Tarea 6. Análisis de audio

Helena Patricia Carrillo Soto
Facultad de Ciencias Físico-Matemáticas, UANL

Maestría Ciencia de Datos: Procesamiento y Clasificación de Datos

Resumen

Se analizaron cuatro grabaciones en dos idiomas diferentes de dos voces distintas. Dichos análisis se comparan para encontrar similitudes y diferencias entre los idiomas y las voces.

Keywords: Voces; Análisis de audio

1. Introducción

Audios En este artículo se realizará un análisis de audio de dos voces: una femenina, voz uno, y una masculina, voz dos. Para ambas voces se analizan dos palabras una en español, "hola", y otra en inglés, "hello".

Análisis a realizar En total se realizaron tres análisis a cada una de las grabaciones: una descripción de MFCC (*Mel Frequency Cepstral Coefficients*), un espectrograma y la obtención de las frecuencias mas representativas .

2. Descripción de los datos

En total se analizaron cuatro grabaciones:

- ***Voz 1. Español***. Voz femenina diciendo la palabra "hola".
- ***Voz 1. Inglés***. Voz femenina diciendo la palabra "hello".
- ***Voz 2. Español***. Voz masculina diciendo la palabra "hola".
- ***Voz 2. Inglésl***. Voz masculina diciendo la palabra "hello".

3. Metodología

Para iniciar el análisis se cortan los audios para eliminar la mayor cantidad de silencio posible y con esto limitar los efectos del ruido de fondo en las grabaciones. Posteriormente para realizar los análisis se utilizó la librería de Python `librosa`.

Como se mencionó con anterioridad los análisis a realizar para cada una de las grabaciones son:

- Análisis MFCC
- Análisis de espectrograma
- Análisis de las frecuencias mas representativas

Análisis MFCC Este análisis consiste en la obtención de coeficientes, usualmente trece, usados para la representación del habla basados en la percepción auditiva humana. Cada coeficiente brinda distinta información:

- **Coeficiente 0:** Representa la energía o energía logarítmica que ayuda a distinguir entre sonidos suaves y sonidos fuertes.
- **Coeficiente 1:** Representa las frecuencias bajas de la voz humana.
- **Coeficiente 2:** Representa las frecuencias bajas a medias que son claves para la claridad del habla.
- **Coeficiente 3:** Representa las frecuencias medias que ayudan a distinguir entre diferentes sonidos del habla.
- **Coeficientes del 1 al 12:** Representan frecuencias altas que ayudan a reconocer variaciones sutiles en la voz y sonidos.

Espectrograma Gráfica que muestra el espectro de frecuencias de una señal por el tiempo en que ocurre el mismo marcando también la intensidad de cada una. Permite identificar variaciones de frecuencia e intensidad a lo largo de un periodo de tiempo.

Frecuencias representativas Este análisis toma aquellas frecuencias que mejor representan a la señal analizada y las presenta en función a su amplitud.

4. Resultados

Análisis individual Cada grabación fue analizada por separado. Los resultados para la grabación de la voz 1 en español se presentan en las figuras 1, 2 y 3.

De los resultados de la figura 1, por ejemplo, podemos destacar como el coeficiente 0, línea más cercana a la base de la gráfica, muestra menos intensidad, más azul, en los momentos de relativo silencio. Esto tiene sentido ya que el coeficiente 0 indica la fuerza del sonido.

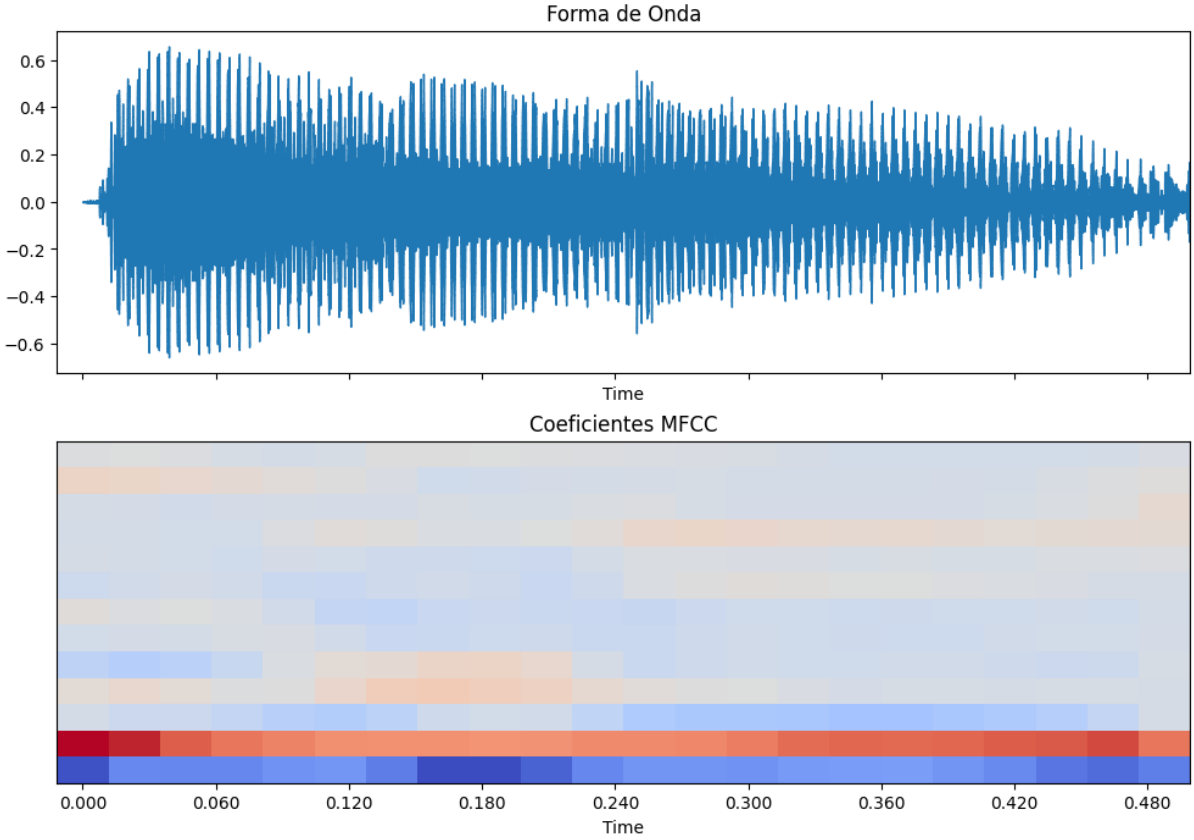

Figura 1: MFCC para la voz 1 en español

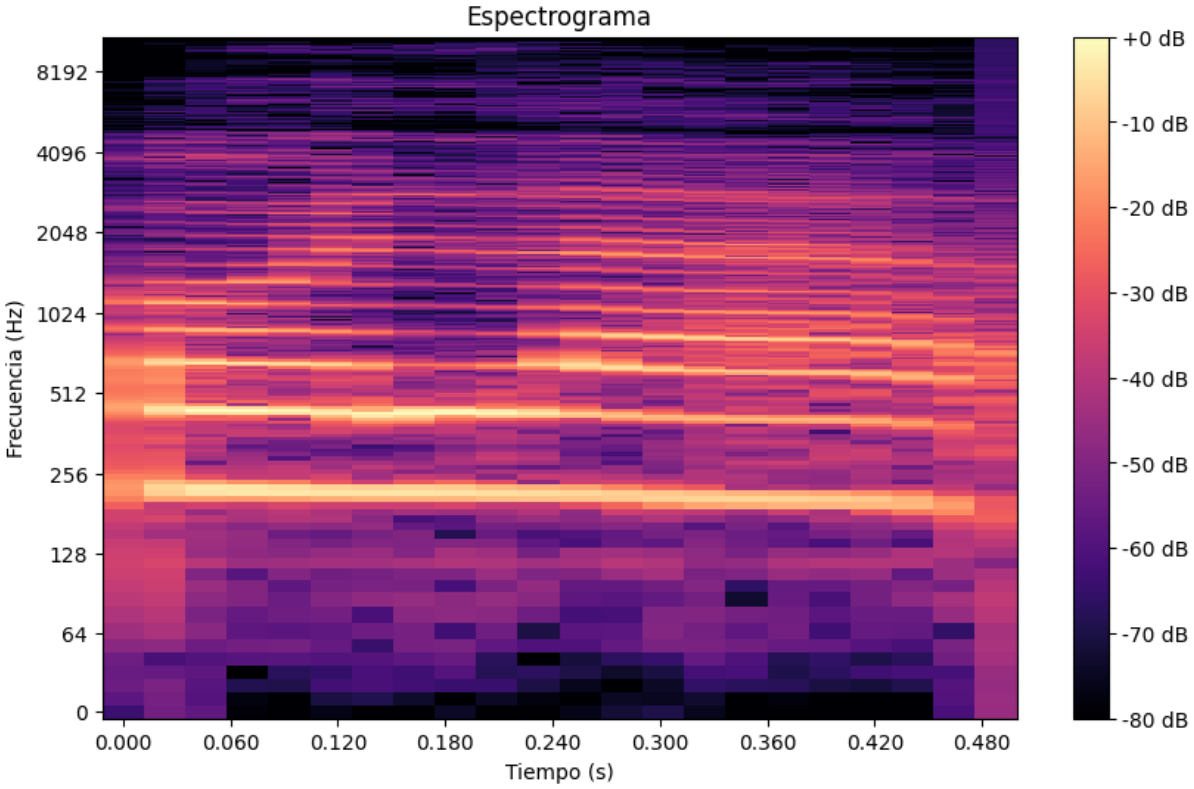

Figura 2: Espectrograma para la voz 1 en español

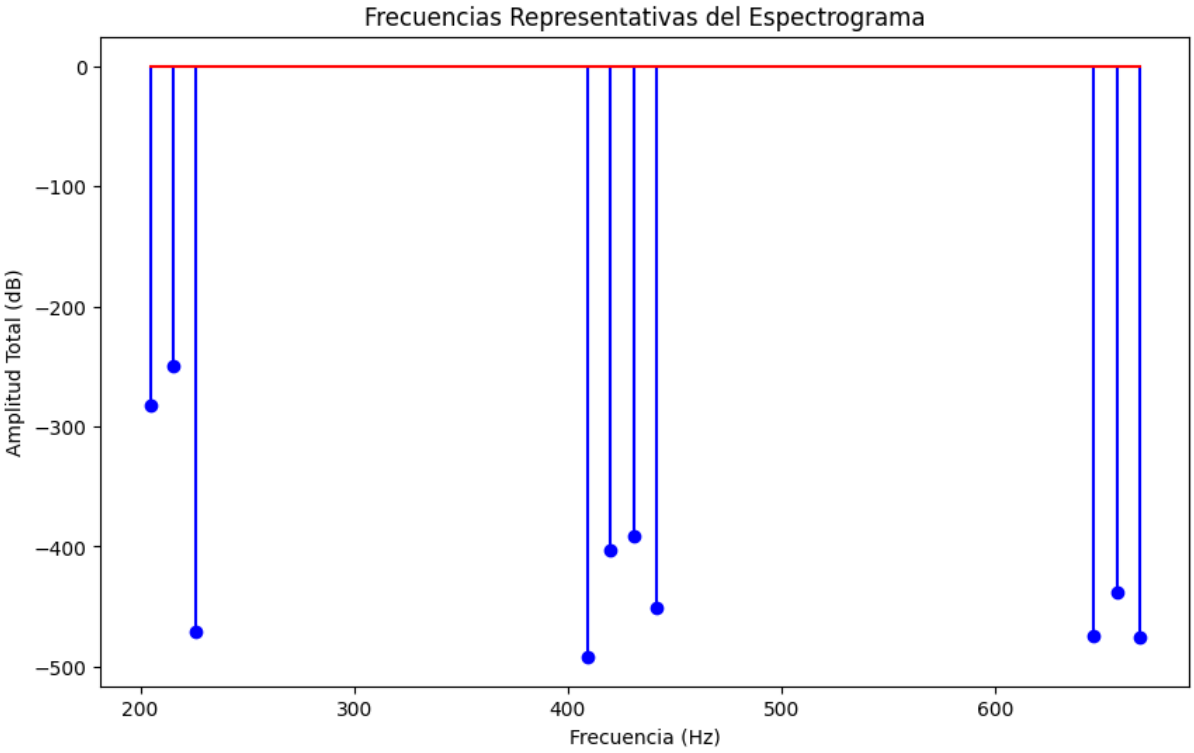

Figura 3: Frecuencias representativas para la voz 1 en español

Comparaciones Después de analizar cada una de las grabaciones se procedió a hacer comparaciones a pares ya sea por voces o por idioma.

En la figura 4 se pueden observar los cuatro análisis MFCC realizados.

En general se puede ver que las grabaciones de la voz dos presentan una mayor intensidad en el coeficiente 2, encargado de las frecuencias bajas. Esto hace sentido ya que la voz dos pertenece a una voz masculina y la voz uno es una voz femenina.

También podemos notar que en los coeficientes superiores de la voz dos se observa más intensidad que la voz uno. Esto podría deberse a ruido en las grabaciones.

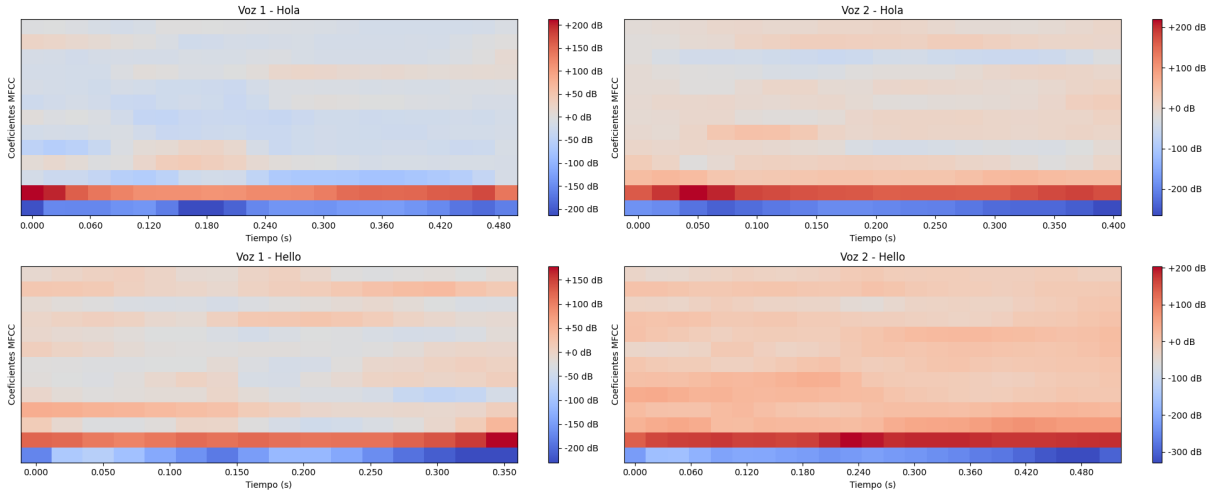

Figura 4: MFCC para las cuatro grabaciones

En la figura 5 se pueden observar los cuatro espectrogramas.

Podemos ver que para la voz uno las frecuencias con mayores intensidades usualmente comienzan alrededor de lso 256 mientras que para la voz dos esto se da entre las frecuencias 84 y 128.

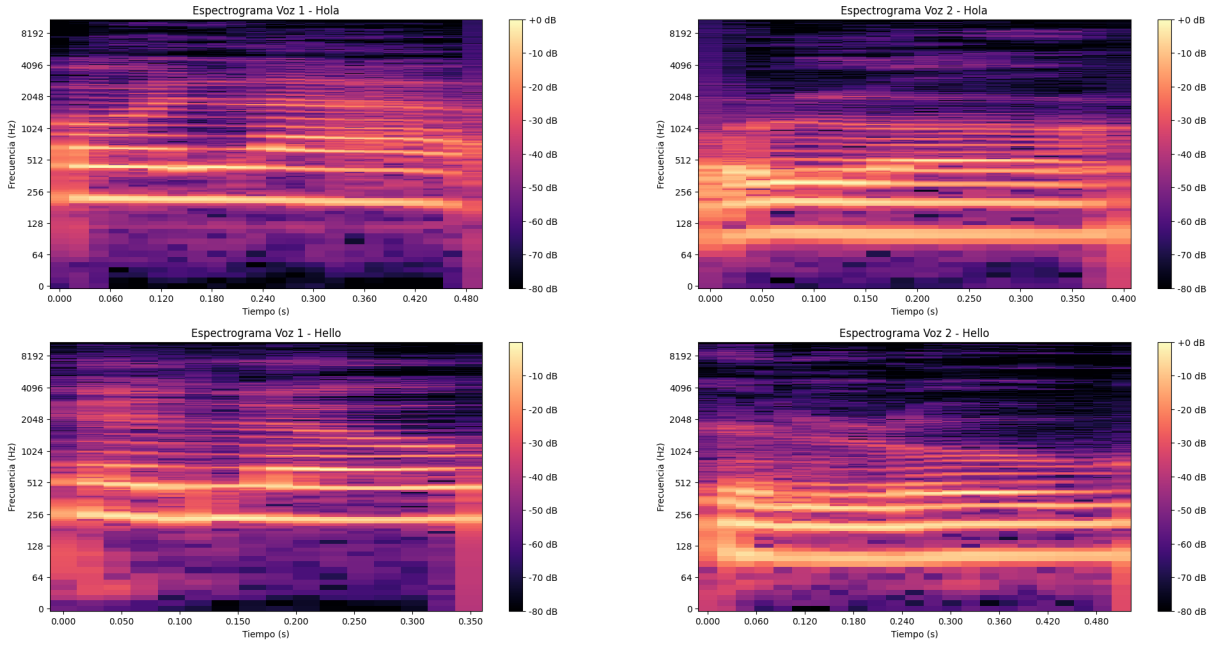

Figura 5: Espectrograma para las cuatro grabaciones

En la figura 6 se pueden observar gráficas comparativas a pares ya sea por voz o por idioma de las cuatro grabaciones.

Podemos ver que se comparten más frecuencias por voz que por idioma. Especialmente la voz dos ya que mucha de las frecuencias distintivas se superponen tanto para español como para inglés. Por otro lado en la voz uno podemos observar más diferencias en ambos idiomas. Esto sugiere un cambio en la tonalidad al momento de cambiar de español a inglés.

En el idioma inglés las mayoría de las frecuencias representativas de ambas voces se concentran por debajo de los 800, especialmente la voz dos, mientras que para español estas pueden llegar hasta 1000.

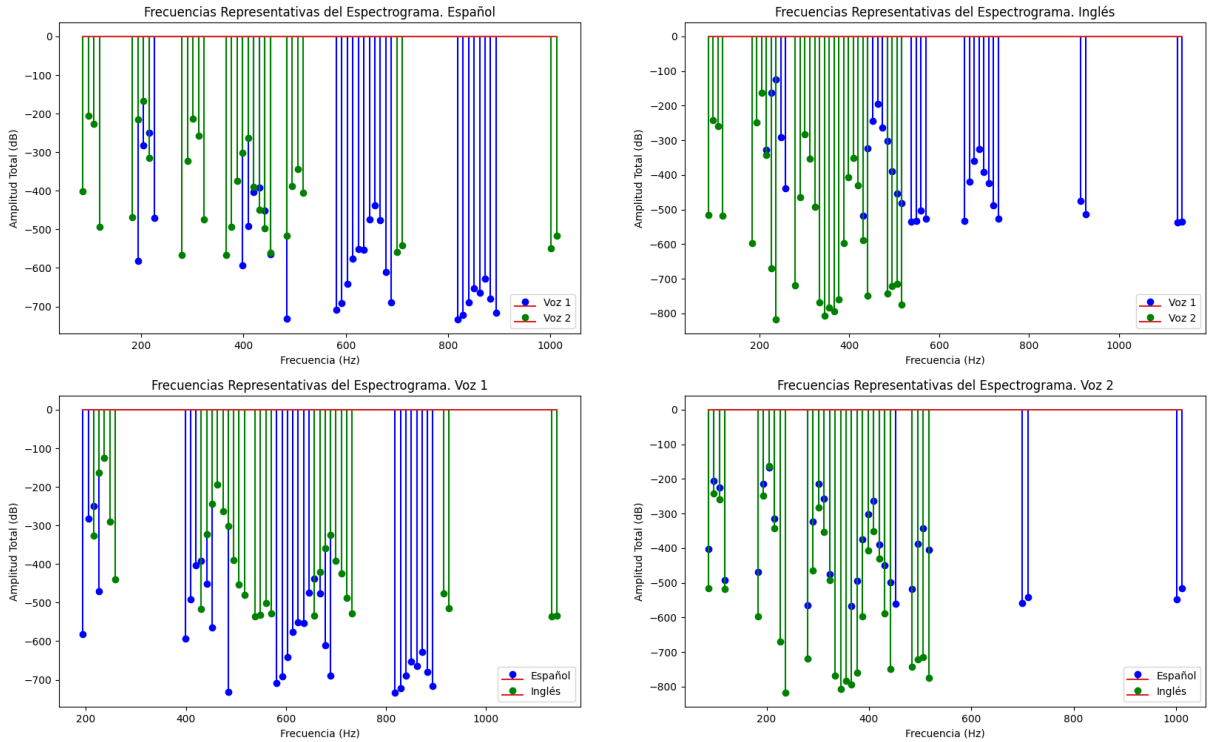

Figura 6: Frecuencias representativas comparativas a pares

5. Conclusiones

Es posible encontrar similitudes en las grabaciones ya sea por idioma o por voces. Sin embargo para realizar un mejor análisis es necesario minimizar el efecto que el ruido pueda tener en las grabaciones. Se propone hacer un preprocesamiento de las grabaciones para eliminar la influencia de factores externos en futuros análisis.