

Extracción de Features en aplicaciones específicas

Taller de Procesamiento de Señales

Agenda

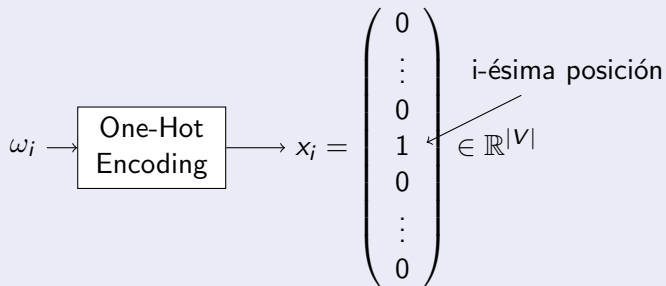
1 Procesamiento de Lenguaje Natural

2 Procesamiento de Sonido

¿Como convertir un texto en un vector?

One-hot Encoding

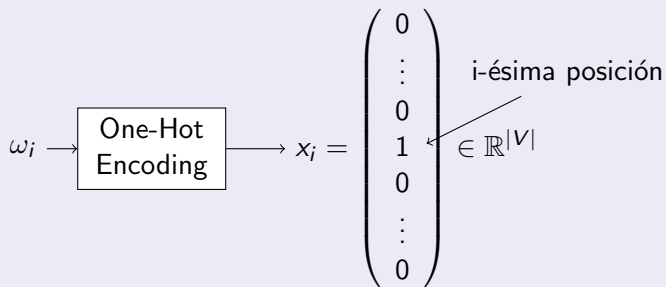
Dado un vocabulario $V = \{\omega_1, \dots, \omega_{|V|}\}$, se puede convertir cada palabra en un vector *one-hot*.



¿Como convertir un texto en un vector?

One-hot Encoding

Dado un vocabulario $V = \{\omega_1, \dots, \omega_{|V|}\}$, se puede convertir cada palabra en un vector *one-hot*.



Bolsa de palabras

Para vectorizar un documento $f(x_1, \dots, x_n)$, la manera más simple es *bolsa de palabras*: $f(x_1, \dots, x_n) = x_1 + \dots + x_n$.

Procesamiento del Lenguaje Natural

Vectorizaciones Sofisticadas

En la práctica suelen utilizarse representaciones pre-entrenadas (ej. FastText).

Procesamiento del Lenguaje Natural

Vectorizaciones Sofisticadas

En la práctica suelen utilizarse representaciones pre-entrenadas (ej. FastText).

Normalizaciones de NLP

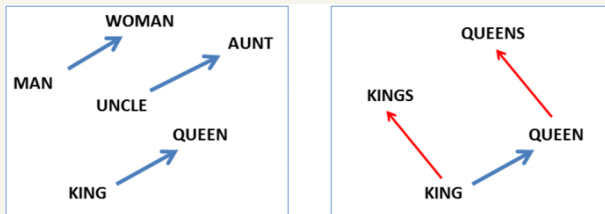
- Eliminar caracteres raros e inusuales
- Convertir todo a minúsculas
- Eliminar palabras no informativas (stop words)
- Descartar las palabras poco observadas
- Descartar las palabras más comunes
- Lemmatization (significado)
- Stemming (quedarse con la raíz)

Term Frequency - Inverse Document Frequency

Transformación tf-idf

Medida numérica que expresa cuán relevante es una palabra para un documento dentro de un dataset. El tf-idf para un término t de un documento d perteneciente a una colección de n documentos es $\text{tf-idf}(t, d) = \text{tf}(t, d) \cdot \text{idf}(t)$. El primer factor $\text{tf}(t, d) = \frac{\#(t \in d)}{\#(d)}$ es la cantidad de veces que aparece el término t en el documento d dividido la cantidad de términos que aparecen en el documento d . El segundo factor $\text{idf}(t) = 1 - \log\left(\frac{\text{df}(t)}{n}\right)$, donde $\text{df}(t)$ es la cantidad de documentos que poseen el término t en su interior.

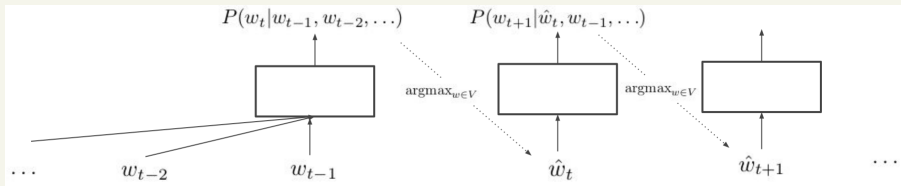
Word Vectors + PCA



(Mikolov et al., NAACL HLT, 2013)

$$\text{vector}(\text{KINGS}) - \text{vector}(\text{KING}) + \text{vector}(\text{QUEEN}) = \text{vector}(\text{QUEENS})$$

Síntesis de texto



Outline

1 Procesamiento de Lenguaje Natural

2 Procesamiento de Sonido

Coeficientes Cepstrum en escala de Frecuencia Mel

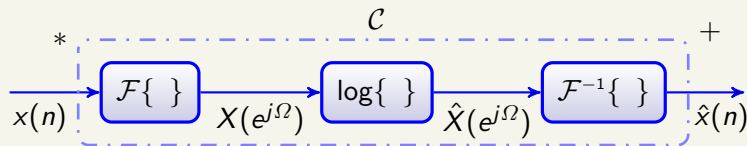
MFCC

Los Mel-Frequency Cepstral Coefficients (MFCC) son transformaciones muy utilizadas en procesamiento de sonido, sobre todo en procesamiento del habla.

- Es una forma alternativa de procesamiento en frecuencia, está basado en el análisis de Fourier de señales.
- Nos permite incorporar varios aspectos del procesamiento biológico del sonido (sistema auditivo externo).
- Genera características descriptivas de los sonidos en una dimensión manejable, apta para representar estadísticamente.

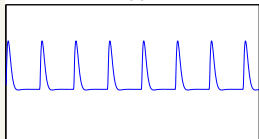
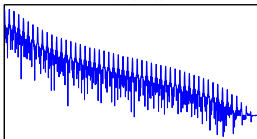
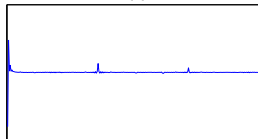
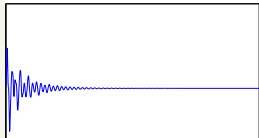
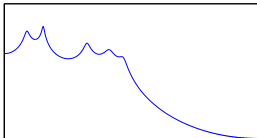
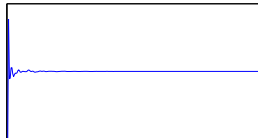
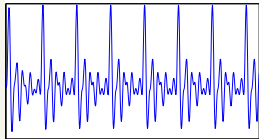
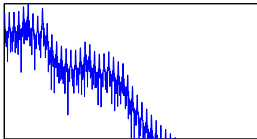
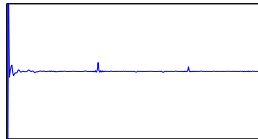
Transformada Cepstrum

La transformada Cepstrum puede transformar convoluciones en sumas:

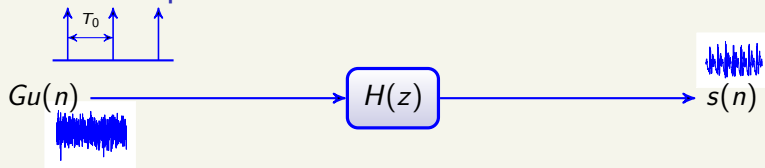


$$\begin{aligned}\mathcal{C}(a[n] * b[n]) &= \mathcal{F}^{-1} [\log |A(e^{j\Omega})B(e^{j\Omega})|] \\ &= \mathcal{F}^{-1} [\log |A(e^{j\Omega})|] + \mathcal{F}^{-1} [\log |B(e^{j\Omega})|] \\ &= \mathcal{C}(a[n]) + \mathcal{C}(b[n])\end{aligned}$$

Transformada Cepstrum

 $x[n]$  $\log(X(e^{j\omega}))$  $\hat{x}[n]$  $h[n]$  $\log(H(e^{j\omega}))$  $\hat{h}[n]$  $x[n] * h[n]$  $\log(X(e^{j\omega})H(e^{j\omega}))$  $\hat{x}[n] + \hat{h}[n]$ 

Modelo de producción del habla



$$H(z) = \frac{1}{A(z)}$$

con

$$A(z) = 1 - \sum_{k=1}^M a_k z^{-k}$$

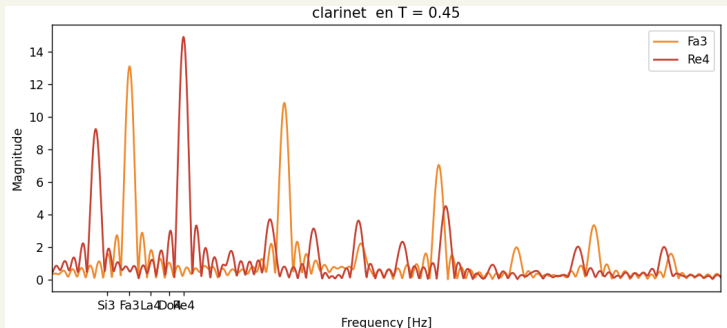
- Es posible modelizar el tracto vocal como un sistema lineal de M polos dado por $H(z)$.
- La entrada a dicho sistema $Gu(n)$ viene dada por un tren de impulsos o ruido blanco. La salida $s(n)$ es la señal de habla modelizada.

Coeficientes Cepstrum en escala de Frecuencia Mel

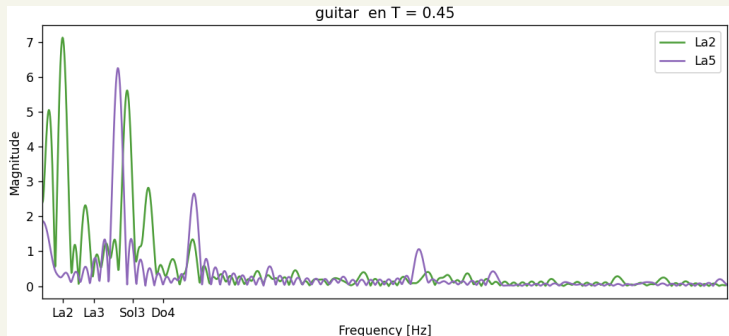
Que información contiene el cepstrum?

- La parte del cepstrum relacionada con el tracto vocal se concentra en la región de bajas **quefrecuencias**.
- La parte del cepstrum relacionada con la excitación glótica se concentra en las quefrecuencias altas.
- Es posible hacer una *deconvolución*, es decir separar excitación de filtro en dos partes separadas, simplemente quedándose con las quefrecuencias que sea pertinente, y volviendo al dominio del tiempo (o de las frecuencias de Fourier).
- El cepstrum permite estimar **la envolvente** del espectro del tracto vocal y el pitch.
- Para volver al dominio de Fourier simplemente se hace una Transformada de Fourier sobre el cepstrum (transformación lineal).

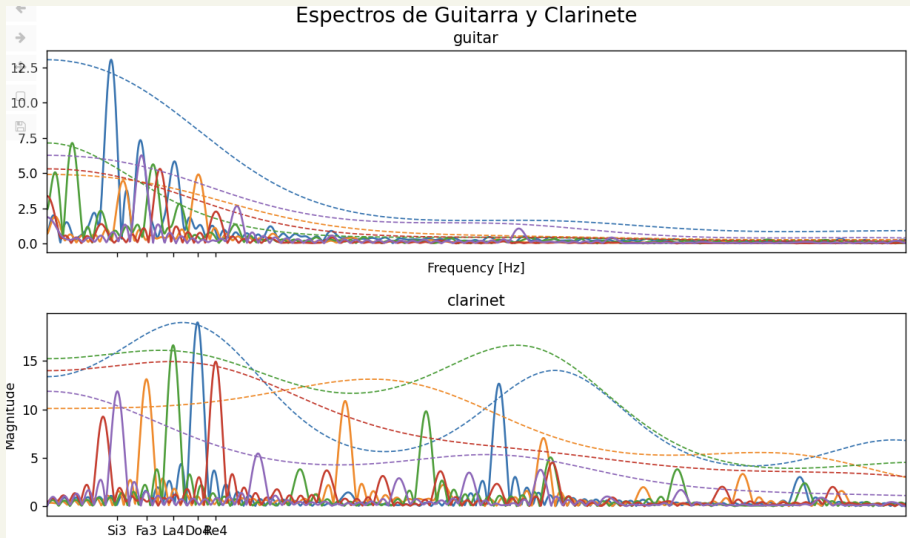
Coeficientes Cepstrum: envolvente vs espectro I



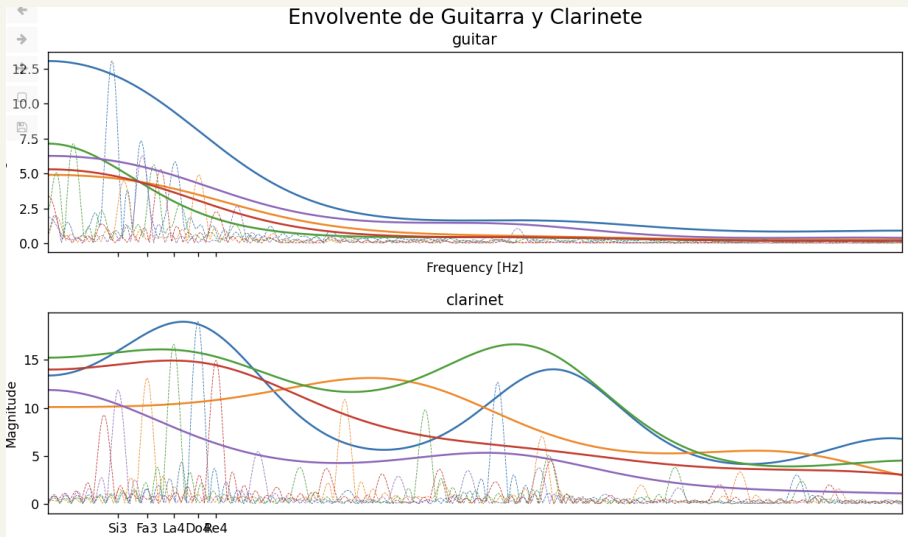
Coeficientes Cepstrum: envoltente vs espectro II



Coeficientes Cepstrum: envoltente vs espectro III

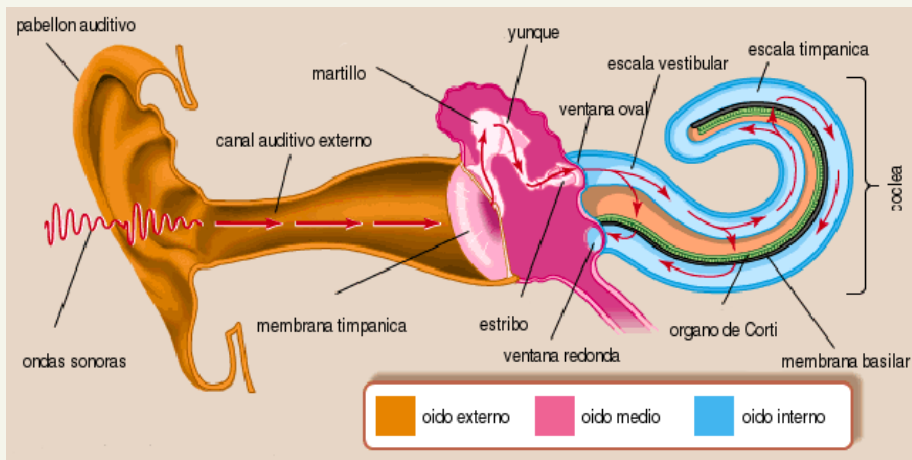


Coeficientes Cepstrum: envolvente vs espectro IV

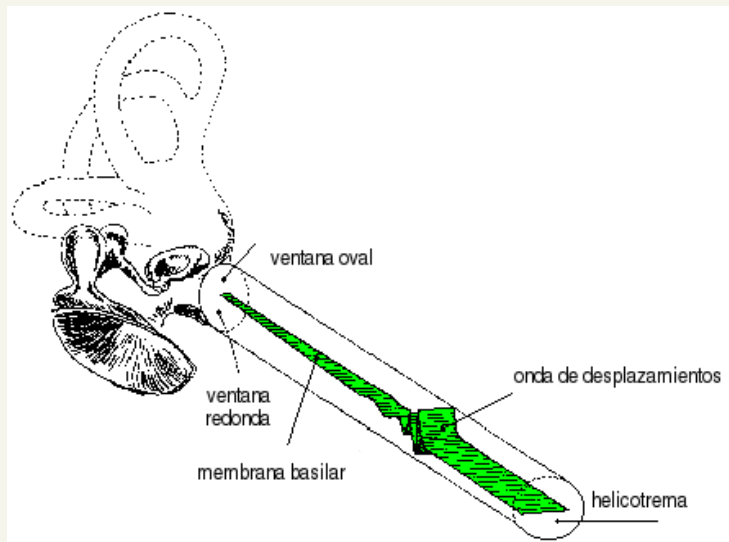


Coeficientes Cepstrum: motivación biológica I

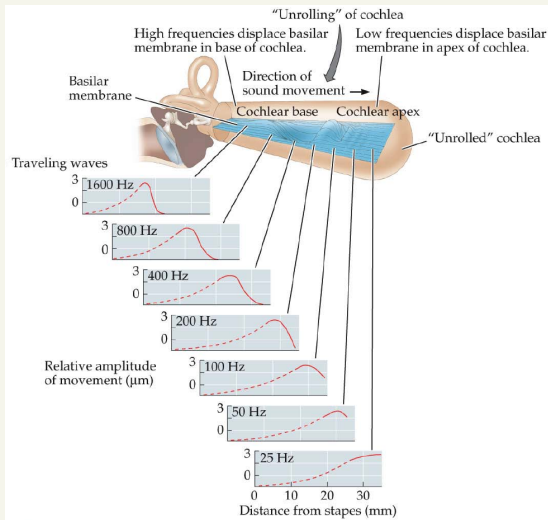
Agregado de información biológica



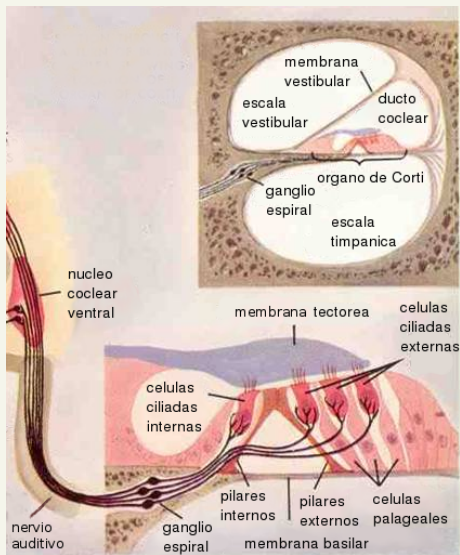
Coeficientes Cepstrum: motivación biológica II



Coeficientes Cepstrum: motivación biológica III

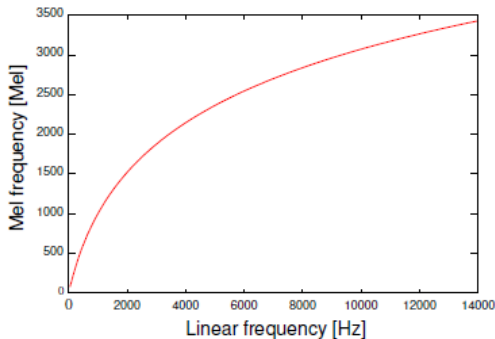


Coeficientes Cepstrum: motivación biológica IV

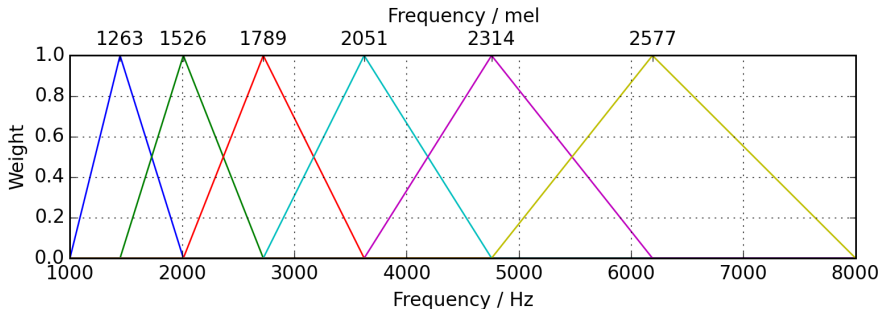


Coeficientes Cepstrum: la escala mel, y los filtros triangulares I

$$M(f) = 1127 \ln(1 + f/700)$$



Coeficientes Cepstrum: la escala mel, y los filtros triangulares II



Coefs. ceps. en escala de frec. mel (MFCC)

- Se calcula la transformada de Fourier (DFT) $X_t(k)$
- Ponderar los coeficientes con los correspondientes filtros triangulares W_m con $m = 1, \dots, M$

$$Y_t(m) = \sum_{k=L_r}^{U_m} |W_m(k)X_t(k)|^2$$

- Obtener el módulo del logaritmo de la salida de los filtros y realizar la transformada *coseno inversa*

$$mfcc(n) = \frac{1}{M} \sum_{m=1}^M \log[Y_t(m)] \cos \left[\frac{2\pi}{M} \left(m + \frac{1}{2}n \right) \right] \quad n = 1, \dots, L$$

- Típicamente $L \approx 13$, $M \approx 24$, $N \approx 512$
- Existen librerías de python que calculan coeficientes MFCC (librosa, Universidad de Columbia.)