**Separador de voz en audio – Etapa 1**

Borys Agustín – Veraza Jeremias

La idea principal del proyecto, es separar y quitar la voz de un archivo de audio o canción, que posea varios instrumentos y una voz cantada. Generando dos nuevos archivos de audio:

* Un audio que incluya solo la voz de la canción, sin instrumentos.
* La pista de la canción sin la voz cantada.

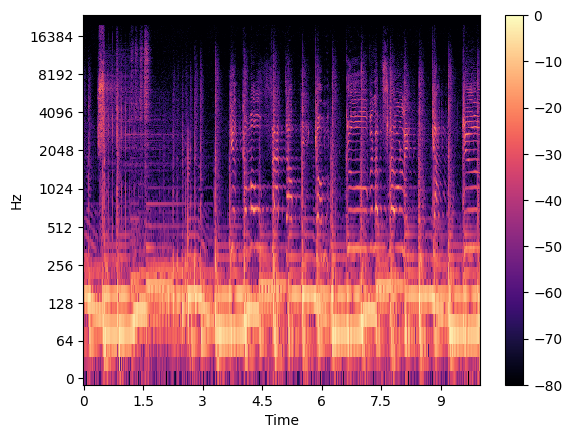
**Etapa 1:**

Utilizando la librería [librosa](https://librosa.org/doc/latest/index.html) procederemos a separar la voz (foreground) de la melodía (background). Para ello, utilizaremos 2 archivos de testeo, una canción conocida como “Come Together” de The Beatles y otra canción open source que nos brinda librosa en los archivos de ejemplo.

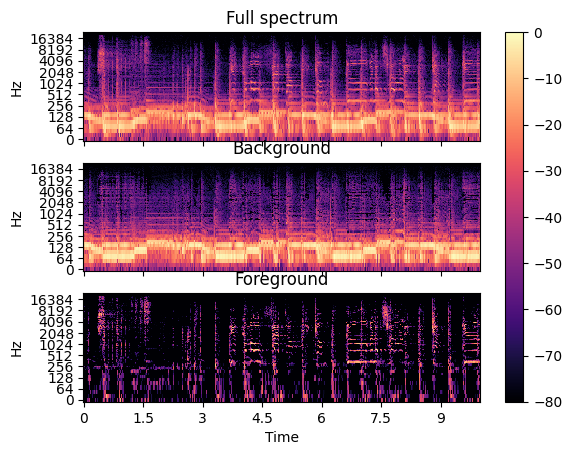
**Proceso de separación:**

El proceso para la separación consta de:

1. Cargar el archivo de audio, el mismo debe tener formato .wav
2. Convertir el archivo a dos arrays de numpy, conteniendo amplitud y fase respectivamente. Esto lo realizamos haciendo utilidad de las funciones de la librería librosa.
3. En este caso decidimos mostrar el espectro de la canción utilizando la librería matplotlib, y obtuvimos lo siguiente:



En el espectrograma se puede observar que las frecuencias centrales (entre los 512 Hz y 4096 Hz), se forman una especie de líneas rectas junto con unas ondulaciones, corresponden a la componente vocal de la canción, mientras que el resto del espectro corresponde a la componente instrumental.

1. Para el filtrado utilizamos una comparación por similitud cosenoidal, incluida en la librería ‘librosa’. Esta busca la similitud por valor medio de la señal, y suprime las partes de la señal que se desvían del promedio del espectro. Esto resulta muy bien para suprimir la voz.
2. Para mejorar la separación de ambas partes deseadas, se pasa la señal por dos mascaras diferentes, una para suprimir la voz por completo, y otro que solo permite el paso de la señal de voz. Obteniendo dos archivos de audio, uno con la voz, y otro con la pista instrumental.
3. Volviendo a graficar el espectro original, y los espectros de ambos archivos de audio procesados, vemos como quedan separadas la voz (Foreground) y la pista musical (Background) correctamente:
4. Una vez verificado el filtrado, procedemos a la recuperación del audio para la componente vocal y para la componente instrumental introduciendo la información de la fase para caso.

**Conclusión**

Como conclusión, pudimos ver que, en algunos casos, probando con diferentes archivos de audio, que quedan diferentes residuos de la señal original en la filtrada. Para la segunda etapa del proyecto, se intentará mejorar la calidad del filtro que busca las similitudes para encontrar el espectro de la voz, y se utilizaran diferentes mascaras para eliminar los residuos de una mejor manera.