

# Laboratorio 2

## Procesamiento Digital de Señales

1<sup>st</sup> Juan José Restrepo Rosero  
Pontificia Universidad Javeriana  
Cali  
Cali, Colombia  
[juanjorestrepo@javerianacali.edu.co](mailto:juanjorestrepo@javerianacali.edu.co)

2<sup>nd</sup> Juan David Vargas Mazuera  
Pontificia Universidad Javeriana  
Cali  
Cali, Colombia  
[juandavid2001@javerianacali.edu.co](mailto:juandavid2001@javerianacali.edu.co)

**Resumen—** En este laboratorio, se analizaron los efectos en frecuencia de una señal usando diferentes ventanas como: triangular, Hanning, y Hamming. Además, se construyó el espectrograma de una señal de audio, con el fin de identificar un fonema especial, por medio de la función de correlación.

**Palabras clave—** Efectos en frecuencia, Ventanas, espectrograma, Correlación.

### I. INTRODUCCIÓN

El desarrollo del laboratorio número 2 se centra en la creación de algoritmos y análisis de la señal seno en frecuencia con la implementación de diferentes ventanas como la triangular, la Hanning, y la Hamming.

Asimismo, analizó una señal de audio usando la librería Librosa, y construyendo un espectrograma. Se extrajo un pedazo del espectrograma el cual represente el fonema “él” de la frase “El coronel no tiene quien le escriba”, y se convirtieron ambas representaciones matriciales a un solo canal para facilitar y finalmente implementar la correlación entre ambas. Se debe de tener en cuenta que para el desarrollo del laboratorio se evalúan los resultados del programa 3 y 6 según ABET, y para la creación de estos algoritmos, el laboratorio brinda las bases y características principales que la función debe de cumplir

### II. OBJETIVOS

1. Estimar el espectro frecuencial de señales usando funciones de ventanas.
2. Diseñar, implementar y evaluar un algoritmo que detecte de manera automática en un espectrograma la ocurrencia de un fonema.

#### Marco Teórico

##### A. Correlación y autocorrelación

La correlación se puede definir como una operación entre dos señales que permite medir el grado de similitud que hay entre ellas, es decir, permite medir qué tan cercanas son dos señales mediante la convolución de estas. Una de las aplicaciones más comunes de la correlación, es para obtener características de una señal por medio de otra que es conocida [1].

$$r_{xy} = x(l) * y(-l) \quad (1)$$

La ecuación que describe la relación entre la correlación de dos señales  $x$ , y su convolución. Teniendo claro el concepto de correlación, la autocorrelación se puede definir como la correlación de una señal con ella misma. La autocorrelación se usa principalmente para identificar el periodo en una señal que se puede encontrar afectada por ruido.

$$r_{xy} = x(l) * x(-l) \quad (2)$$

Un dato importante es que cuando  $l=0$ , la correlación alcanza su máximo.

$$|r_{xy}(l)| \leq r_{xx}(0) = E_x \quad (3)$$

##### B. Transformada discreta directa de Fourier

La Transformada de Fourier es una operación matemática creada por Joseph Fourier alrededor del año 1747, la cual hoy es indispensable no solo en el procesamiento de señales, sino también en diferentes disciplinas. El término “transformada” se refiere al cambio de dominio, bien sea del dominio del tiempo al de frecuencia o viceversa, por métodos matemáticos. La forma matemática de la Transformada de Fourier tiene 2 ecuaciones, la de análisis, la cual lleva del dominio de tiempo a frecuencia:

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-jft} dt \quad (1)$$

Y la ecuación de síntesis, que lleva del dominio del tiempo a frecuencia se representa como:

$$X(t) = \frac{1}{2\pi} * \int_{-\infty}^{\infty} x(f)e^{jft} df \quad (2)$$

##### C. Propiedad de traslación en el dominio de la frecuencia

La propiedad de traslación en frecuencia es una de las más útiles al querer realizar procesamiento de señales en el dominio de la frecuencia, especialmente si se quiere hacer una traslación de frecuencia. Esta propiedad relaciona una multiplicación de una exponencial compleja de la señal en el dominio del tiempo,

con el efecto que tiene esta en el dominio de la frecuencia de la siguiente forma [1].

### III. PROCEDIMIENTO

El trabajo consiste en dos etapas. La primera consiste en procesar una señal usando ventanas triangulares, Hanning, y Hamming. La segunda consiste en procesar el espectrograma de una señal de audio para hallar los instantes dónde se encuentra un fonema que determinaremos.

Para el desarrollo de este laboratorio se utilizó como lenguaje de programación Python, el cual, a lo largo de los años ha tenido un constante crecimiento logrando ser uno de los lenguajes más utilizados en aplicaciones como análisis de datos o Procesamiento Digital de señales. Además, Python cuenta con gran cantidad de librerías que hacen que las tareas complejas sean cada vez más fáciles; un ejemplo de estas es Matplotlib, que contiene funciones para aplicaciones graficas.

#### A. Truncamiento de señales

Utilizar ventanas de suavizado para mejorar las características espectrales de una señal muestreada. Al realizar un análisis de Fourier o espectral en datos de longitud finita, puede usar ventanas de suavizado para minimizar las discontinuidades de las formas de onda truncadas, reduciendo así la fuga espectral. La cantidad de fuga espectral depende de la amplitud de la discontinuidad. A medida que se aumenta la discontinuidad, aumenta la fuga espectral (*spectral leakage*) y viceversa. Las ventanas de suavizado reducen la amplitud de las discontinuidades en los límites de cada período y actúan como filtros predefinidos de paso bajo de banda estrecha.

El proceso de “ventaneo” de una señal implica multiplicar el registro de tiempo por una ventana de suavizado de longitud finita cuya amplitud varía suave y gradualmente hacia cero en los bordes. La duración, o intervalo de tiempo, de una ventana de suavizado se define en términos de número de muestras. La multiplicación en el dominio del tiempo es equivalente a la convolución en el dominio de la frecuencia. Por lo tanto, el espectro de la señal con ventana es una convolución del espectro de la señal original con el espectro de la ventana de suavizado.

Inicialmente, se pedía graficar las funciones y los espectros frecuenciales de las ventanas triangular, Hanning y Hamming. Los resultados se ven a continuación:

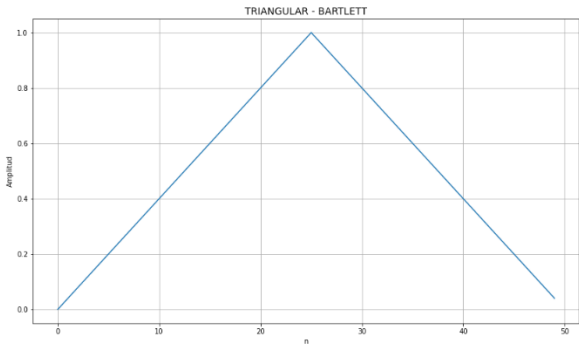


Figura. 1: Ventana Triangular – Bartlett

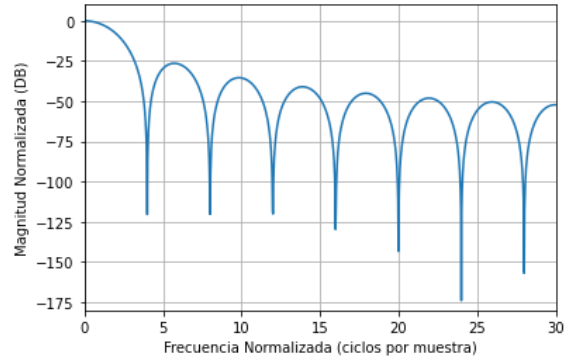


Figura. 2: Respuesta en frecuencia ventana Triangular

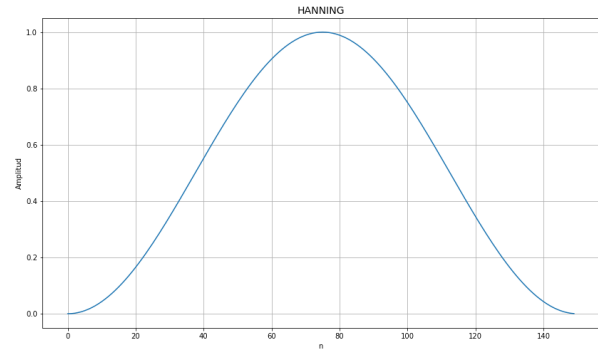


Figura. 3: Ventana Hanning

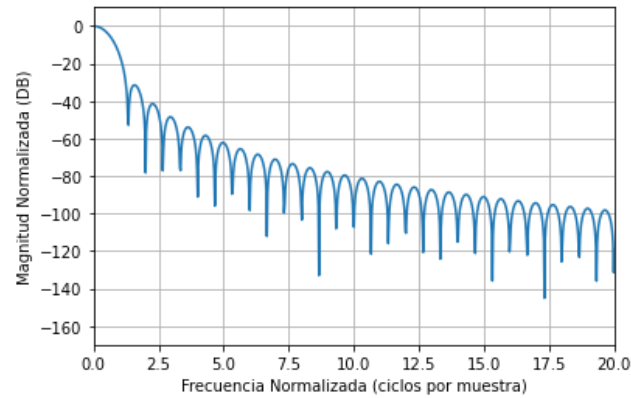


Figura. 4: Respuesta en frecuencia ventana Hanning

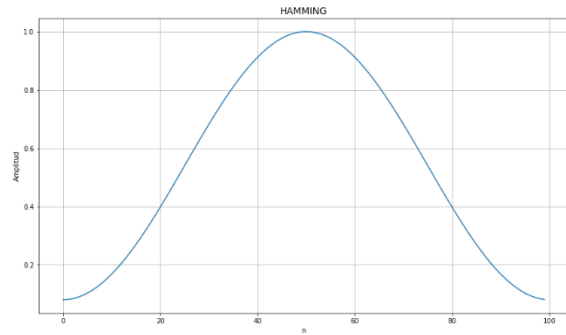


Figura. 5: Ventana Hamming

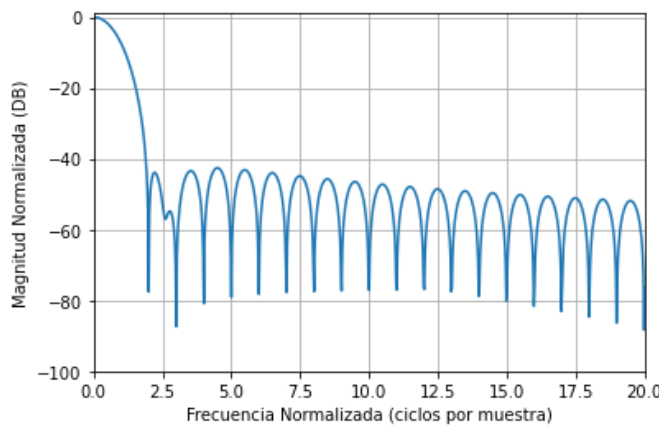


Figura. 6: Respuesta en frecuencia ventana Hamming.

Una vez obtenidos los espectros frecuenciales de las tres ventanas anteriores, se procede a calcular el espectro frecuencia de una función senoide mediante el truncamiento con las ventanas. La función es la siguiente:

$$x(n) = 2 \sin(2000\pi \frac{n}{8000}) \quad (3)$$

Para realizarlo, primero muestreamos la señal sinusoidal con una frecuencia de muestreo  $f_{sampling} = 8000 \text{ Hz}$ . La señal resultante es la que se muestra en la siguiente figura:

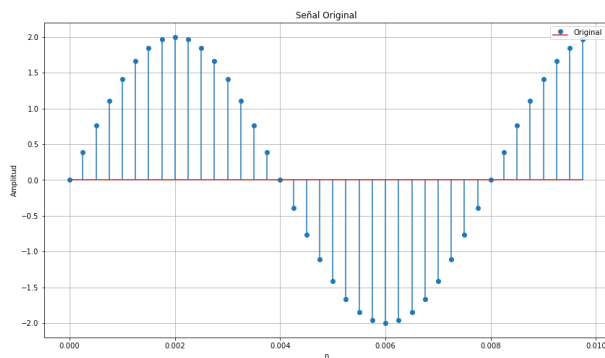


Figura 7. Señal sinusoidal muestreada.

Luego se calcula la transformada discreta de Fourier haciendo uso de la rutina `fft` y se graficar el espectro en frecuencia. De igual forma, se realizó una función llamada `TDF_directa()` para calcular la transformada, al igual que la parte real, imaginaria y fase. Las graficas resultantes se pueden ver en las siguientes figuras:

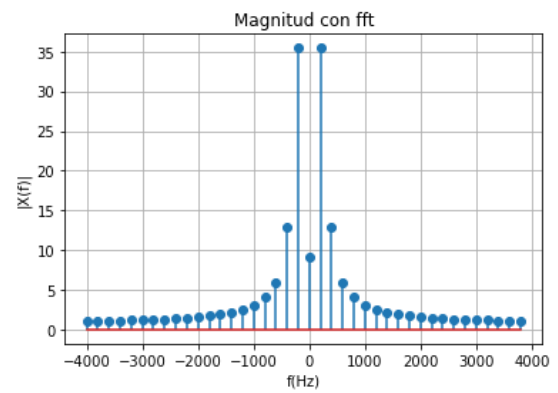


Figura 7. Espectro frecuencia obtenido con la rutina `fft` de Python

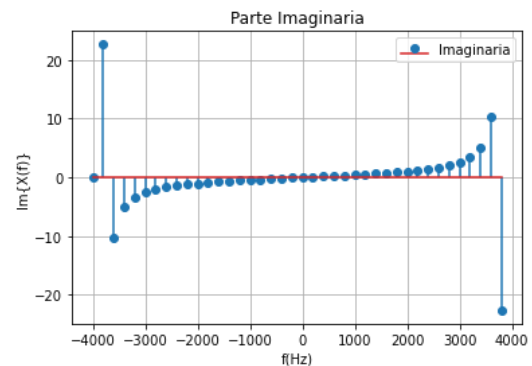


Figura 8. Parte imaginaria de la transformada de la señal sinusoidal

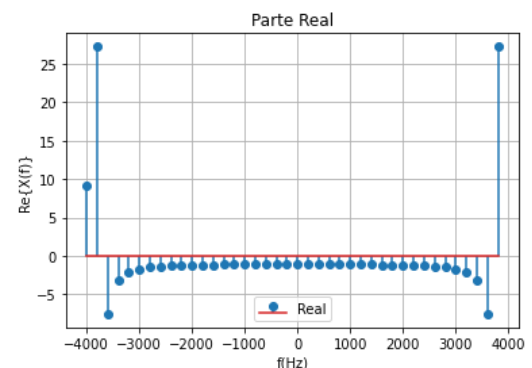


Figura 9. Parte real de la transformada de la señal sinusoidal

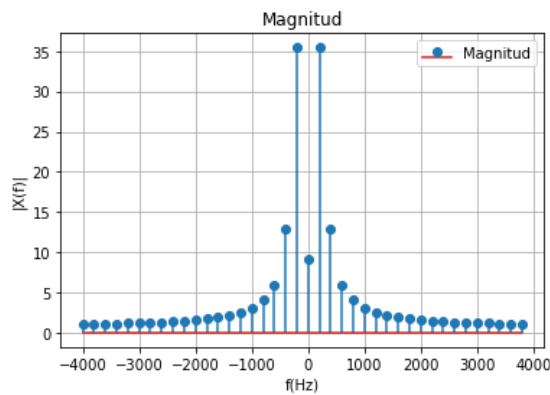


Figura 10. Espectro frecuencial de la transformada de la señal sinusoidal con función *TDF\_directa()*

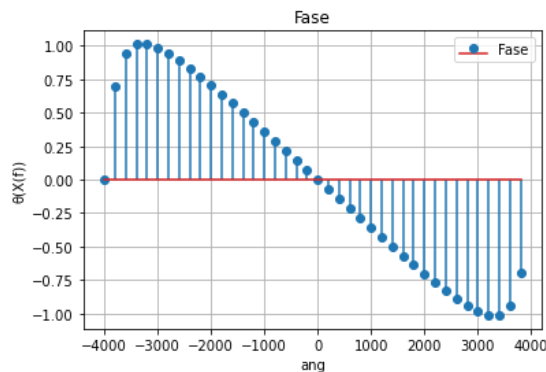


Figura 11. Fase de la transformada de la señal sinusoidal

### B. Identificación de un fonema en una señal de audio

Se requiere identificar un fonema especial en el espectrograma de una señal de audio. La señal de audio es grabada y posteriormente transformada a formato .wav para su manejo por parte de la librería Librosa. Esta señal contiene la grabación de voz de la frase “el coronel no tiene quien le escriba”, dónde el fonema elegido corresponde el sonido de la palabra “el”, la cual corresponde a los primeros segundos de grabación. Para este objetivo, se siguió el procedimiento mostrado en la figura 12.

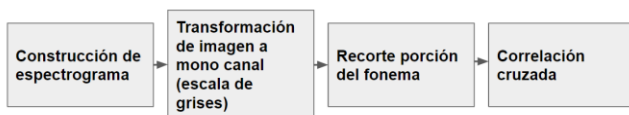


Figura 12. Procedimiento seguido para la identificación del fonema en el espectrograma

Primero que todo, para la obtención del espectrograma se usó la “Short Time Fourier Transform”, para la cual se usó una frecuencia de muestreo de 16,000 Hz, lo cuál puede ser corroborado con el software Audacity, dónde la máxima frecuencia es de 10,000 Hz, valor que se usa para calcular la frecuencia de muestreo, la cual suele ser el doble de la frecuencia máxima, así, en este caso sería 20,000 Hz la

frecuencia de muestreo. En la figura 13 se puede visualizar el espectrograma conseguido con Audacity, mientras que en la figura 14 se puede ver el espectrograma conseguido con Python y Librosa.

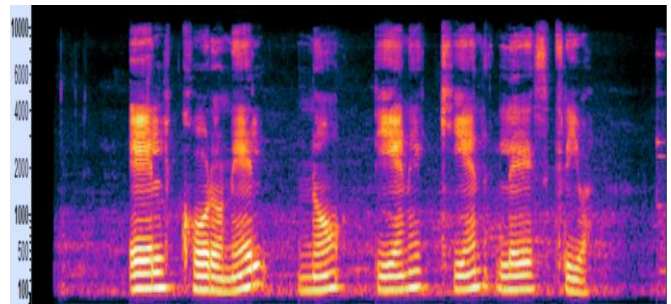


Figura 13. Espectrograma de la señal de audio conseguida con Audacity

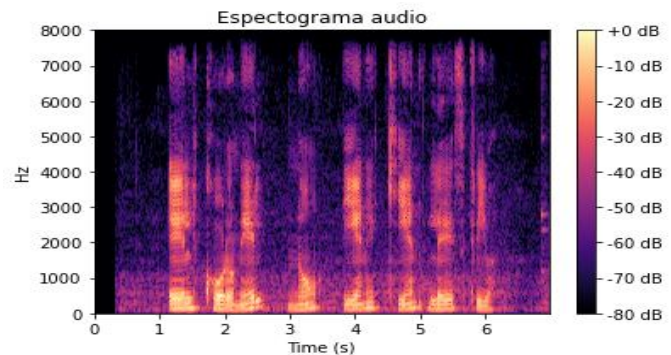


Figura 14. Espectrograma de la señal de audio conseguida con Librosa

Al visualizar esto, es notorio que ambos resultados son muy similares, lo que da tranquilidad por poder continuar con el procedimiento. Es notorio, además, que el resultado del espectrograma es congruente con la grabación de la frase, debido a que se resaltan los momentos dónde cada fonema es dicho. Por ejemplo, el fonema escogido fue escuchado a detalle, pudiendo notar que comprende aproximadamente lo segundos de 1 segundo a 1,5 a aproximadamente, pudiendo esto ser visto con claridad en el espectrograma.

Posteriormente, se procedió a guardar la imagen, recortar, y convertir la representación matricial de la imagen a un solo canal. Esto debido a que la representación retornada por Librosa es Quadmesh, la cual no puede ser recorrida como una matriz. Al guardar, cargar, y recortar con la librería PIL (Python Imaging Library) se logró obtener la representación matricial de la imagen. Sin embargo, esta representación, debido al color, es en cuatro canales. Por esto, se usó igualmente una función de PIL, para convertir la imagen a escala de grises y un único canal. Al realizar esto, se procedió a recortar el pedazo correspondiente al fonema “el”. Para dicho objetivo se observó el espectrograma, y se corroboró el periodo de tiempo escuchando el audio. Se recortó un recuadro de 50 x 50 pixeles

correspondientes al intervalo dónde se ve más resaltado en el espectrograma.

Posterior a esto se usó un código propio para encontrar la correlación entre el espectrograma y el recuadro correspondiente al fonema.

IV. RESULTADOS

A. Análisis en frecuencia de una señal usando ventanas triangular, Hanning, y Hamming

Después de haber obtenido los espectros en frecuencia de la señal seno, se procedió a hacer el truncamiento de la señal con cada una de las tres ventanas. Cabe aclarar que para que esto fuera posible, se realizó un redimensionado de la longitud de la señal seno, mediante la variación de los ciclos para que tuviese el mismo tamaño que las ventanas. Para el caso de la ventana Triangular, se tomó con un tamaño de 50 muestras, con la Hamming 100 muestras y la Hanning 150 muestras. El truncamiento resultante en el dominio del tiempo y de la frecuencia se puede visualizar en las siguientes figuras:

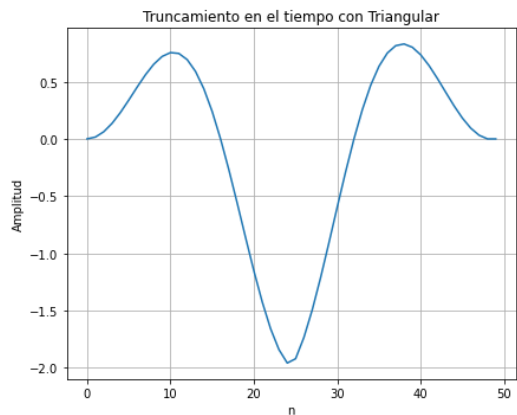


Figura 15. Truncamiento con la ventana triangular en el dominio del tiempo

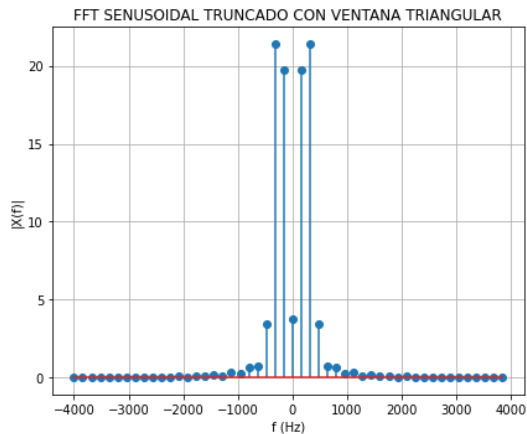


Figura 16. Truncamiento con la ventana triangular en el dominio frecuencial

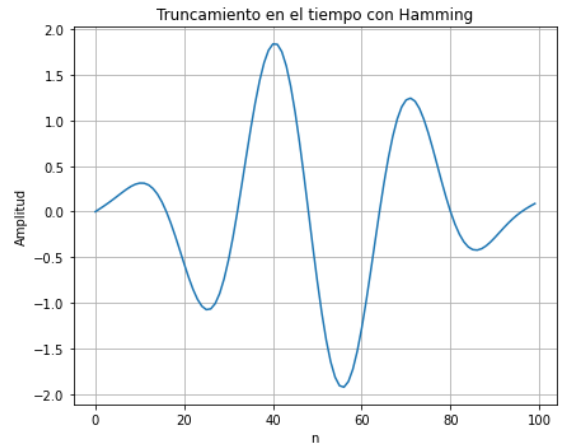


Figura 17. Truncamiento con la ventana Hamming en el dominio del tiempo

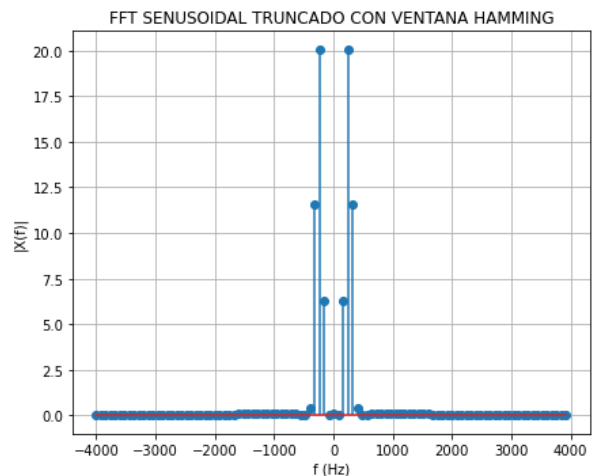


Figura 18. Truncamiento con la ventana Hamming en el dominio frecuencial

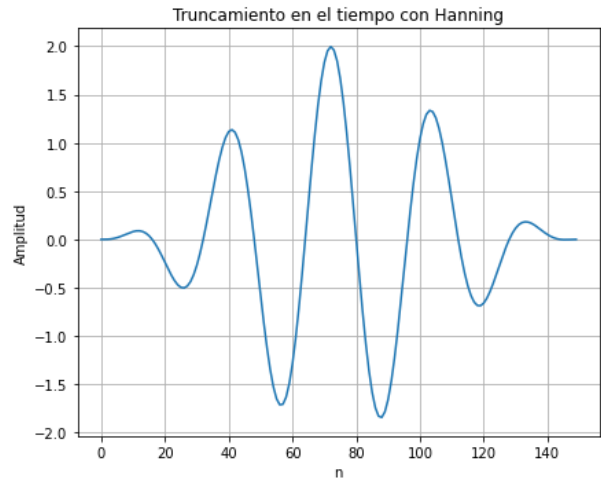


Figura 19. Truncamiento con la ventana Hanning en el dominio del tiempo



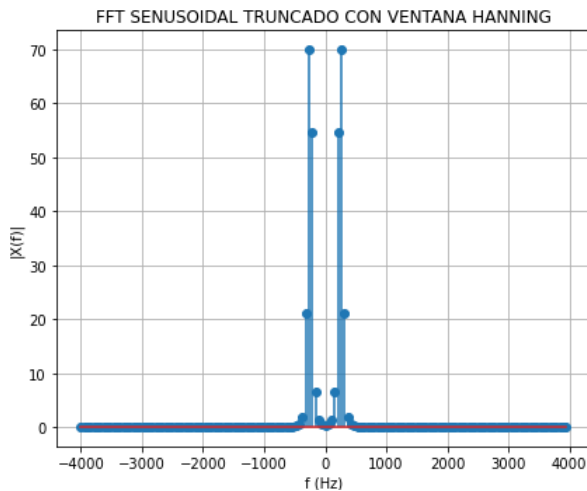


Figura 20. Truncamiento con la ventana Hanning en el dominio frecuencial

Al comparar los espectros de frecuencia entre las ventanas, es posible darse cuenta que a pesar que las ventanas Hamming y Hanning se ven muy similares, la Hanning tiende a representar mejor el espectro, es decir, con una mayor resolución, al igual que en la gráfica del dominio temporal y esto se debe a que la ventana de Hamming tiene menos términos, lo que termina influyendo en el cálculo del espectro con una menor precisión, por el hecho de tener 50 términos adicionales en el caso de la Hanning.

Por otro lado, esto también se ve reflejado en el espectro de la ventana rectangular, en donde si se aumenta el tamaño de esta, hay una mejor resolución del espectro. No obstante, la ventana rectangular generalmente se usa cuando desea procesar una porción más pequeña de un vector de datos mucho más grande, se debe hacer un redimensionamiento o rellanado con ceros con el fin

En el dominio de la frecuencia, una ventana rectangular convoluciona una señal más larga con una función Sinc, lo que produce artefactos de ventana que a veces se denominan "fugas" espectrales si la señal más larga no es estrictamente/exactamente periódica entera en la longitud de la apertura de la ventana, lo que significa que su espectro se distorsionó y la energía originalmente concentrada en  $f_0$  se dispersó en dos bandas de frecuencia más amplias.

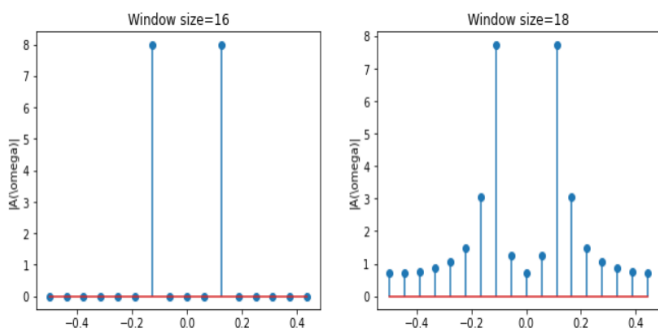


Figura 21. Espectro frecuencial ventana rectangular

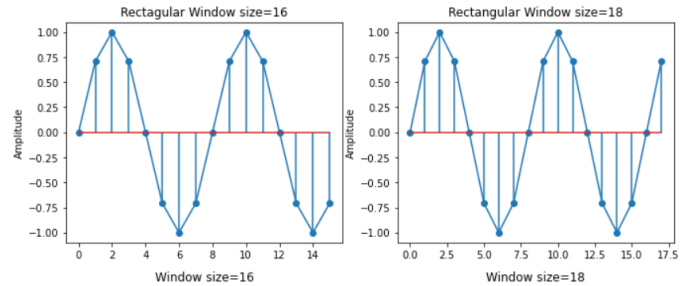


Figura 22. Truncamiento con ventana rectangular en el dominio del tiempo.

Por el teorema de muestreo se puede saber que, no importa cuán alta sea la frecuencia de muestreo, siempre que la señal se trunca, inevitablemente causará *aliasing*, por lo tanto, el truncamiento de la señal conducirá inevitablemente a algunos errores.

### B. Identificación de un fonema en una señal de audio

La primera etapa del procesamiento consiste en convertir la imagen del espectrograma a una imagen a escala de grises y a un solo canal, esto para facilitar el procesamiento por medio de la correlación. En la figura 10 se observa el espectrograma en escala de grises.

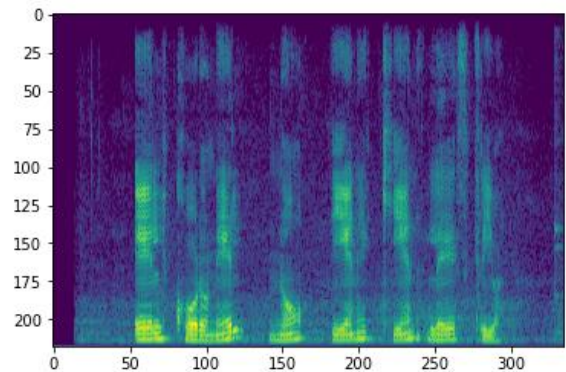


Figura 10. Espectrograma de la señal de audio en un solo canal(imagen)

De aquí, se procedió a calcular los recuadros correspondientes al intervalo dónde se produce el fonema "el". Al hacer la correspondencia entre los segundos dónde esto transcurrió y el número de píxeles del espectrograma, se consiguió recortar el fonema usando la librería PIL. La figura 11 muestra el recorte.

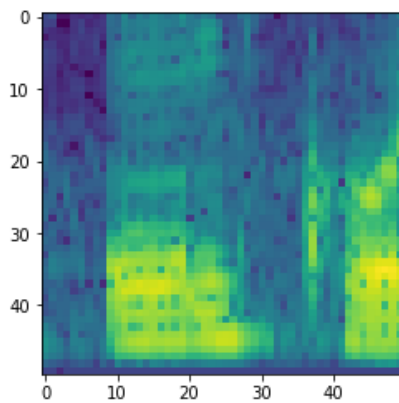


Figura 11. Parte del espectrograma en un solo canal correspondiente al fonema “el” (imagen)

Finalmente, se construyó un programa de correlación entre dos matrices. Este programa consigue realizar la operación de correlación, dónde se usó un “zero padding” para lograr que no hubiera errores al correlacionar los pixeles más cercanos a los bordes. Este relleno usando ceros produce una imagen, la cual se muestra en la figura 12.

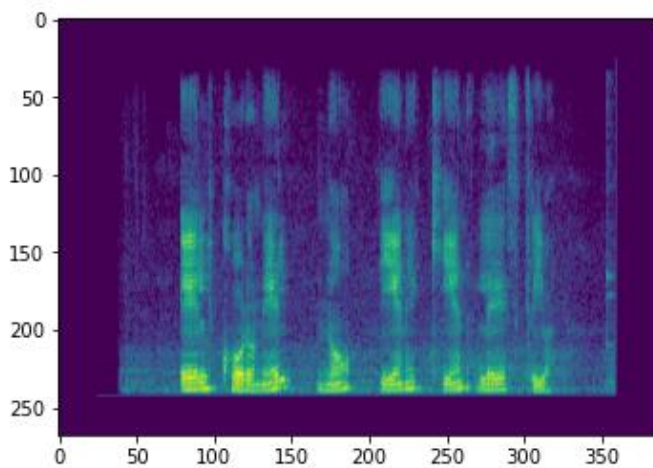


Figura 12. Espectrograma de la señal de audio después de proceso de “zero padding”

A esto, finalmente se hace la correlación entre ambas imágenes, lo cual produce el resultado mostrado en la figura 13.

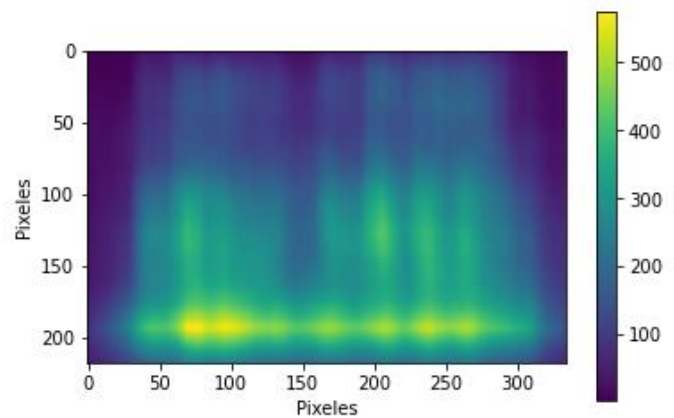


Figura 13. Correlación entre el espectrograma de la señal de audio y del fonema

Al observar estos resultados, se evidencia como efectivamente se ilumina principalmente y con más intensidad el intervalo de dónde se eligió el fonema. Además, se evidencia cierta correlación con los intervalos dónde se dijeron los otros fonemas, aunque sin tanta intensidad. Esto tiene sentido, debido a que comparten similitudes, al ser ruidos producidos por la voz humana. Sin embargo, tiene diferencias en el sonido, entonación, u otros que sean características de cada fonema, en este caso “el”.

Se realizó, además, un programa que detecta el máximo absoluto y evalúa los puntos en su cercanía, hasta 25 puntos de distancia desde el punto máximo, debido a que esto construiría una cuadrícula 50 x 50 que es del mismo tamaño del pedazo del espectrograma extraído para el fonema. Este programa acota los puntos y los incluye como el segmento en el que se detectó la correlación si corresponde a un valor de al menos el 90% del valor máximo detectado de correlación. Haciendo este procedimiento, y hallando los externos de estos puntos, se obtuvo que la correlación se detectó principalmente entre 1,4 segundos y 1,96 segundos. Además, se detectó entre 770 Hz y 1,14 kHz.

## V. CONCLUSIONES

### A. Análisis en frecuencia de una señal usando ventanas triangular, Hanning, y Hamming

Luego de haber realizado el truncamiento con las tres ventanas y de comparar sus espectros en frecuencia con los de la ventana rectangular, se puede apreciar que además de notar que la ventana Hamming tiene menos términos que la Hanning, existen mayor cantidad de lóbulos en las dos ventanas anteriores en comparación con la triangular. Si estos se reducen, se aumentaría la eficiencia del ventaneo, lo que significaría una menor pérdida de energía y una mayor precisión a la salida. De igual manera, se puede apreciar que la ventana rectangular tiende a provocar más “Spectral Leakage” a la salida que las demás.

### B. Identificación de un fonema en una señal de audio

Al analizar el espectrograma de una señal de audio, es posible reconocer los diferentes momentos de distorsión, en este caso cuando se pronuncia cada fonema. Además, se puede visualizar los rangos de frecuencia donde predominantemente se entona cada fonema. Además, es posible observar la semejanza de los resultados de espectrograma de programas especializados como Audacity, así como al usar la librería Librosa. Al realizar la correlación de la señal, se identifica con mayor intensidad el intervalo dónde se extrae el fonema, de forma que es resulta sencillo, usando una ventana del espectrograma de un fonema, identificar a cuál momento de la señal de audio pertenece. Igualmente, se observa que se resaltan, aunque en menor medida, los otros momentos dónde fueron entonados los otros fonemas, lo cual tiene sentido ya que al hallar la correlación se haya los instantes que tienen una mayor similitud.

## VI. DISTRIBUCIÓN DEL TRABAJO

Cada miembro del equipo tiene experiencia en la realización de las distintas partes del laboratorio. Sin embargo, a continuación, se describirán los aportes distintivos de cada miembro:

**A. Juan José Restrepo:** Desarrolló el programa que da solución al problema el análisis en frecuencia usando las ventanas. Organizó el cuaderno en Jupyter.

**B. Juan David Vargas:** Desarrolló el programa que da solución al problema de la identificación del fonema en el espectrograma de una señal de audio.

Adicionalmente, cada miembro colaboró en el informe desde lo que realizó en Python. Los resultados fueron compartidos y en muchos casos se ejecutaron los códigos en medio de las reuniones de equipo.

## VII. REFERENCIAS

- [1] Benítez, H., 2022. Correlación de señales.
- [2] Benítez, H., 2022. Introducción.
- [3] PIL. (2019). *Crop a part of the image with Python, Pillow (trimming)* / *note.nkmk.me*. <https://note.nkmk.me/en/python-pillow-image-crop-trimming/>
- [4] Lewis, C. (2021). *How to Create & Understand Mel-Spectrograms* / by *Christopher Lewis* / *Medium*. <https://importchris.medium.com/how-to-create-understand-mel-spectrograms-ff7634991056>
- [5] MLK. (2021). *4 Ways to Convert Image to Grayscale in Python using Skimage, Pillow and OpenCV - MLK - Machine Learning Knowledge*. <https://machinelearningknowledge.ai/ways-to-convert-image-to-grayscale-in-python-using-skimage-pillow-and-opencv/>
- [6] MTU PHISYCS. (2018). Physics of music - notes. Frequencies for equal-tempered scale. [online] Available at: <https://pages.mtu.edu/~suits/notefreqs.html>.
- [7] Benítez, H., 2022. Transformada discreta de Fourier.
- [8] Benítez, H., 2022. Transformada rápida de fourier (FFT).