

UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE INGENIERÍA
INGENIERÍA ELECTRÓNICA Y TELECOMUNICACIONES
TRATAMIENTO DE SEÑALES I
LABORATORIO 1
CORRELACIÓN.

Justificación

La correlación es un procedimiento matemático que permite comparar señales en el dominio del tiempo y cuantificar su similitud. Las aplicaciones en ciencias e ingeniería son incontables, transformándola en una de las técnicas básicas en muchos procedimientos dentro del análisis de señales. Como un acercamiento a este método, se usará la correlación para determinar la frecuencia fundamental de una señal de voz y, en el caso de Pearson y Spearman, para conocer la relación existente entre dos conjuntos de datos.

Objetivos

- Afianzar el concepto y el uso de la función de correlación mediante un ejemplo práctico.
- Conocer y poner en práctica diferentes funcionalidades de paquetes de Python útiles para la abstracción de información relevante de una señal.
- Conocer otros métodos existentes para correlacionar datos.

Teoría:

Los valores de correlación de Pearson y Spearman nos permite cuantificar la dependencia que hay entre 2 grupos de datos: X y Y. El coeficiente de correlación de Pearson es una medida de la fuerza de una asociación lineal entre dos variables y se denota por r :

$$r_{x,y} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

Con

σ_{xy} Covarianza entre las variables X y Y
 σ_x : Varianza de la variable X
 σ_y : Varianza de la variable Y

El valor de correlación de Pearson cumple con las siguientes condiciones: $r \in [-1,1]$, si $r < 0$, las dos variables tienen una asociación lineal negativa, lo que significa que si una variable aumenta, la otra variable disminuye. Si $r = 0$, se dice que no hay asociación entre las dos variables y si $r > 0$, existe una asociación lineal positiva, lo que significa que si el valor de una variable aumenta, el valor de la otra variable también aumenta.

La correlación de Spearman es una prueba no paramétrica que se utiliza para medir el grado de asociación entre dos variables. El coeficiente de correlación de Spearman es el análisis de correlación apropiado cuando las variables se miden en una escala que es por lo menos ordinal y viene definido por:

$$\rho = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}$$

Con

$$d_i: \text{diferencia de los rangos de } X \text{ y } Y$$

$$n: \text{número de elementos}$$

El valor de correlación de Spearman ρ , tiene las siguientes características: Su valor va de -1 a +1 y la ventaja de la correlación de Spearman sobre la correlación de Pearson es que Spearman es útil para encontrar correlaciones no lineales entre las variables.

Procedimiento

Parte 1: Función de correlación Básica.

Para el desarrollo de la práctica, deberá tener un archivo .wav de duración promedio 7 segundos, grabado por usted mismo para calcular la frecuencia fundamental de su voz. Dado el caso que no le sea posible grabar el archivo .wav, se adjunta un audio junto con la guía de este Laboratorio.

1. Cargar un archivo de audio y graficarlo

a. Use la función `read` de la librería `scipy.io.wavfile` para cargar el archivo de audio en formato .wav. Tenga en cuenta que la función `read` retorna dos valores, los valores de amplitud de la señal de audio **x**, y la frecuencia de muestreo **fs**.

```
fs,x =read('audio.wav')
```

b. La idea en general, es tener una representación de la señal de audio **x** en función del tiempo. Ya que la variable **x** solo tiene información de la amplitud, es necesario obtener el vector de tiempo, para que cada punto en la señal **x** esté representado como un valor en amplitud y un valor en el tiempo:

```
t=np.arange(0, len(x)/fs, 1.0/fs)
```

fs es la cantidad de puntos que se toman por unidad de tiempo. Así, si se divide la cantidad de datos (puntos) por la frecuencia de muestreo, se obtiene la duración del audio en segundos.

c. Finalmente use el comando `plot` de la librería `matplotlib.pyplot` para graficar la señal en función del tiempo, deberá ver algo similar a la imagen mostrada en la Figura 1:

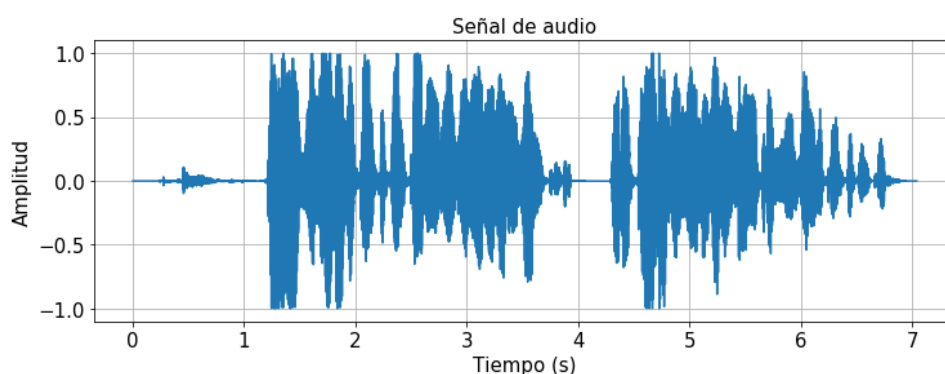


Figura 1. Señal de audio en función del tiempo.

2. Observe el grafico obtenido y almacene en una variable un segmento de la señal de audio donde haya registro de voz. Grafique el segmento de la señal junto con la señal original, para esto utilice la función *subplot*, debe observar algo similar a la imagen de la Figura 2.

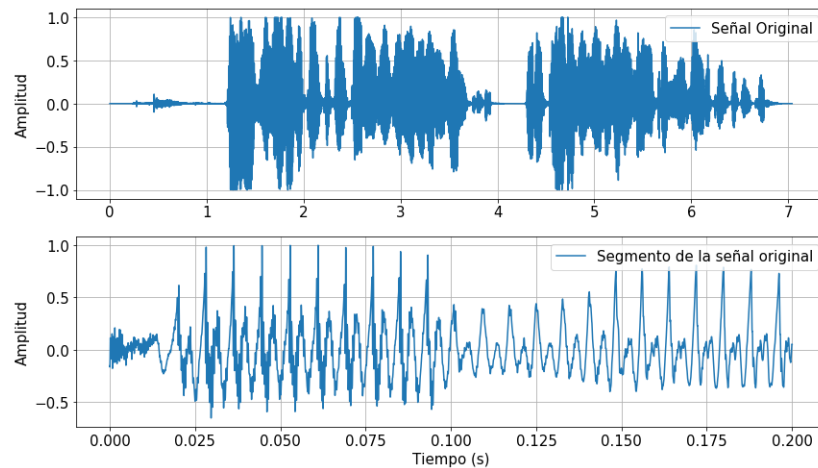


Figura 2. Señal original y segmento de 0.2 segundos en función del tiempo.

Recuerde que la correlación mide la similitud entre dos señales. A continuación, en la Figura 3, se muestra gráficamente, la autocorrelación de un segmento aleatorio de la señal.

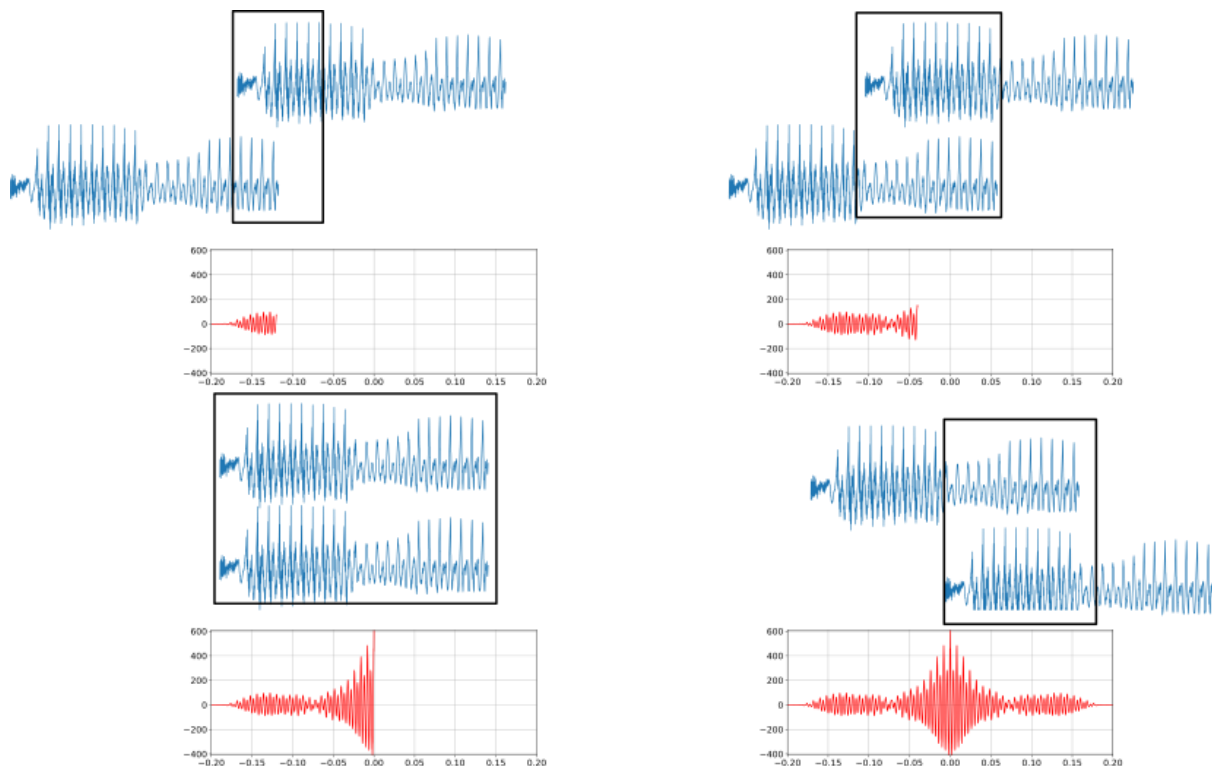


Figura 3. Ejemplo de autocorrelación de manera gráfica de un segmento del audio original elegido aleatoriamente.

Note que, a medida que el área de comparación se va haciendo mayor, igualmente se obtienen picos con mayor energía. Note además, que cada que un ciclo de la señal está comparándose con un ciclo exactamente igual, el resultado de la autocorrelación es un máximo local.

3. Realice la autocorrelación del segmento de la señal (se recomienda usar la función *correlate* de la librería *numpy*). Grafique en una misma gráfica la señal original, el segmento y el resultado de la autocorrelación del segmento, deberá observar algo similar a lo que se muestra en la Figura 4. Explique lo obtenido.

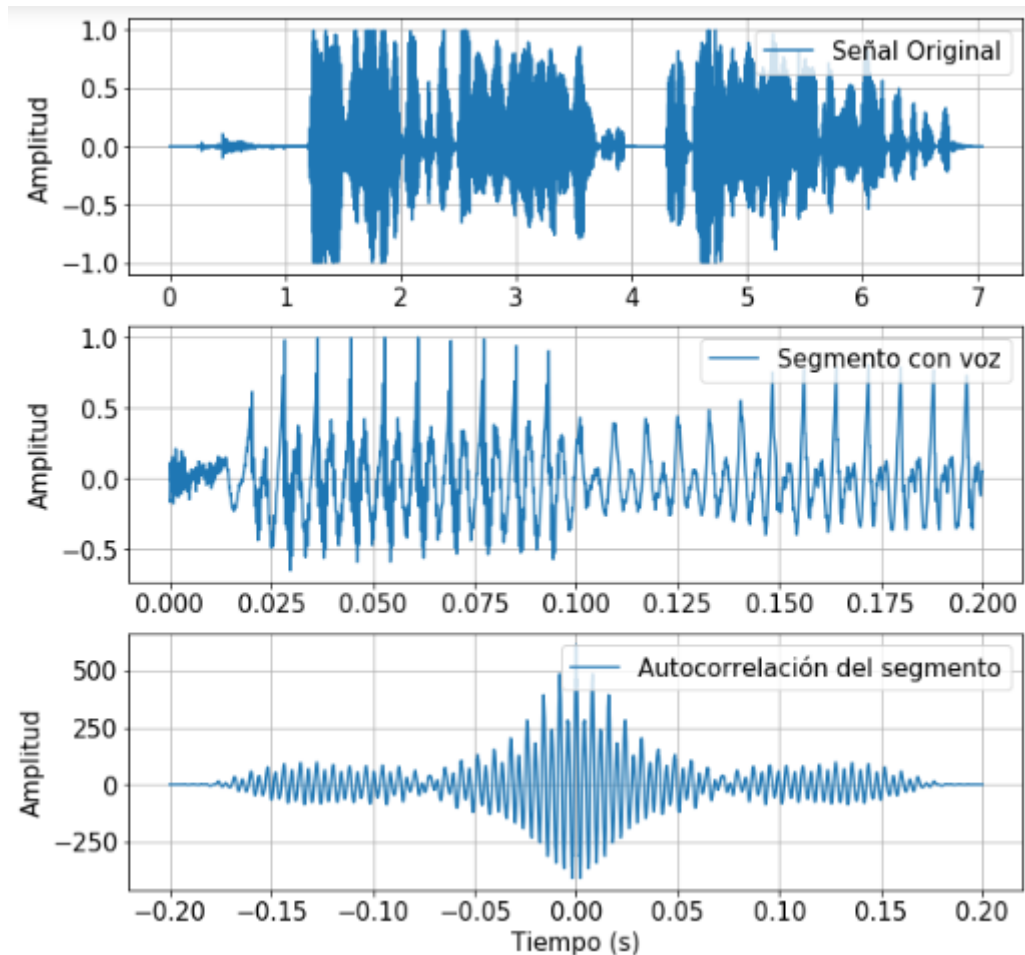


Figura 4. Señal original, segmento y autocorrelación del segmento.

4. Realizar el mismo procedimiento para un segmento sordo (segmento donde vea que no hay registro de voz). Explique los resultados obtenidos.

Parte 2: Radar

5. Comenzaremos creando el pulso $x(n)$ que el transmisor envía. Para ello cree una señal de 1000 puntos en el intervalo de tiempo entre -10 y 10, (puede usar la función *linspace* de la librería *numpy* para las abscisas). Para la señal tenga en cuenta que los puntos 500 a 550 estén en valor 1 y el resto en 0. Calcule el ciclo de dureza de la señal y grafíquela.

6. Ahora simulará la afectación que el canal produce sobre la señal. Inicialmente atenué la amplitud del pulso $x(n)$ en un 50%. Para simular el tiempo transcurrido desde el envío del pulso hasta que encuentra el objeto y retorna al receptor, haga un desplazamiento circular de la señal de 200 puntos. Para ello puede utilizar la función *roll* de la librería *numpy*. Grafique el resultado.

7. Ahora introduzca la afectación del canal con relación al ruido. Genere ruido gaussiano de media 0 y desviación estándar 0.3. Sume el ruido a la señal atenuada y desplazada. Realice la gráfica. ¿Resulta fácil la identificación del pulso en la señal de retorno?

8. Calcular la correlación:

a. Calcule la correlación entre la señal original y el pulso modificado.

b. Grafique el resultado de la correlación. ¿Cuáles son las coordenadas del pico máximo de correlación? ¿Cuál es la energía de la señal?

c. Repita el literal a y el literal b usando una desviación estándar de 0.7 ¿Qué se puede concluir?

Parte 3: Correlación Pearson y Spearman.

De un grupo de 55 estudiantes universitarios quienes están asistiendo al curso de Cálculo I e Informática I, se recolectó en la siguiente tabla la edad del estudiante y la nota que sacó entre 0-20 en el último parcial, siendo 20 la mayor nota. Los datos se encuentran en el archivo adjunto con esta guía "DatosCorrelacion.xlsx".

9. Genere un vector columna para cada uno de los datos de la tabla: Edad, nota cálculo y nota informática. Lea los datos para cada vector desde el archivo "DatosCorrelacion.xlsx".

10. Realizar una gráfica de cálculo contra informática y edad contra cálculo. ¿Según las gráficas obtenidas, se puede inferir el valor al que va tender su correlación? ¿De ser posible, cuál es ese valor?

11. Calcular el coeficiente de correlación de Pearson y Spearman entre notas cálculo e notas informática, y entre edad y notas cálculo. ¿Hay una relación lineal en las notas de ambas materias? O ¿entre la edad y las notas de cálculo? ¿El resultado de ambas correlaciones concuerda con lo esperado en el punto anterior?

12. Explique. ¿Qué diferencia puede encontrar entre ambos tipos de correlación?

Entregables

- Informe en el Jupyter Notebook (archivo .ipynb) que contiene el código (solución a lo pedido), respuesta a las preguntas y las conclusiones del laboratorio.

Bibliografía

- <https://docs.scipy.org/doc/scipy-0.14.0/reference/generated/scipy.io.wavfile.read.html>
- <https://numpy.org/doc/stable/reference/generated/numpy.correlate.html>
- <https://cmdlinetips.com/2019/08/how-to-compute-pearson-and-spearman-correlation-in-python/>