

Ingeniería de
Sonido

**V Jornadas de Acústica,
Audio y Sonido**

UNTREF
UNIVERSIDAD NACIONAL
DE TRES DE FEBRERO



**Workshop de Python
Procesamiento de Señales en Python**

Facultad de Ingeniería del Ejército - Infiniem Labs Acústica

29 y 30 de agosto de 2019

Chungo, F., Groisman A., y Yommi, M.
infiniemplab.dsp@gmail.com



LO MÁS IMPORTANTE

GitHub

[/infiniemplabs-acustica](#)



infiniemplab.dsp@gmail.com



infiniemacustica.com

AUDIO FEATURES

UNTREF

UNIVERSIDAD NACIONAL
DE TRES DE FEBRERO

La extracción de **características relevantes** de audio originales es una etapa importante en el **análisis de audio**. En general, es un paso previo en el reconocimiento de patrones y machine learning.

También se puede ver como una **reducción de volumen** de datos, porque queremos que nuestros algoritmos de análisis se basen en un número relativamente pequeño de características.

PARA ESTO ES IMPORTANTE TENER UN BUEN CONOCIMIENTO DEL DOMINIO DE APLICACIÓN, PARA SABER QUÉ CARACTERÍSTICAS SON MÁS REPRESENTATIVAS.

EJEMPLOS

UNTREF

UNIVERSIDAD NACIONAL
DE TRES DE FEBRERO



<https://www.youtube.com/watch?v=d-JMtVLUSEg>

EJEMPLOS

UNTREF

UNIVERSIDAD NACIONAL
DE TRES DE FEBRERO



**CONCIENCIA
SONORA**

GitHub

<https://github.com/infiniemlabs-acustica/conciencia-sonora>

EJEMPLOS

UNTREF

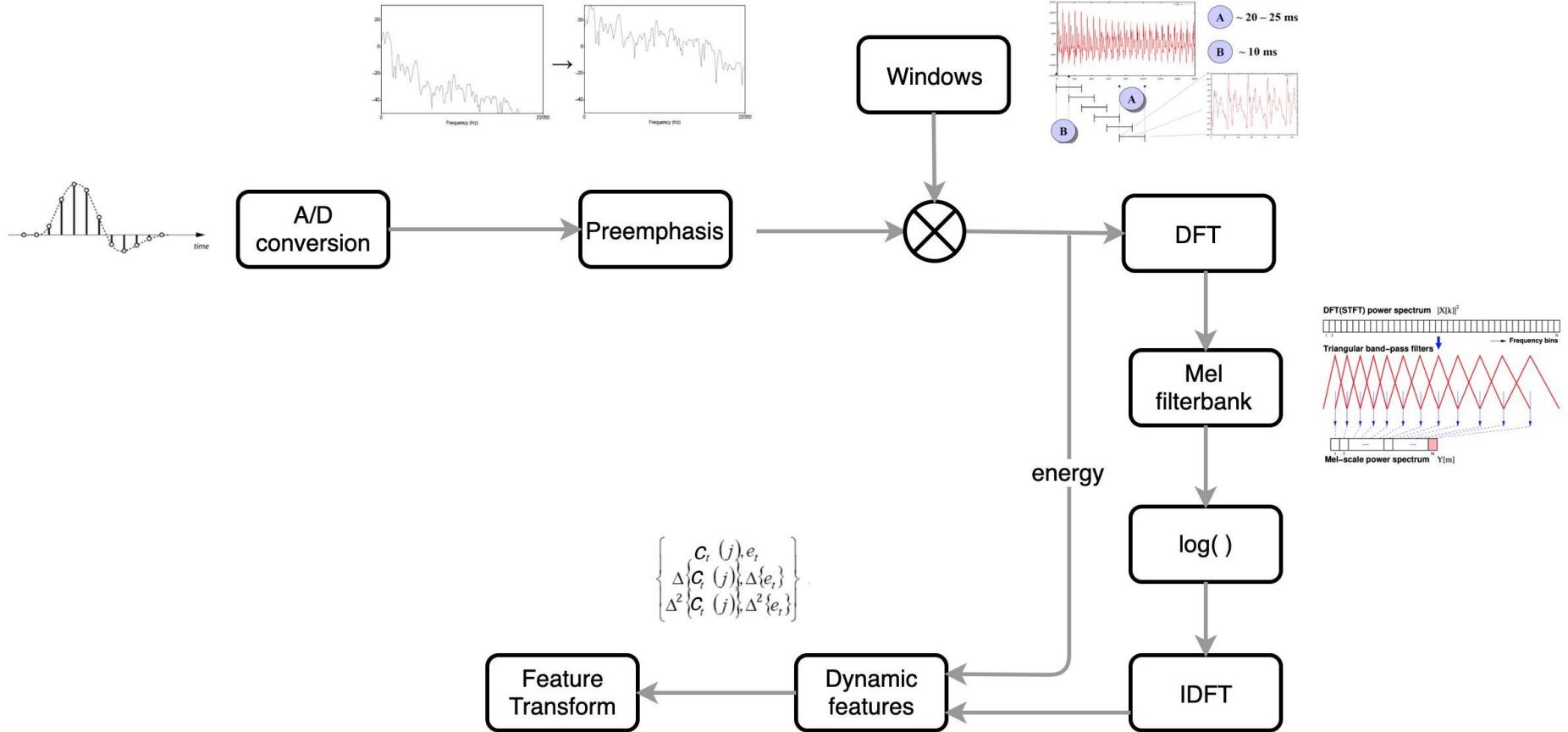
UNIVERSIDAD NACIONAL
DE TRES DE FEBRERO



EJEMPLOS

UNTREF

UNIVERSIDAD NACIONAL
DE TRES DE FEBRERO



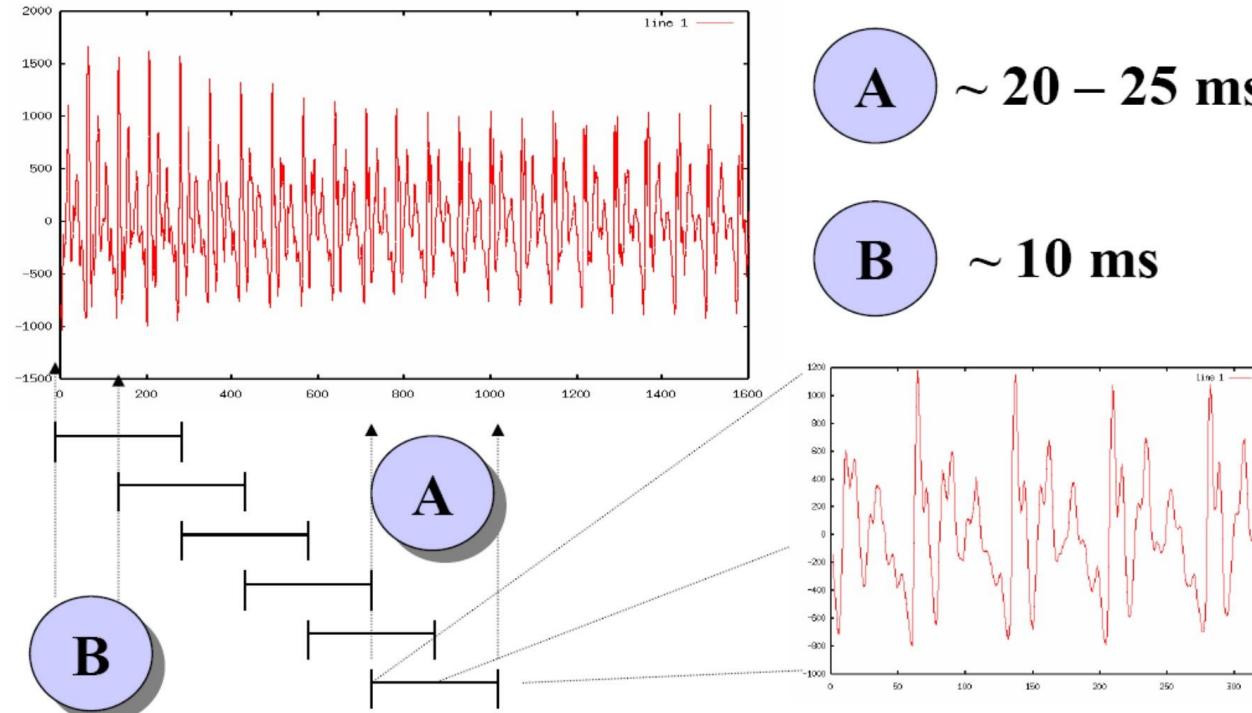
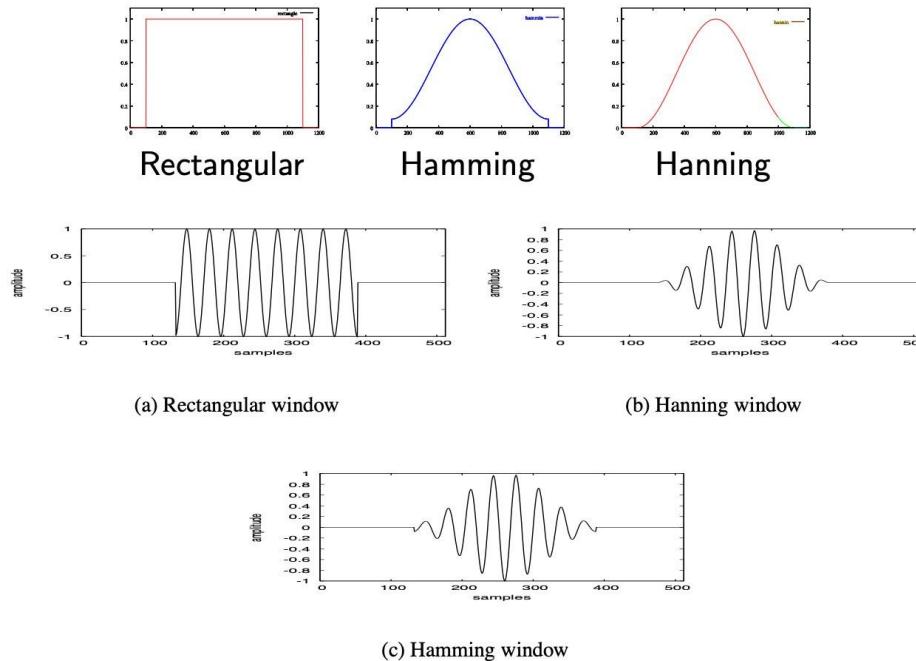


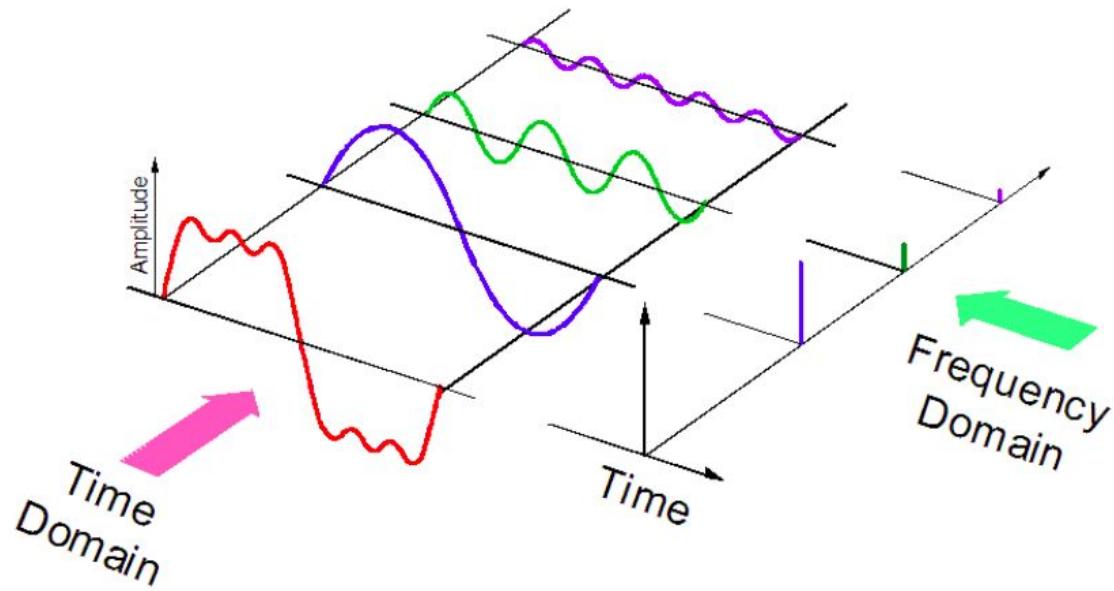
Image from Bryan Pellom

Pero no podemos simplemente cortar en el borde del frame, esto genera mucho ruido que se mostrará en el dominio de la frecuencia. Para cortar el audio, la amplitud debería disminuir gradualmente cerca del borde de un frame. Algunas alternativas para las ventanas son: Hamming y Hanning.



(Taylor, fig 12.1)

The central idea



DOMINIO TEMPORAL

Las características de audio en el dominio temporal se extraen directamente de las muestras de audio. Estas características ofrecen una forma simple de analizar las señales de audio, aunque generalmente es necesario combinarlas con características de dominio de frecuencia más sofisticada.

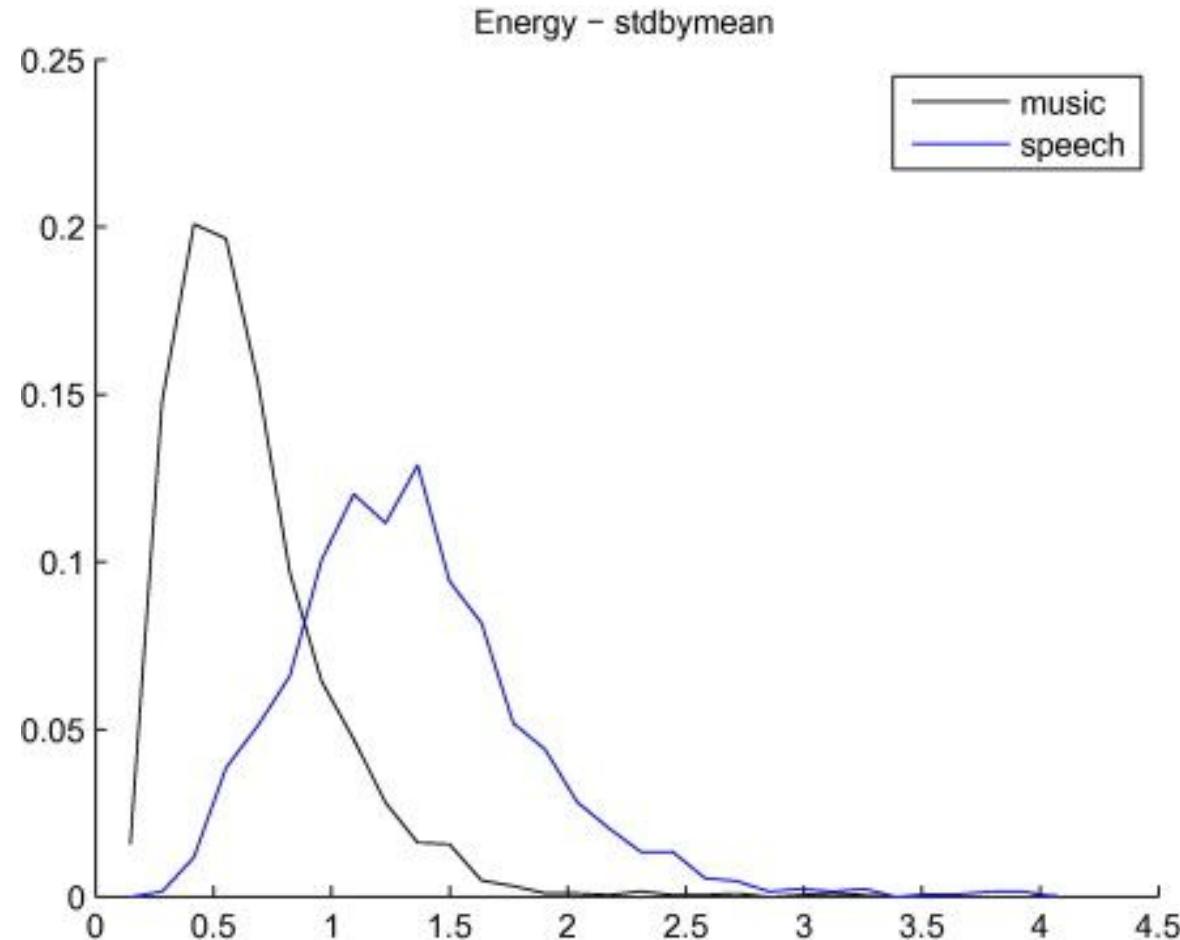
$$x_i(n), n = 1, \dots, W_L$$

ENERGÍA

W_L será la secuencia de muestras de audio del i -ésimo cuadro, donde W_L es la longitud del cuadro. La energía a corto plazo se calcula de acuerdo con la ecuación

$$E(i) = \sum_{n=1}^{W_L} |x_i(n)|^2.$$

$$E(i) = \frac{1}{W_L} \sum_{n=1}^{W_L} |x_i(n)|^2.$$



TASA DE CRUCES POR CERO

La tasa de cruce por cero de un frame, es la tasa de cambios de signos de la señal durante dicho frame. En otras palabras, es la cantidad de veces que la señal cambia el valor, de positivo a negativo y viceversa, dividido por la longitud del frame.

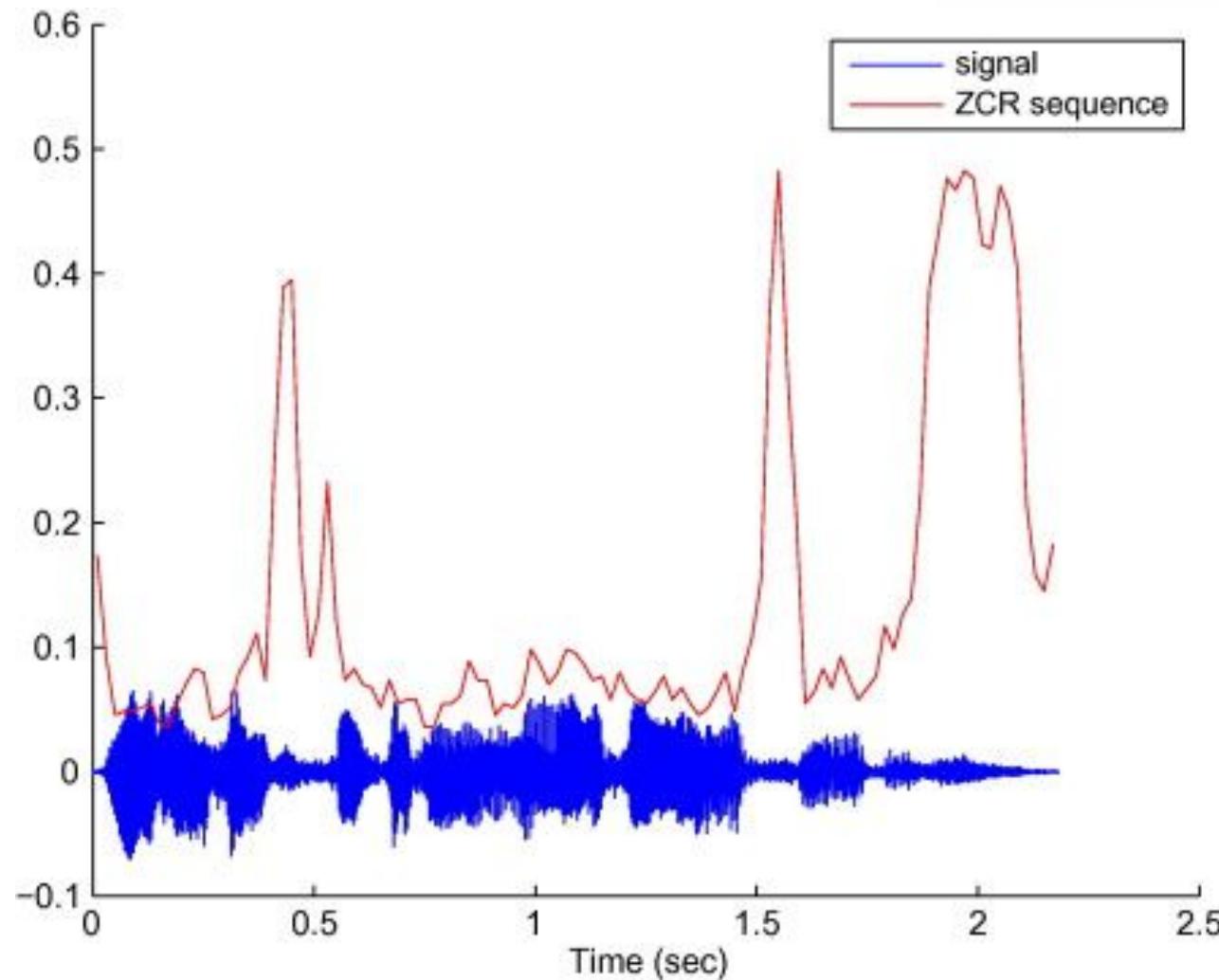
$$Z(i) = \frac{1}{2W_L} \sum_{n=1}^{W_L} | sgn[x_i(n)] - sgn[x_i(n-1)] |$$

$$sgn[x_i(n)] = \begin{cases} 1, & x_i(n) \geq 0 \\ -1, & x_i(n) < 0 \end{cases}$$

TASA DE CRUCES POR CERO

UNTREF

UNIVERSIDAD NACIONAL
DE TRES DE FEBRERO



Time-domain features

- Energy
 - usually normalized by window length
 - high variation over successive speech frames (std statistic)

$$E(i) = \frac{1}{W_L} \sum_{n=1}^{W_L} |x_i(n)|^2.$$

- Zero Crossing Rate

- rate of sign changes during the frame
- measure of noisiness
- high values for noisy signals

$$sgn[x_i(n)] = \begin{cases} 1, & x_i(n) \geq 0 \\ -1, & x_i(n) < 0 \end{cases} \quad Z(i) = \frac{1}{2W_L} \sum_{n=1}^{W_L} | sgn[x_i(n)] - sgn[x_i(n-1)] |$$

- Energy Entropy

- measure of abrupt changes in the signal's energy
- divide frames to K sub-frames and compute (normalized) sub-energies (e_{subframe_k})
- compute entropy of e_{subframe_k} sequence

$$e_j = \frac{E_{\text{subFrame}_j}}{E_{\text{shortFrame}_i}} \quad E_{\text{shortFrame}_i} = \sum_{k=1}^K E_{\text{subFrame}_j}.$$

$$H(i) = - \sum_{j=1}^K e_j^2 \cdot \log_2 (e_j^2)$$

Extraido de:

<https://docs.google.com/presentation/d/1heH7rKGEEySVh3sK583MuwqINwACiAerHQw4JQTntI4/edit?usp=sharing>

Frequency Domain (Spectral) Features

- Let X be the abs(FFT)
- Spectral Centroid
 - Center of gravity of the spectrum
- Spectral spread
 - 2nd central moment of the spectrum
- Spectral entropy
 - Divide spectrum into L sub-bands
 - Compute normalized sub-band energies (E_f)
 - Compute entropy
- Spectral Flux
 - Spectral change between two successive frames
- Spectral Rolloff
 - Freq below which a percentage of the mag distribution of the spectrum is concentrated
 - If the m -th DFT coefficient is the spectral rolloff

$$C_i = \frac{\sum_{k=1}^{Wf_L} (k+1) X_i(k)}{\sum_{k=1}^{Wf_L} X_i(k)}$$

$$S_i = \sqrt{\frac{\sum_{k=1}^{Wf_L} ((k+1) - C_i)^2 X_i(k)}{\sum_{k=1}^{Wf_L} X_i(k)}}$$

$$n_f = \frac{E_f}{\sum_{f=0}^{L-1} E_f}, f = 0, \dots, L-1 \quad H = - \sum_{f=0}^{L-1} n_f \cdot \log_2(n_f)$$

$$EN_i(k) = \frac{X_i(k)}{\sum_{l=1}^{Wf_L} X_l}$$

$$Fl_{(i,i-1)} = \sum_{k=1}^{Wf_L} (EN_i(k) - EN_{i-1}(k))$$

$$\sum_{k=1}^m X_i(k) = C \sum_{k=1}^{Wf_L} X_i(k)$$

Cepstral Domain

- Mel-Frequency Cepstral Coefficients
 - Compute DFT
 - Mel-scale filter bank application
 - Compute O_k as the power of the output of each filter
 - Compute MFCCs as the discrete cosine transform coefficients of the mel-scaled log-power spectrum
- Usually select the first 13 MFCCs (considered to carry enough discriminative information especially for speech classification tasks)
- Cepstrum in general (not mel):
 - Inverse fft of the log fft

$$f_w = 1127.01048 * \log(f/700 + 1)$$

$$c_m = \sum_{k=1}^L (\log \widetilde{O}_k) \cos[m(k - \frac{1}{2})\frac{\pi}{L}], \quad m = 1, \dots, L$$

$$: \left| \mathcal{F}^{-1} \left\{ \log \left(|\mathcal{F}\{f(t)\}|^2 \right) \right\} \right|^2$$



- librosa (Python)
 - <https://librosa.github.io>
 - Implements various audio features (mfccs, chroma, beat, etc)
 - Audio fx (e.g. pitch shift)
 - Some ML components
- pyAudioAnalysis (Python)
 - <https://github.com/tyiannak/pyAudioAnalysis>
 - Implements various audio features
 - Built-in training/testing of classifiers (using scikit-learn)
 - Clustering (speaker diarization)
 - Visualization
- Opensmile (C++)
 - <https://audeering.com/technology/opensmile/>
 - Richest set of audio features

Extraido de:

<https://docs.google.com/presentation/d/1heH7rKGEEySVh3sK583MuwqINwACiAerHQw4JQTntI4/edit?usp=sharing>

¿TE IMAGINAS EXTRAER AUDIO FEATURES DE SPOTIFY?



<https://developer.spotify.com/documentation/web-api/reference/tracks/get-audio-features/>

<https://medium.com/@RareLoot/extracting-spotify-data-on-your-favourite-artist-via-python-d58bc92a4330>

<https://spotipy.readthedocs.io/en/latest/>

Ingeniería de
Sonido

**V Jornadas de Acústica,
Audio y Sonido**

UNTREF
UNIVERSIDAD NACIONAL
DE TRES DE FEBRERO



Muchas gracias por su atención!

Facultad de Ingeniería del Ejército - Infiniem Labs Acústica

29 y 30 de agosto de 2019

Chungo, F., Groisman A., y Yommi, M.
infiniemplab.dsp@gmail.com

