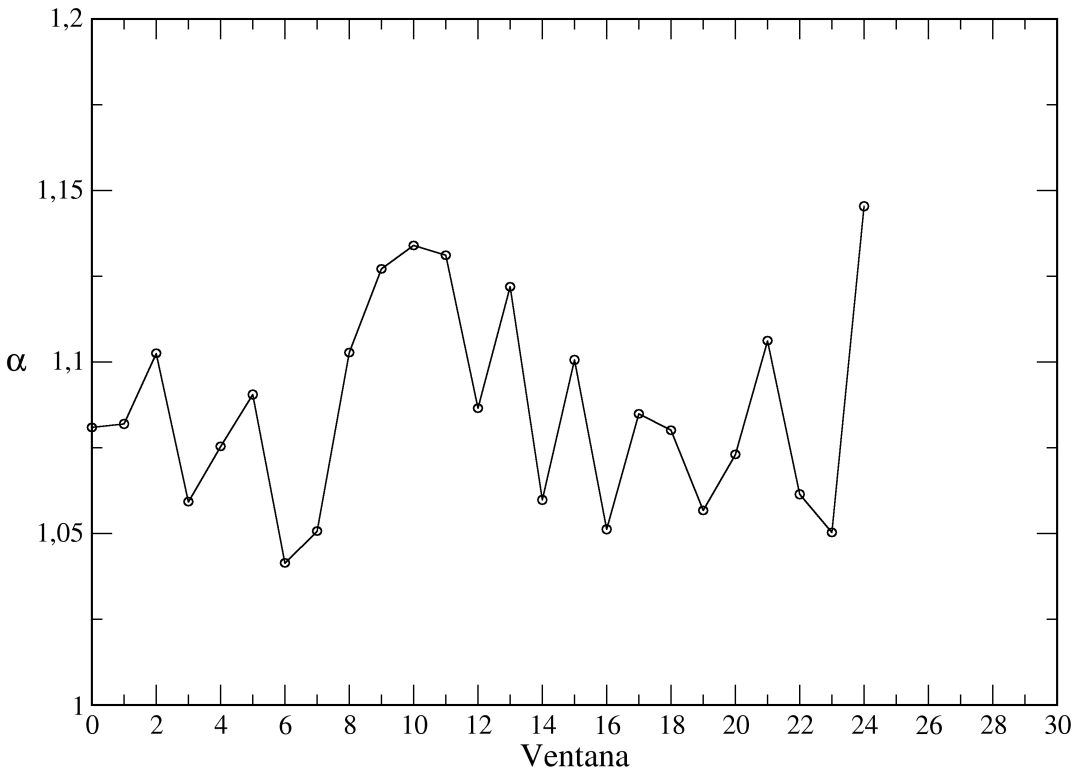
# Don Quijote de la Mancha, 2nd. chap.



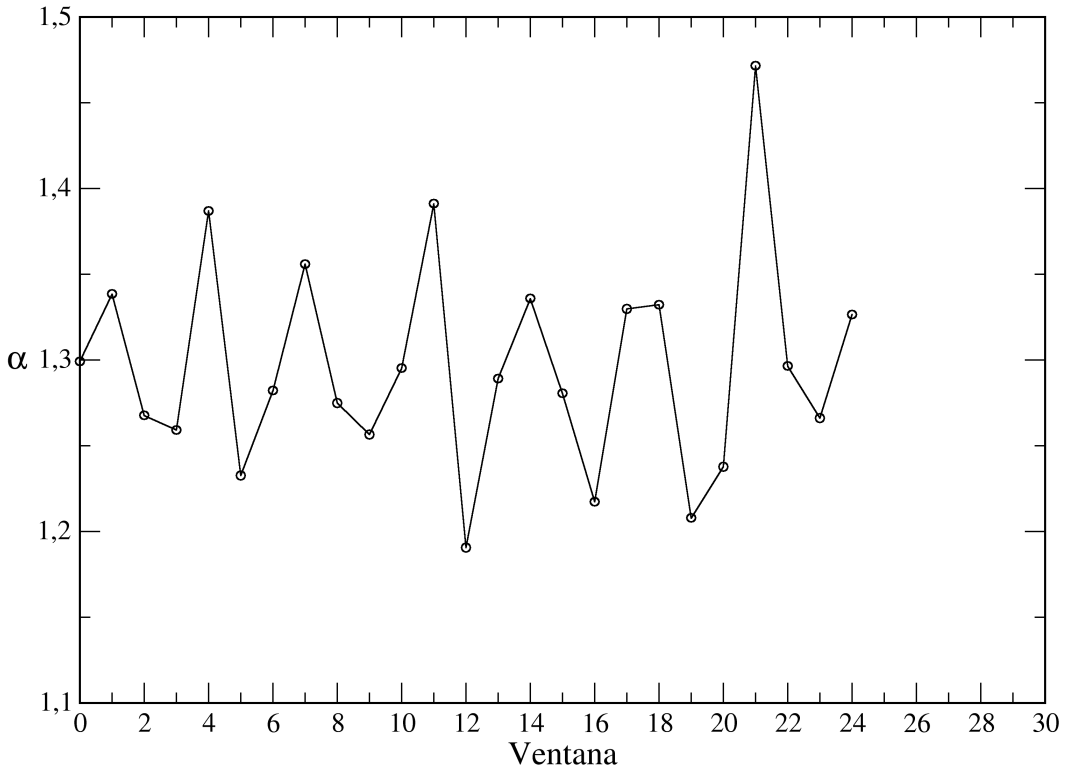
# Optimized Parameters

- Unit increment: variance of 2048 samples.
- Windows of 256 increments
- Window's duration $256 \times 2048 = 524288\,\text{samples} = 11,88s$
- DFA4

# Pink Floyd, Have a Cigar



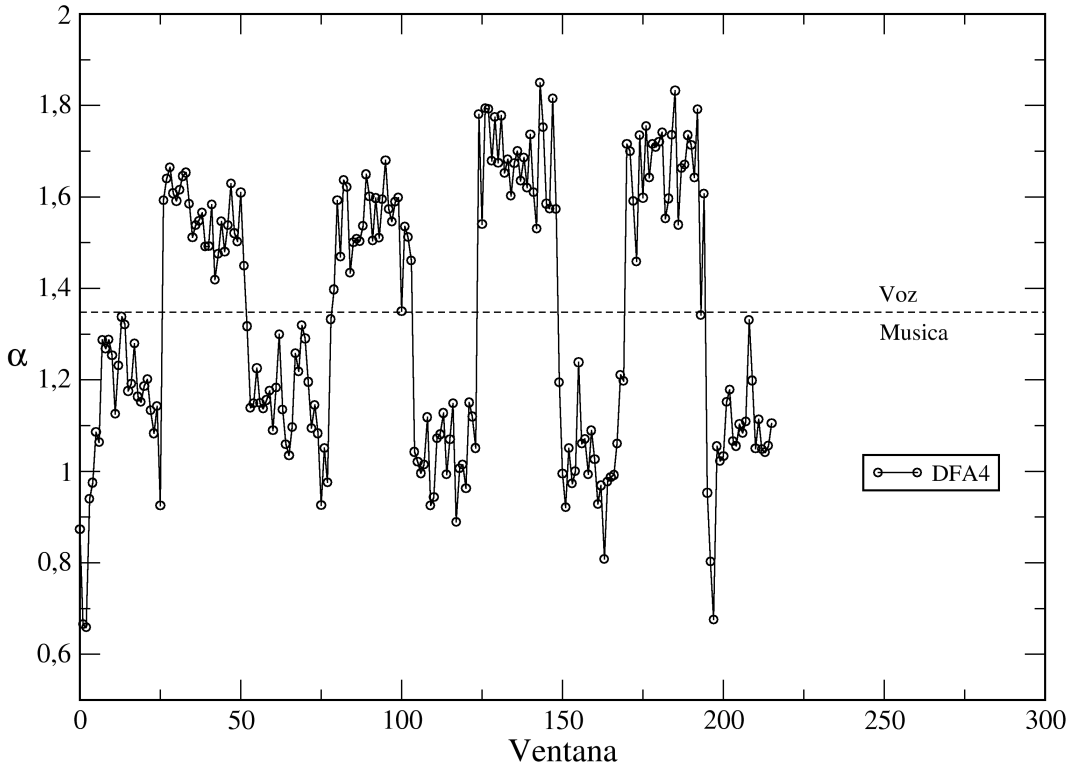
# Harry Potter and The Deathly Hallows



# Difference between Music and Voice

- Music $\alpha \in (0,6; 1,5)$
- Voice $\alpha \in (1,1; 1,9)$

# Alternate fragments



# Radio Podcast – Effectiveness 88 %