TALLER DE PROCESAMIENTO DE SEÑALES

2do Cuatrimestre 2026 - Trabajo Práctico Nº 7 - PCA, EM

IMPORTANTE: En este ejercicio se valorarán las implementaciones propias, pero las mismas no son obligatorias.

Se desea agrupar sonidos. Para ello se cuenta con la base de datos **ESC-50**: 2000 grabaciones, de 5 segundos cada una, sobre 50 tipos de sonidos diferentes.

(**a**) *Análisis espectral:*

1. Leer la documentación de la base de datos `https://github.com/karolpiczak/esc-50`.
2. Descargar la base de datos. Dicha descarga debe estar autocontenida en el código.
3. Utilizando `load` (librosa) cargar al menos tres audios de la base de datos para una primera exploración: uno de una sirena, uno de una aspiradora y uno de bocinas (elegir uno que tenga más de un bocinazo) .
4. Graficar las señales temporales en función del tiempo (en segundos).
5. Reproducir los audios cargados previamente utilizando `Audio` (IPython).
6. Calcular la transformada de Fourier de corto tiempo de cada señal utilizando `stft` (librosa). Configurar dicha transformada para utilizar una FFT cuya cantidad de puntos coincida con el largo de la ventana, un solapamiento del 50 % y descartar la parte simétrica del espectro. Los parámtros a seleccionar serán el tipo (rectangular, hamming, hann o blackman) y largo de la ventana (entre 128 y 8192 puntos).
7. Elegir una ventana (tipo y largo) a partir de un análisis cualitativo de los espectrogramas (en escala logarítmica) de las señales. Se debe priorizar:
   - Resolución en frecuencia de la variación espectral de la sirena.
   - Visualización de la componente tonal asociada al motor de la aspiradora y posiblemente sus armónicos.
   - Separación temporal clara de los bocinazos.
8. Para cada ventana de tiempo, se desea descomponer el espectro de las señales en *bandas de energía*. Dividir el espectro en 20 bandas del mismo largo, y en cada una calcular la energía de la banda $i-$ésima $\mathcal{B}_i$ como:
$$E[i] = \sum_{k \in \mathcal{B}_i} |X[k]|^2$$
9. Convertir la energía en dB como:
$$E_{dB} = 10 \log_{10} \left(\frac{E}{E_{\text{máx}}} \right)$$
donde $E_{\text{máx}}$ es el máximo valor de energía de toda la señal.

(**b**) *Reducción de dimensión:*

1. Separar el conjunto de datos en train y test con las proporciones 80 % y 20 %.
2. Preprocesar los datos para quedarse con las bandas de energía en dB. : Cada muestra debe tener varias ventanas de tiempo (correspondientes a los 5 segundos) y 20 bandas de energía.
3. Aplanar los datos en un vector unidimensionar por muestra.
4. Entrenar un algoritmo PCA para conservar el 95 % de la energía (durante el entrenamiento). ¿Cuantas componentes fueron necesarias?

(**c**) *Clustering:*

1. Entrenar un algoritmo EM de mezcla de gaussianas.
2. Clusterizar los datos de testeo (*hard*).
3. Explicar que es el *Fowlkes-Mallows index* (FMI). Indicar el FMI de testeo.
4. Indicar la log-verosimilitud de testeo.
5. Probar variantes del pipeline e hiperparámetros para mejorar el resultado.