



66.74 - Señales y Sistemas
FIUBA

TRABAJO PRÁCTICO

Análisis y procesamiento de la
señal de habla

Tomás Augusto Bert - 95606

Introducción

El presente trabajo práctico se desarrolló en **Python 3.6** con las herramientas disponibles para el procesamiento de señales y cálculos matemáticos. Estas son:

- **Matplotlib**: Herramienta para realizar gráficos y visualizaciones
- **Numpy**: Paquete fundamental para la ciencia.
- **Scipy**: Un nivel más a numpy. Con herramientas para la ciencia, matemáticas e ingeniería
- **FIIR (<https://fiiir.com/>)**: Web interactiva para generar filtros con numpy en Python.

Para la grabación de sonidos y extracción de los fonemas se utilizo Audacity.

Ejercicios

1 - Grafique la señal de voz del archivo hh15.wav, ubicando en ella porciones de señales que se o correspondan con fonemas sonoros y sordos. Segmentar y etiquetar en forma aproximada cada uno de los fonemas presentes en la señal

Para segmentar los fonemas se editó el audio extendiendo el tiempo, de esta forma al reproducir el sonido la locutora habla de forma más lenta pudiendo interpretar qué letra está diciendo. Desde ahí, de forma, manual se marcaron los tiempos en lo que la locutora pronunciaba cada fonema.

En cuanto a los segmentos sonoros y sordos, se grafican líneas verticales con mayor intensidad en el color, dependiendo de la amplitud del sonido. De esta forma se obtienen amarillos más intensos cuando la amplitud es mayor y amarillos transparentes cuando la amplitud es menor.

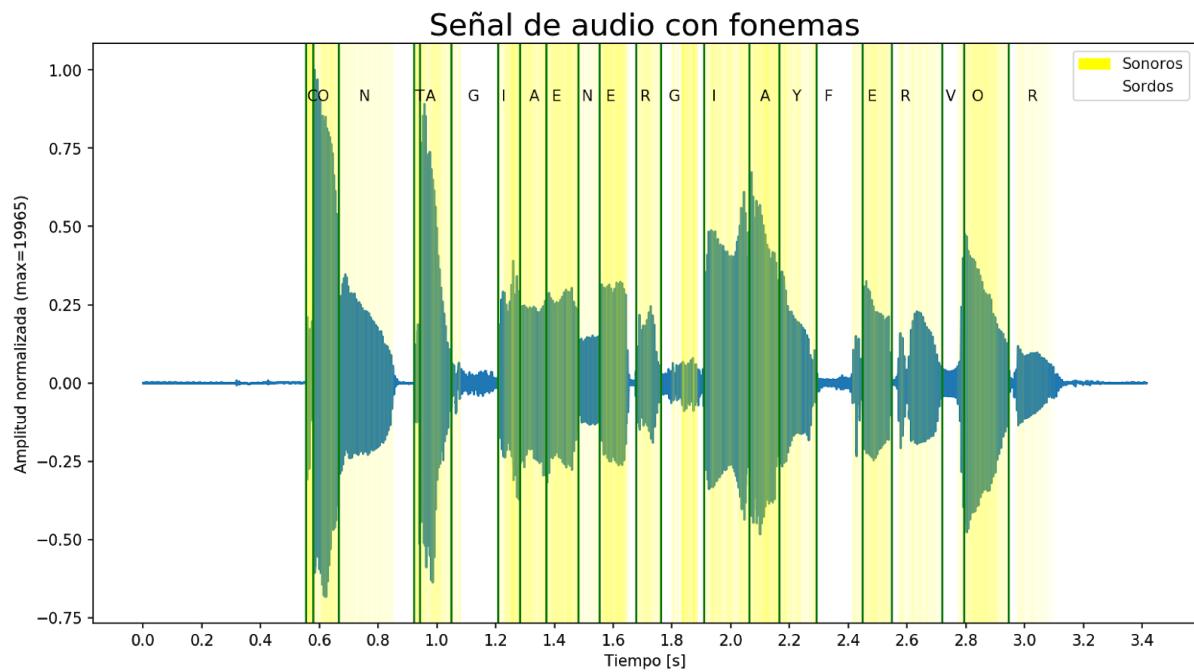


Figura 1. Señal de voz del archivo “hh15.WAV” con fonemas segmentados y etiquetados

2 - Con la segmentación realizada en el ejercicio 1 de la señal hh15.wav, encuentre los coeficientes de Fourier de un período del segmento de señal correspondiente a un fono [a]. Repetir el cálculo para varios períodos de la vocal.

Inicialmente se toman dos muestras de la vocal “A” y se marca un periodo. Puede notarse que gráficamente ambas señales no son similares. Pero comparando sus máximos y sus períodos pueden encontrarse similitud.

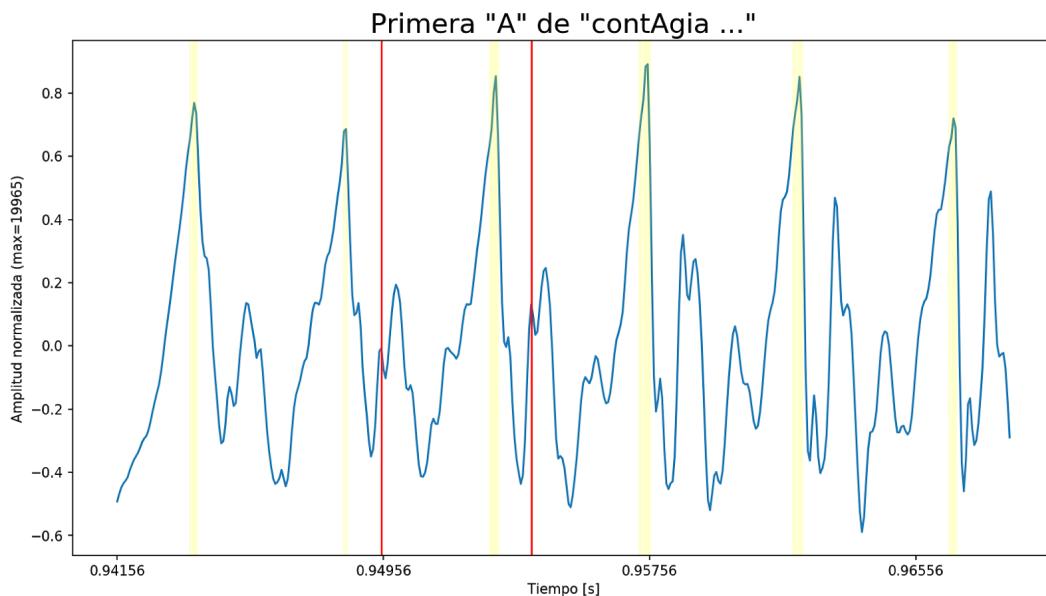


Figura 2. Señal del primer fono [a] proveniente de la locutora

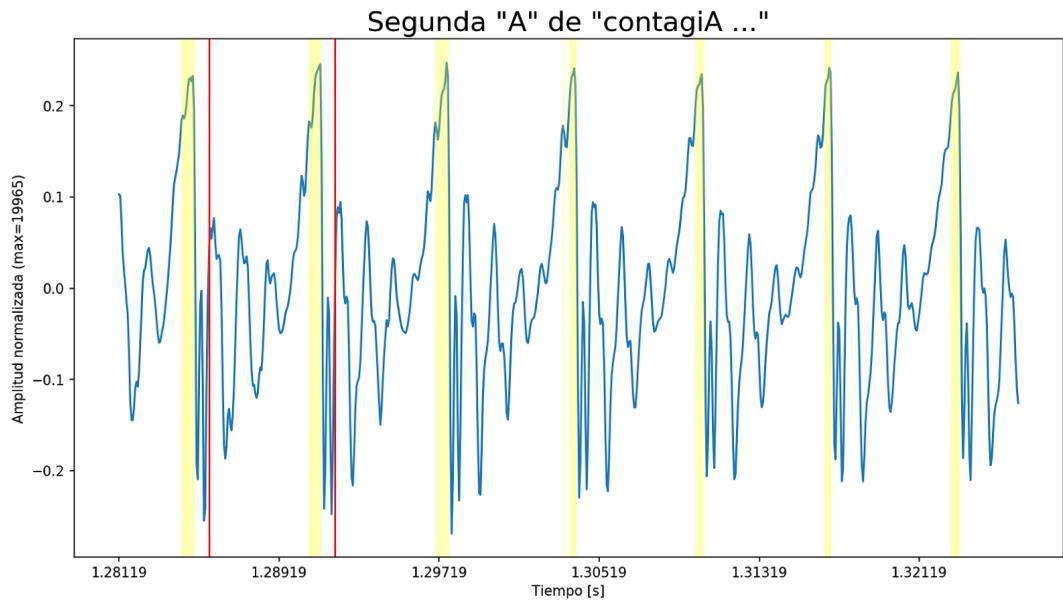


Figura 3. Señal del segundo fono [a] proveniente de la locutora

Para el análisis de los coeficientes de Fourier se toma la primera “A” de la palabra “Contagia”. Se busca el periodo de forma manual y se grafica. A continuación se ve el segundo periodo del fonema, el periodo es el marcado en la figura [n], entre las líneas verticales rojas.

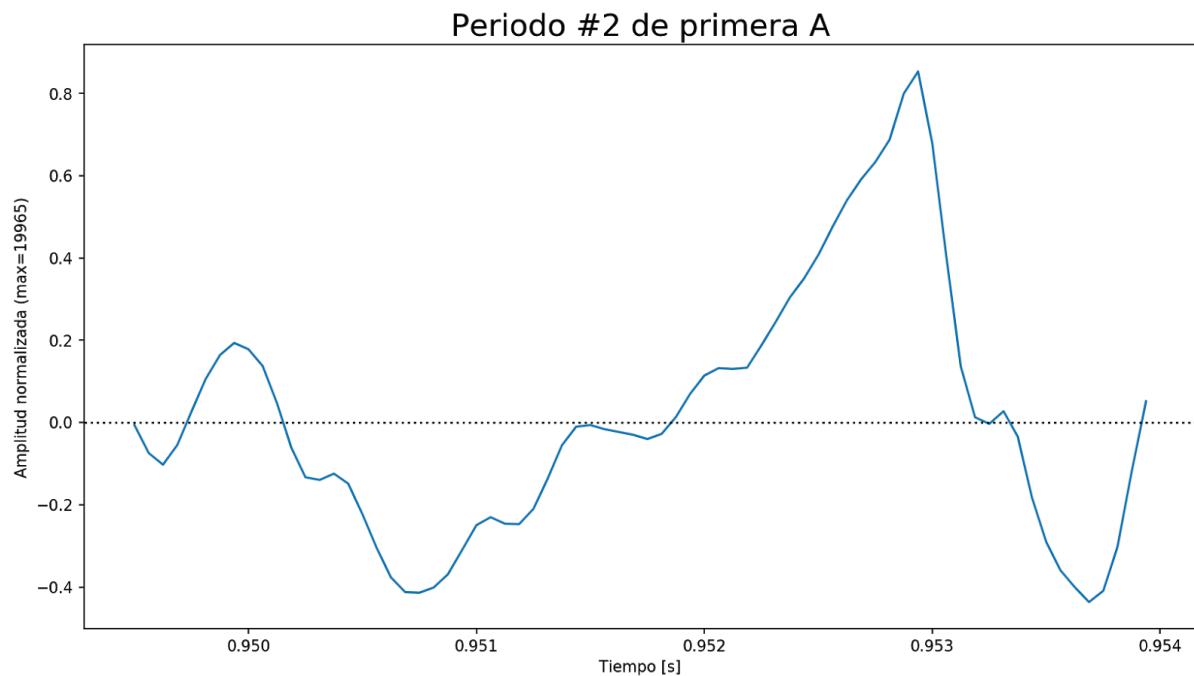


Figura 4. Periodo de la señal del primero fono [a] (Figura 2)

Una vez obtenido el periodo, se procede a realizar FFT de la señal. Obteniendo así los coeficientes de fourier. Puede observarse que en el rango aproximado entre $k=15$ y $k=65$ nos “ak” toman valores muy cercanos al cero.

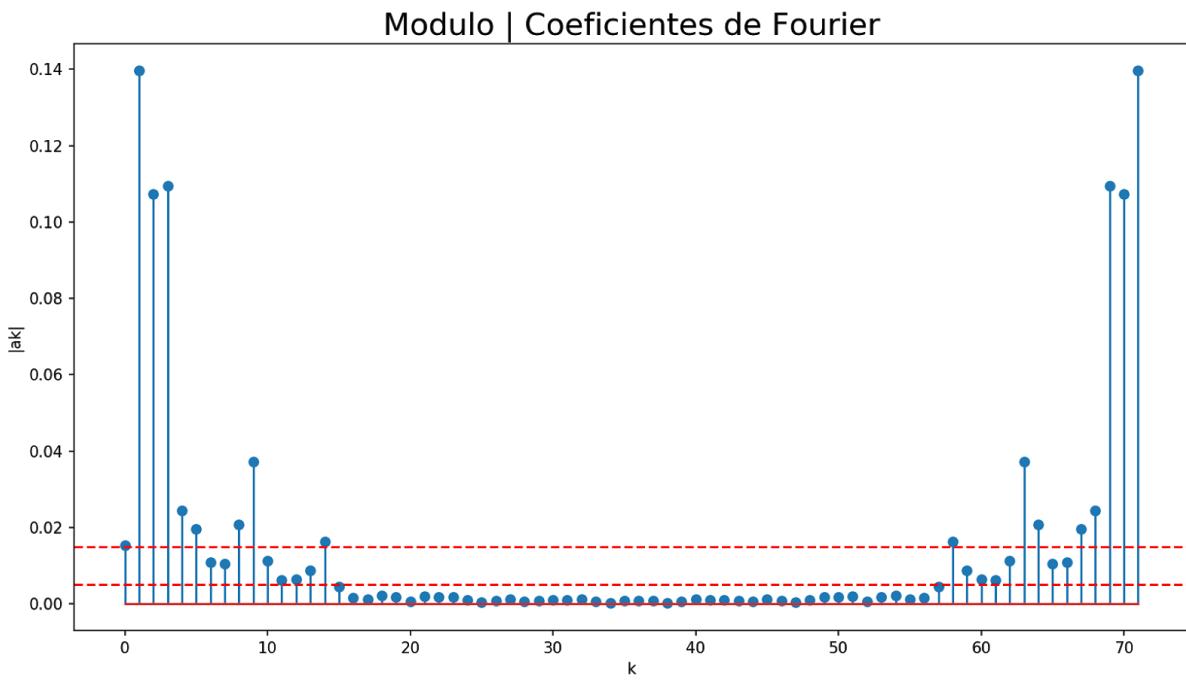


Figura 5. Coeficientes de Fourier de la señal de la Figura 4

También, se realiza el diagrama de fases de los coeficientes de fourier. La fase indica el ángulo de cada “ a_k ” ya que pertenece a los números imaginarios.

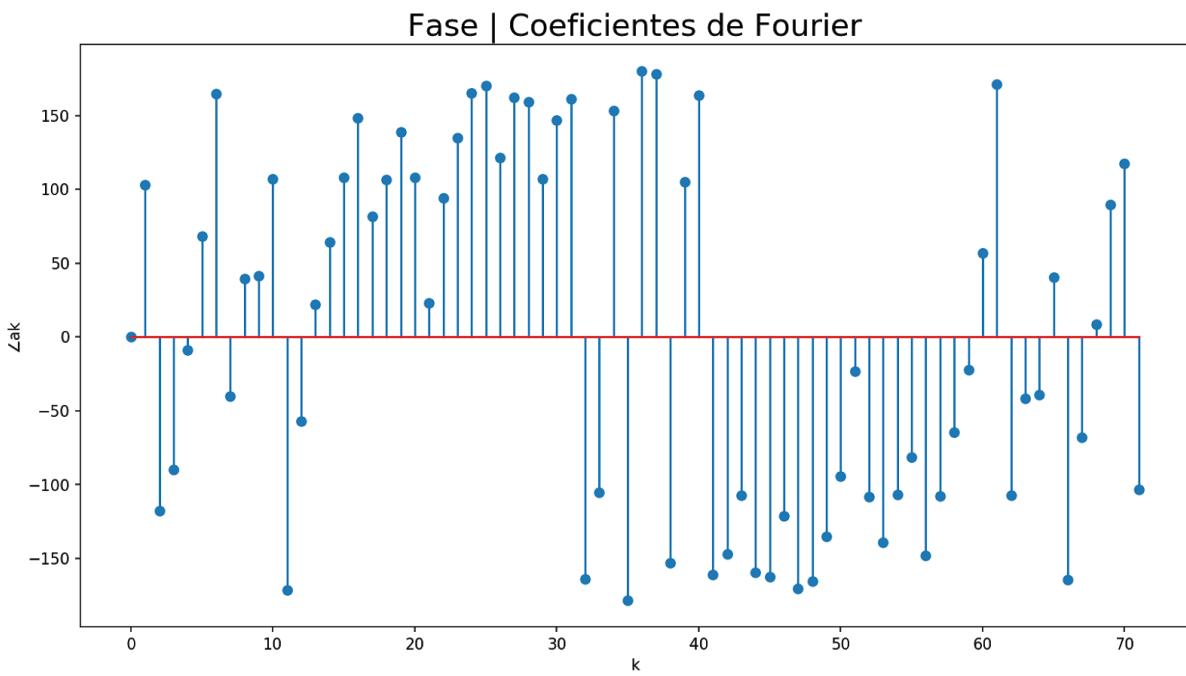


Figura 6. Fase de los coeficientes de Fourier de la Figura 5

3 - Reconstruya la señal temporal a partir de los coeficientes calculados. Escuche y compare las distintas reconstrucciones correspondientes a coeficientes de Fourier tomados de distintos períodos. Compárelas también con la señal original. ¿Qué observación se puede hacer sobre la periodicidad de los fonemas vocálicos?

A los resultados obtenidos en el punto anterior, se realiza la inversa de la DFT (*ifft*) y se lo concatena sucesivamente varias veces para que quede resultante un segmento con duración de aproximadamente dos segundos. Ese resultado se vuelva sobre un archivo WAV dentro de la carpeta “out”.

Para realizar distintas reconstrucciones se plantea utilizar un umbral para eliminar coeficientes cercanos al cero. Estos se pueden visualizar en la Figura 5. Los resultados obtenidos son distintas reconstrucciones de la vocal y la variación al eliminar coeficientes debajo de los umbrales es mínima, ya que auditivamente se distinguen silbidos de fondo en cada una de las reconstrucciones.

4 - Grabe la misma frase del ejercicio 1. Mencionar las diferencias entre ambas señales.

La grabación se realizó con el software Audacity. Se intentó conseguir una señal sincronizada respecto del audio original. A continuación se muestra el gráfico de las señales con los fonemas marcados tal como se realizó en el ejercicio 1.

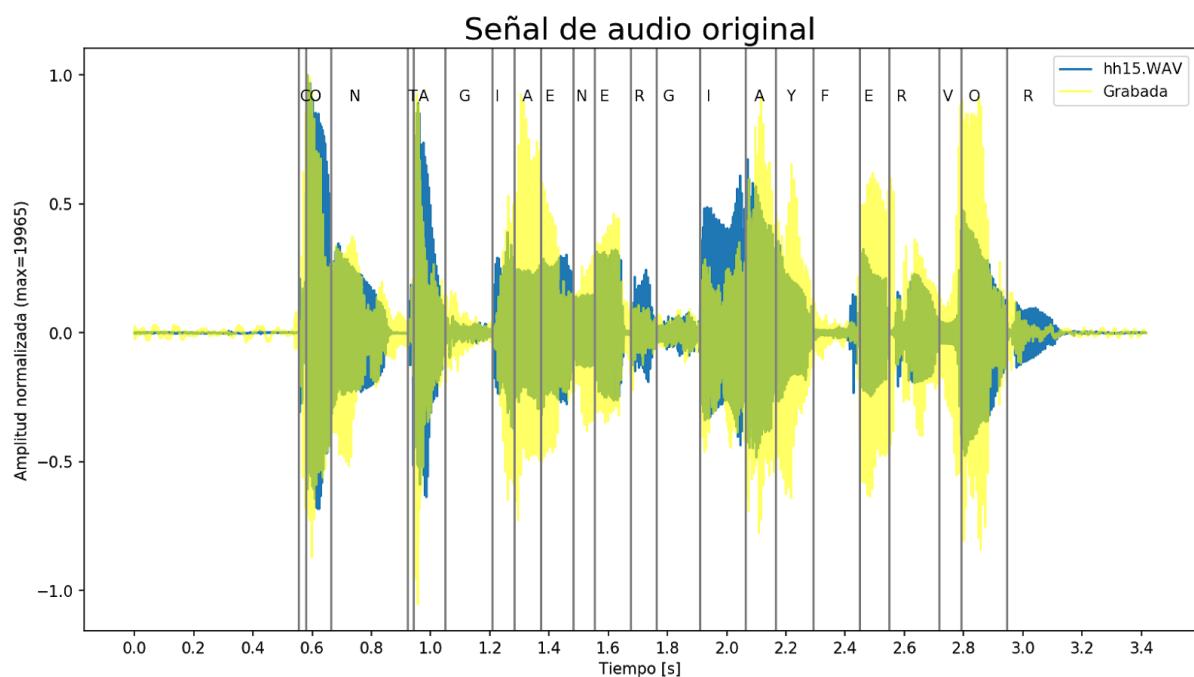


Figura 7. Comparación señal grabada y señal original

Puede notarse a primera vista la diferencia de amplitudes en las vocales. Por un lado la grabación muestra mayores amplitudes para las vocales “A”, “E”, “O”. Mientras que la “I” muestra menor amplitud en ambas apariciones.

La diferencia entre las señales se puede describir con las diferencias investigadas entre la voz del hombre y la mujer. Principalmente la diferencia fisiológica donde se describe que el tracto vocal común de una mujer es menor a la de un hombre, dando como resultado que el tono de voz de un hombre sea menor a la de una mujer. En estudios se ha observado que la mujer ronda los 200 Hz mientras que el hombre los 100 Hz.

5 - Grafique los espectrogramas de banda angosta de los segmentos de señal correspondientes a tres vocales presentes en la señal hh15.wav. Compare y analice las diferencias.

Para el análisis de espectrogramas voy a tomar las vocales “O”, “I” y “E”. banda angosta Se puede notar como el ancho de la ventana afecta la resolución temporal. Es decir, mientras más ancha, menor resolución temporal (horizontal) y mayor resolución espectral (vertical).

En cuanto al cálculo de la ventana se realiza a partir de la frecuencia de muestreo (FS), la duración de un periodo del fono (vocal) y la cantidad de periodos del mismo. Por ejemplo, para el fono “O”, se midió el periodo y resulta de una duración de 4.6 milisegundos, El fono está compuesto por aproximadamente 8 periodos (lo que se puede observar a ojo). Por lo tanto la ventana utilizada para realizar el espectrograma va a ser de:

$$T \cdot nT \cdot FS = 0.0046 \cdot 8 \cdot 16000 = 588$$

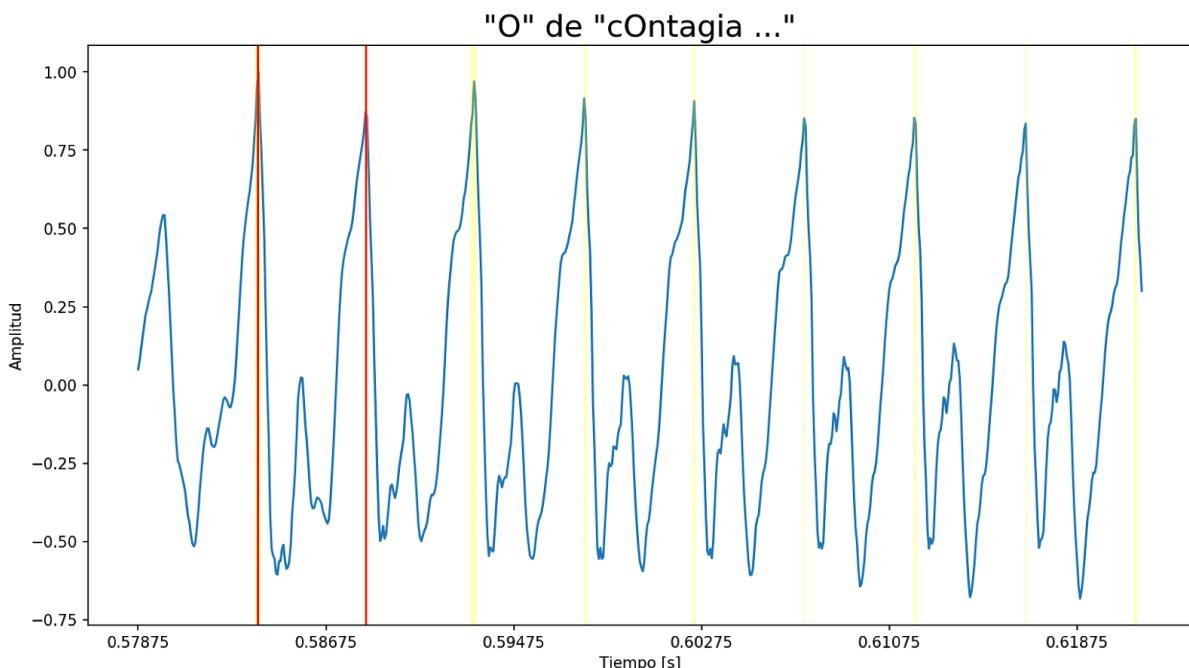


Figura 8.1 Señal del fono [o] correspondiente a la palabra “contagia”

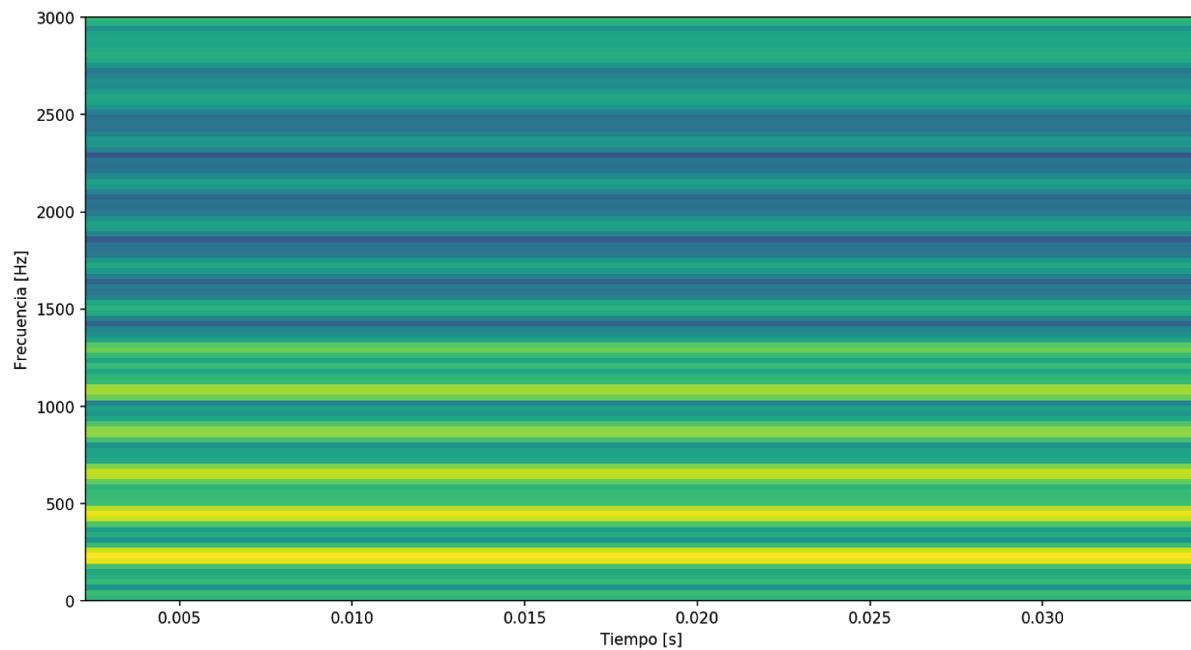


Figura 8.2 Espectro de banda angosta de la Figura 8.1

Para el análisis de la segunda vocal seleccionada, realizando los mismos cálculos que se realizaron en el fono anterior tenemos que el ancho de la ventana está definido por:

$$T \cdot nT \cdot FS = 0.005481 \cdot 16000 \cdot 6 = 526$$

Donde 5,5 ms es la duración del periodo y en la Figura 10.1 se pueden ver los 6 períodos.

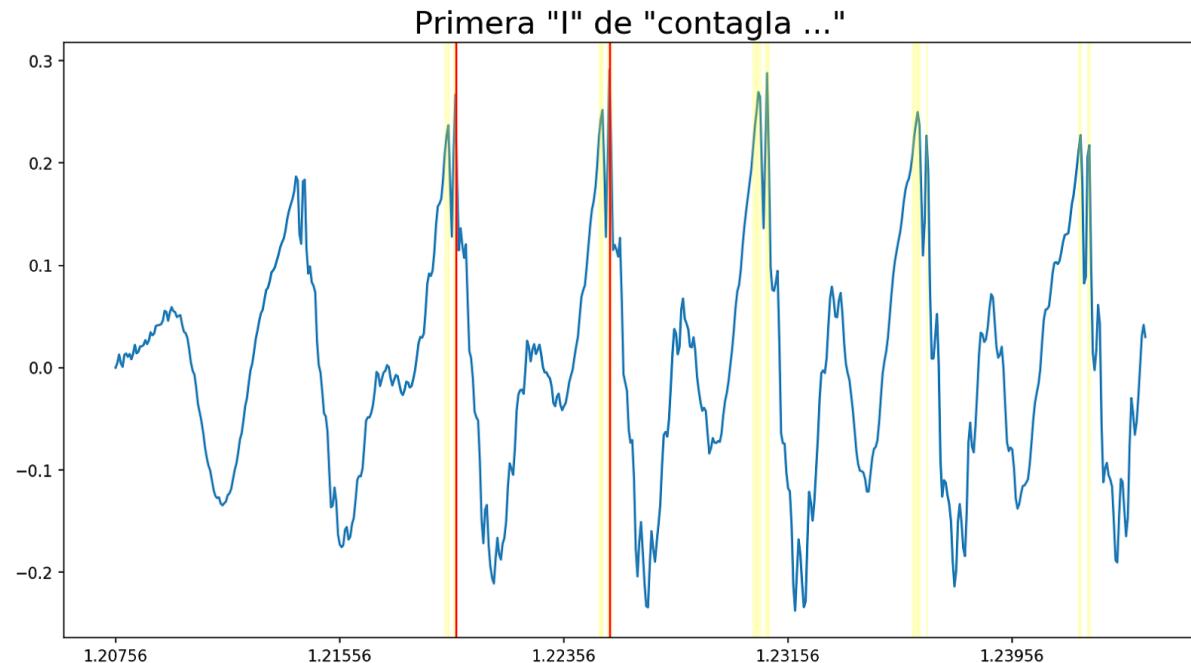


Figura 9.1. Señal del fono [i] correspondiente a la palabra “contagia”

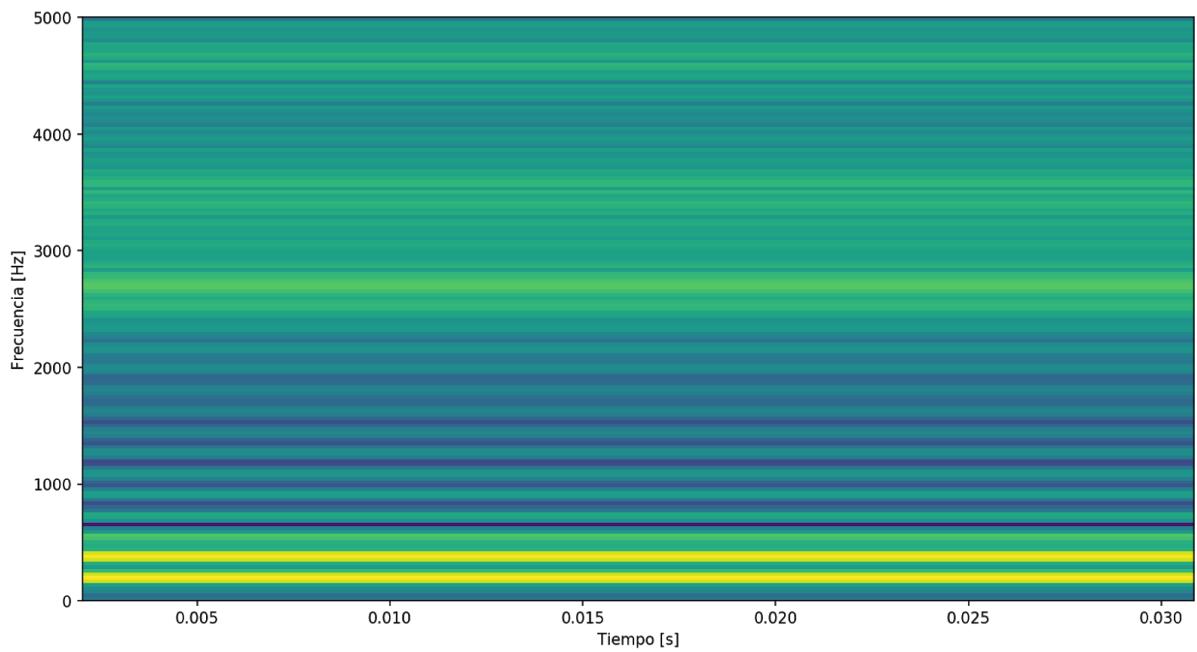


Figura 9.2 Espectro de banda angosta de la Figura 9.1

Por último, para la vocal E. El cálculo realizado para el tamaño de la ventana es:

$$T \cdot nT \cdot FS = 0.006281 \cdot 16000 \cdot 9 = 904$$

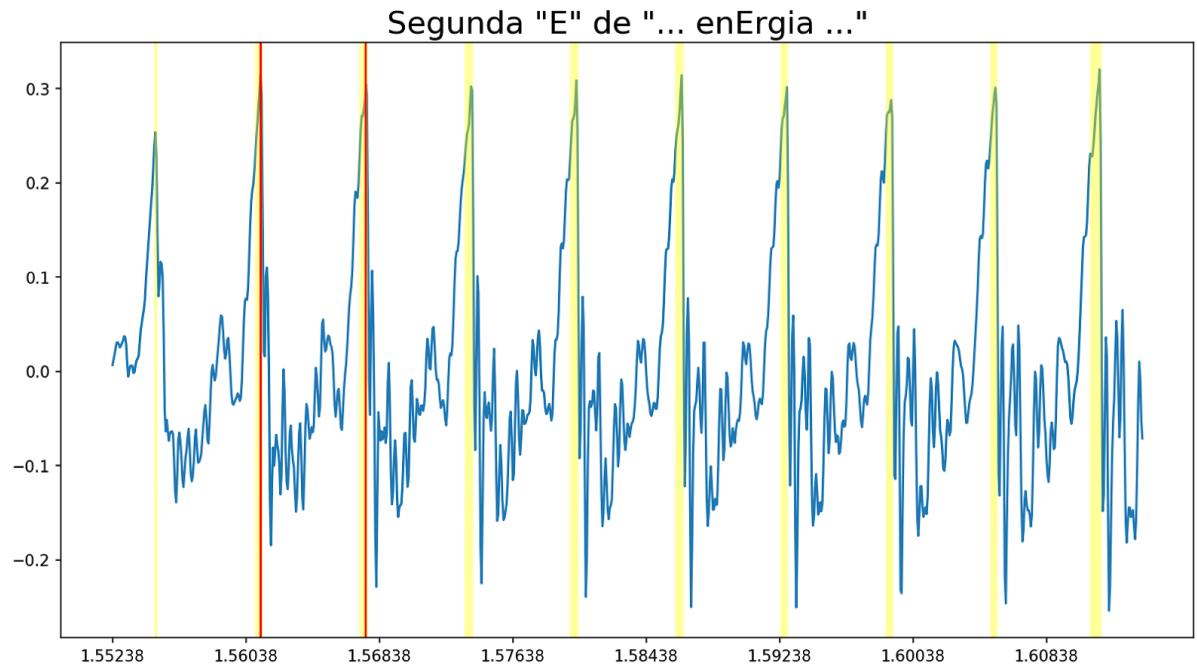


Figura 10.1 Señal del fono [i] correspondiente a la palabra “energia”

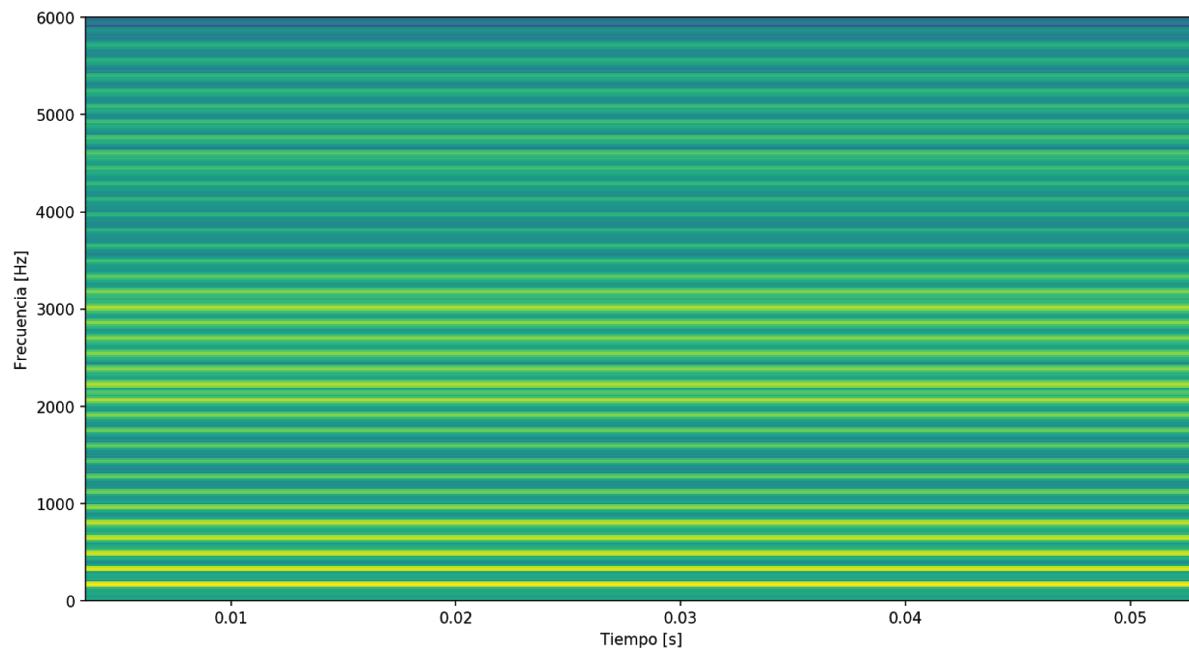


Figura 10.2. Espectro de banda angosta de la Figura 10.1

Se puede observar en los distintos espectrogramas las frecuencias predominantes para cada vocal. En el fono “O” el rango de frecuencias va desde 0 a 1500 Hz, en la “I” el rango está por debajo de los 500 Hz y en la vocal “E” las frecuencias están distribuidas en el rango 0 a 3000 Hz.

6 - Genere diez ciclos del tren de pulsos glóticos según los modelos de Rosenberg. Tomar una frecuencia $F_0 = 200$ Hz, y fases de apertura y cierre de 40% y 16%, respectivamente, de la duración de un pulso. Considerar una amplitud máxima de 1. A los efectos de la simulación, considerar una frecuencia de muestreo de 16 kHz. Estimar su espectro de amplitud y explicar su contenido. Grafique en forma superpuesta el espectro de un pulso y del tren de pulsos. Justifique los resultados observados.

Para el cálculo del tren de pulso glótico, se utiliza el modelo de Rosenberg. Ingresando los parámetros Amplitud 1, $T_p = 0.001975s$ y $T_n = 0.00079s$. El resultado obtenido es el siguiente:

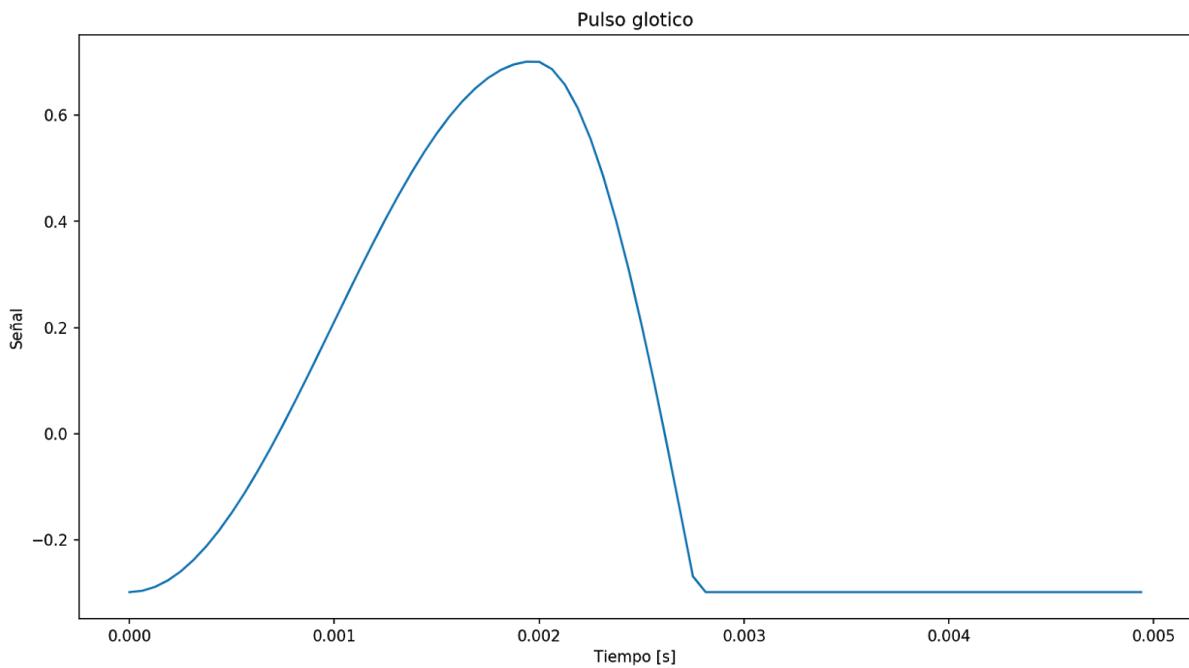


Figura 11. Pulso glótico con valor medio restado.

Para realizar el Tren de pulso glótico, se repite diez veces el resultado anterior. Obteniendo:

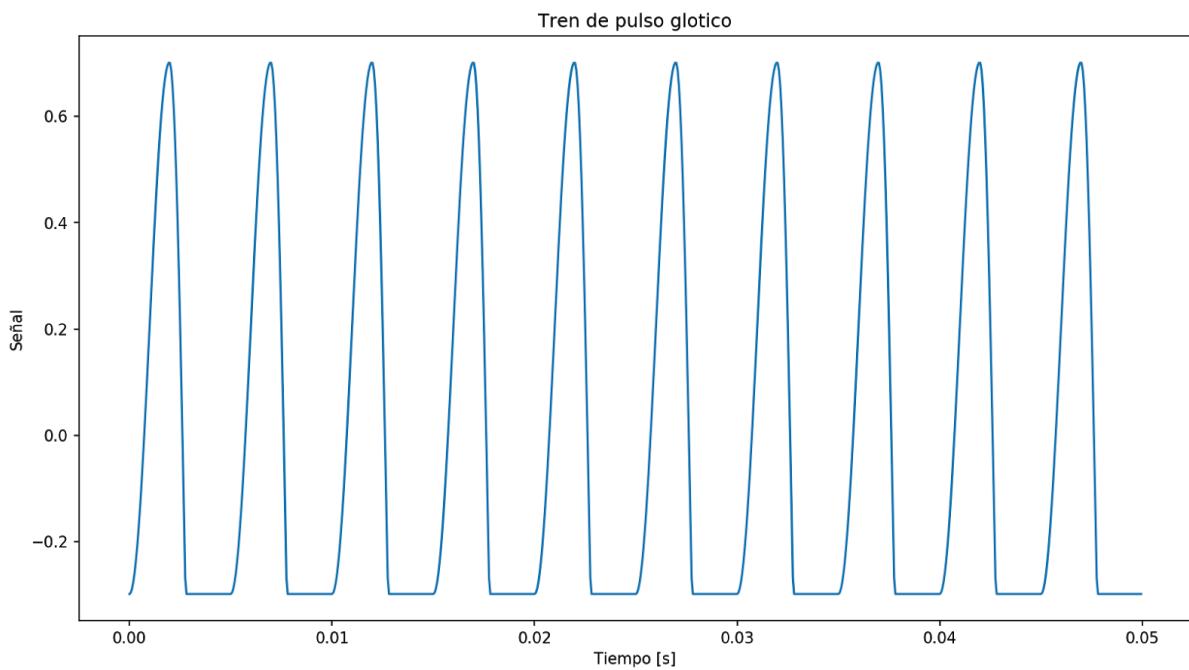


Figura 12. Tren de pulso glóticos (10 veces la figura 11)

A continuación se muestra el espectro de amplitud de ambas señales. Puede notarse como como se describe que el espectro de un periodo son deltas de dirac con envolvente superpuestas.

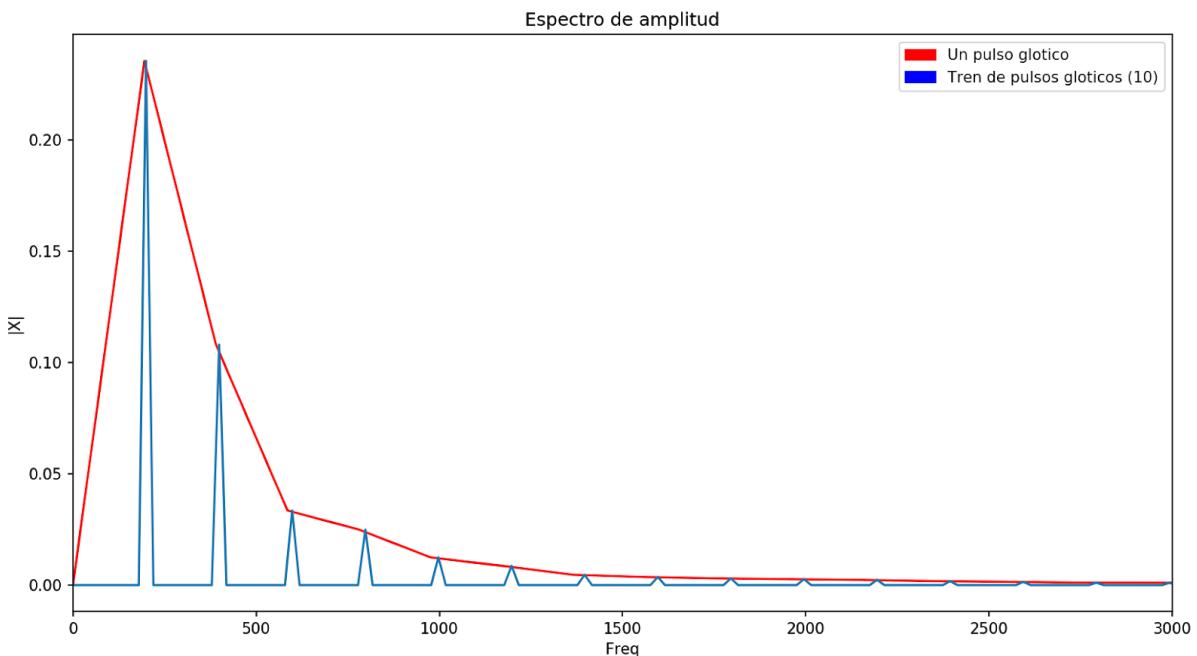


Figura 13. Espectro de amplitud correspondientes al pulso y al tren de pulsos.

7 - Utilizando las ec. 3 y 4, generar un modelo de tracto vocal para cada uno de los siguientes conjuntos de valores de parámetros, que se corresponde con una vocal emitida por una locutora. Graficar diagrama de polos y ceros, y la respuesta en frecuencia de cada vocal, compare.

Para el cálculo de la respuesta en frecuencia de cada vocal se itero cada n (F_n , B_n) de cada vocal de la tabla. Para cada par F_n - B_n se calculó el tracto vocal (H) y por último se multiplicaron entre ellas, es decir, se genera una respuesta en frecuencia total que son las respuesta en frecuencia de cada n en cascada. El resultado obtenido es el siguiente:

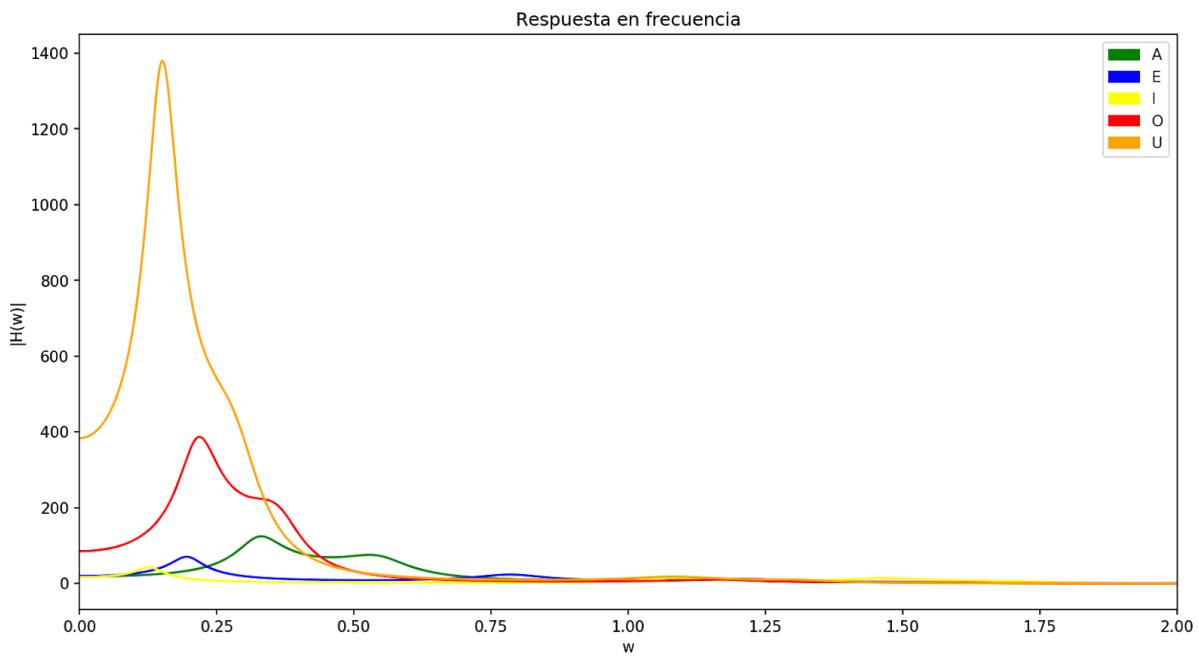


Figura 14. Respuesta en frecuencia (tracto vocal) de las vocales

En cuanto a los polos y cero de cada vocal se parte de la fórmula del tracto vocal.

$$H_n(z) = \frac{1}{(1-p_n z^{-1})(1-p_n^* z^{-1})}$$

Dado que no tiene ceros (el numerador es 1). Los polos para cada F_n , B_n son p_n y p_n^* (conjugado). Los diagramas de polos obtenidos son:

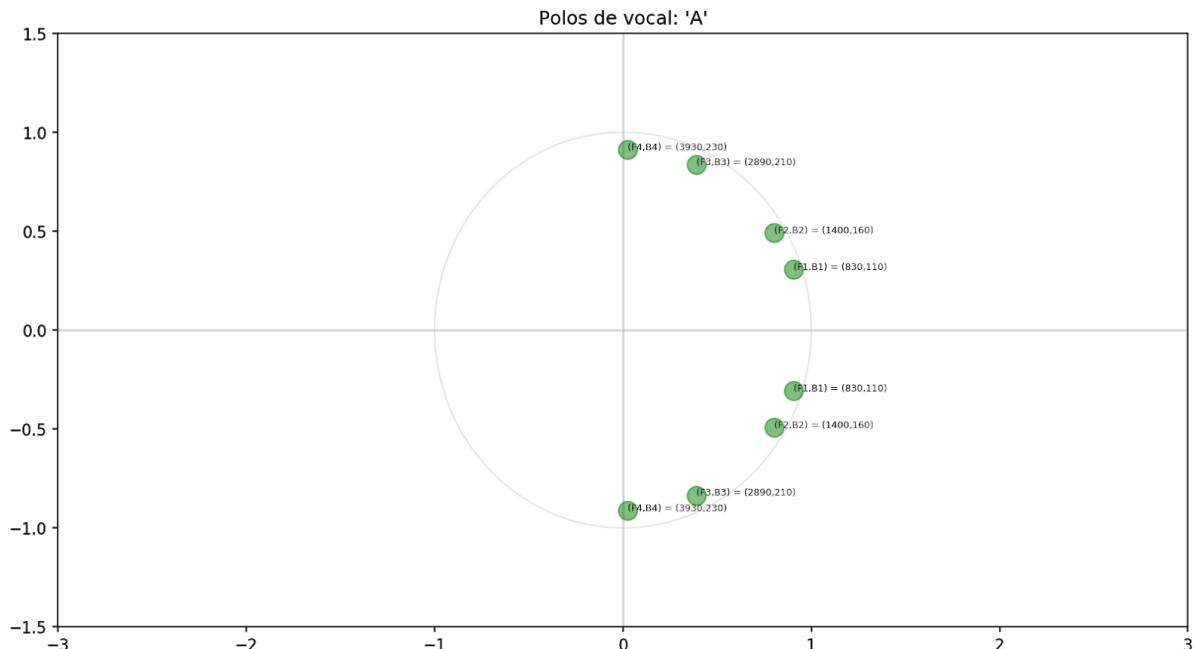


Figura 15.1. Polos correspondiente a la vocal A

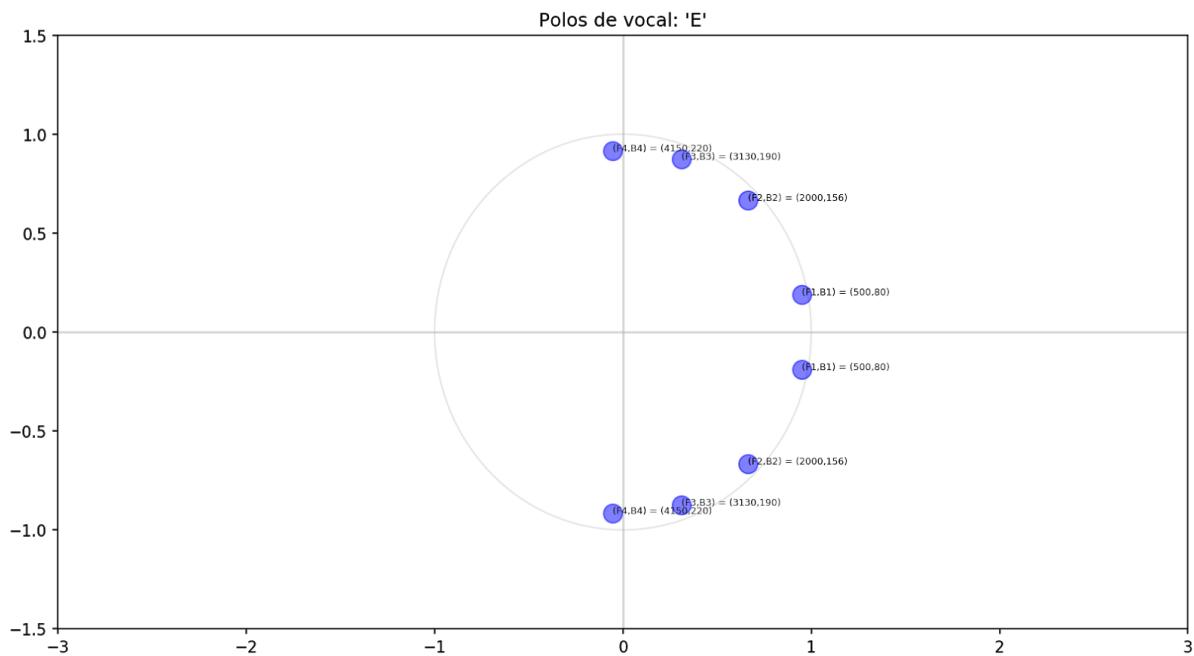


Figura 15.2 Polos correspondiente a la vocal E

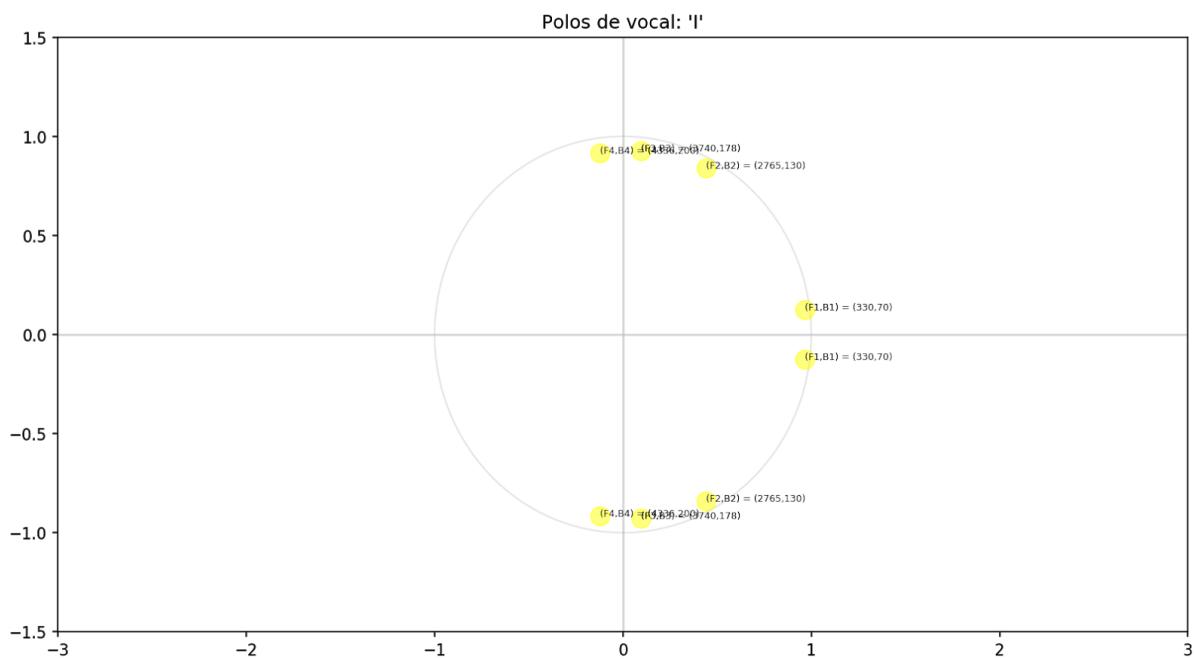


Figura 15.3 Polos correspondiente a la vocal I

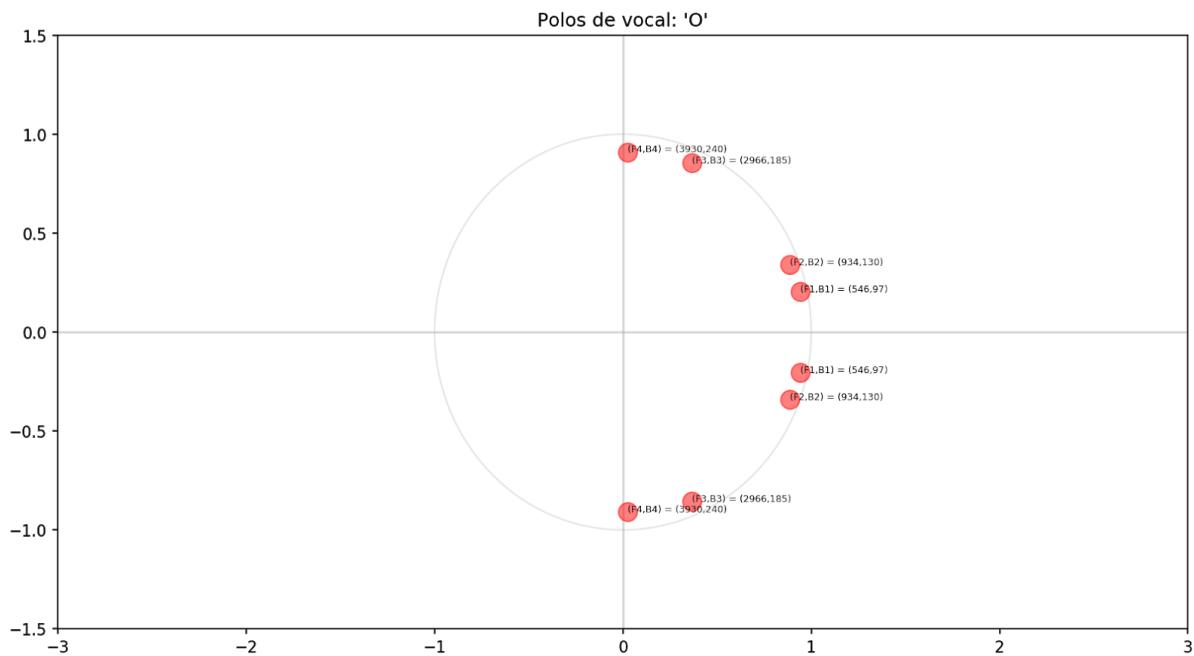


Figura 15.4 Polos correspondiente a la vocal O

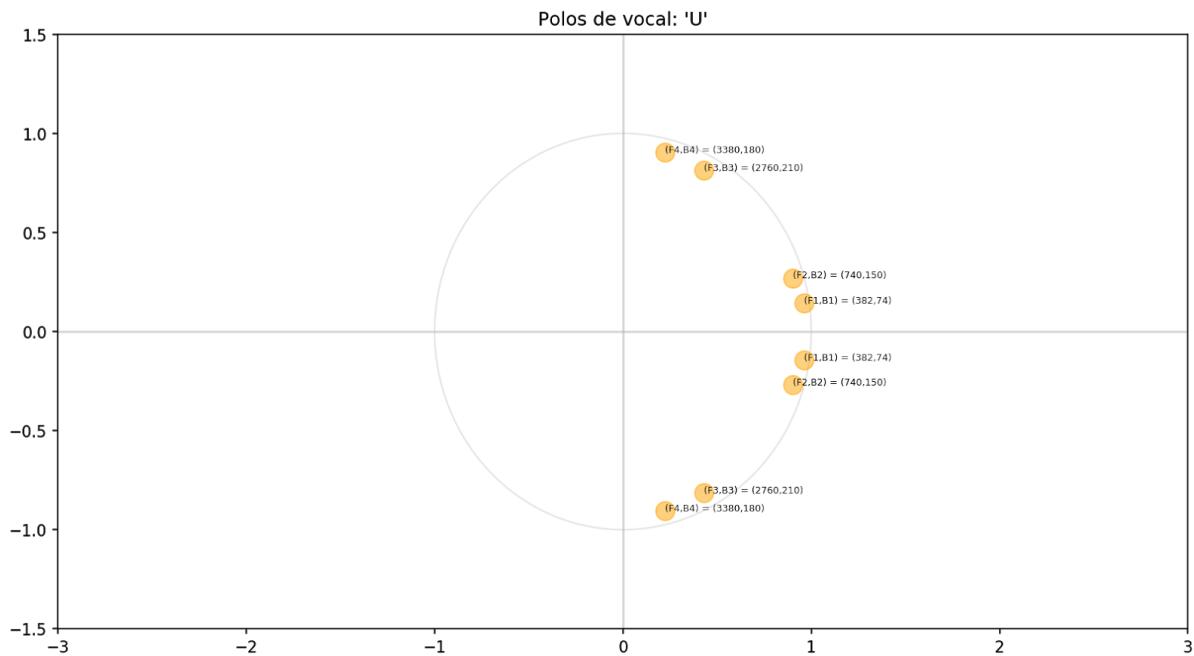


Figura 15.5 Polos correspondiente a la vocal U

8 - Utilizando los resultados de los dos último ejercicios, sintetice un segundo de las cinco vocales. Escuche y grafique. Haga un análisis en frecuencia, y en tiempo-frecuencia.

Vocal "A"

A partir de los polos obtenidos en el ejercicio anterior, se realiza un filtro donde se le pasa el pulso glótico. El resultado de este filtro y esta excitación es una vocal sintetizada. Gráficamente se obtiene:

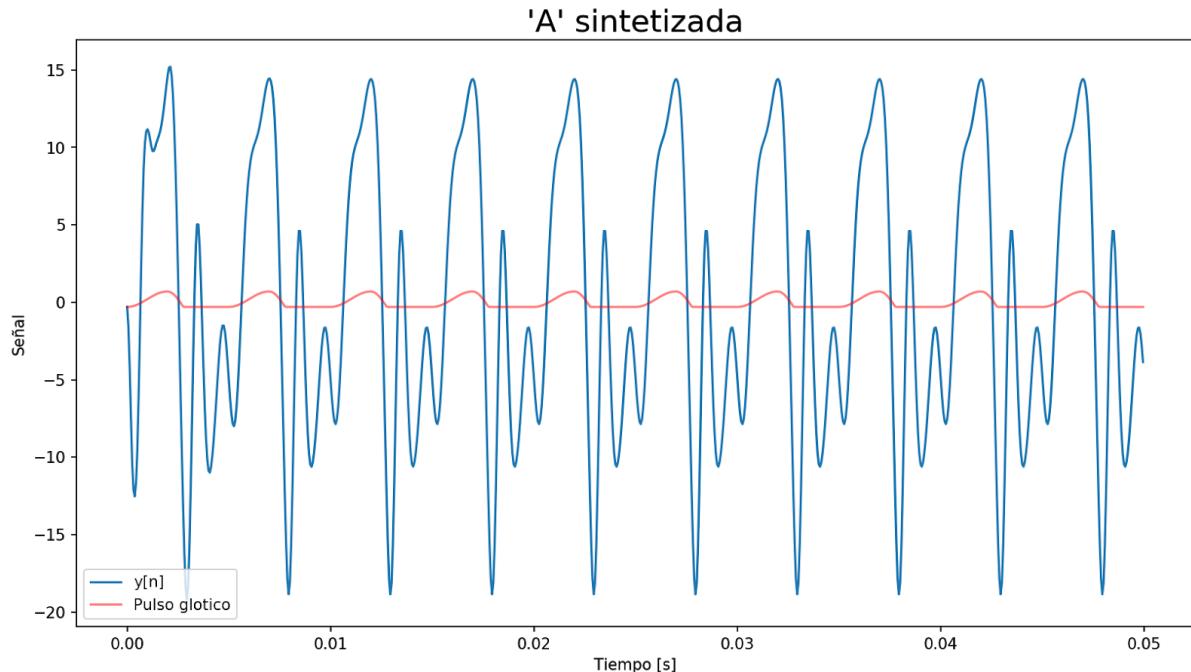


Figura 16. Señal resultante. A sintetizada y de fondo el pulso glótico

Puede notarse una similitud gráfica de la señal sintetizada con la Figura 3. Se observa que al realizar una comparación con la DFT de dicha señal los picos sean aproximadamente iguales en comparación a los de la vocal sintetizada. Puede decirse que estos picos describen cada vocal.

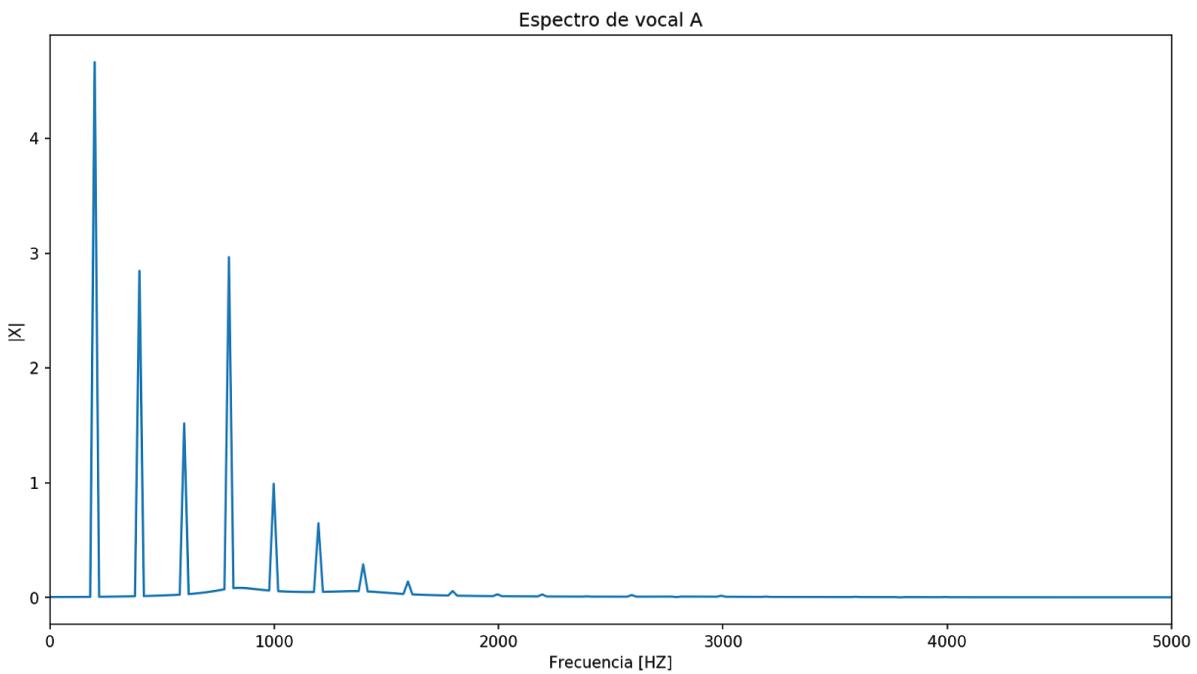


Figura 17.1 Espectro de la vocal A sintetizada.

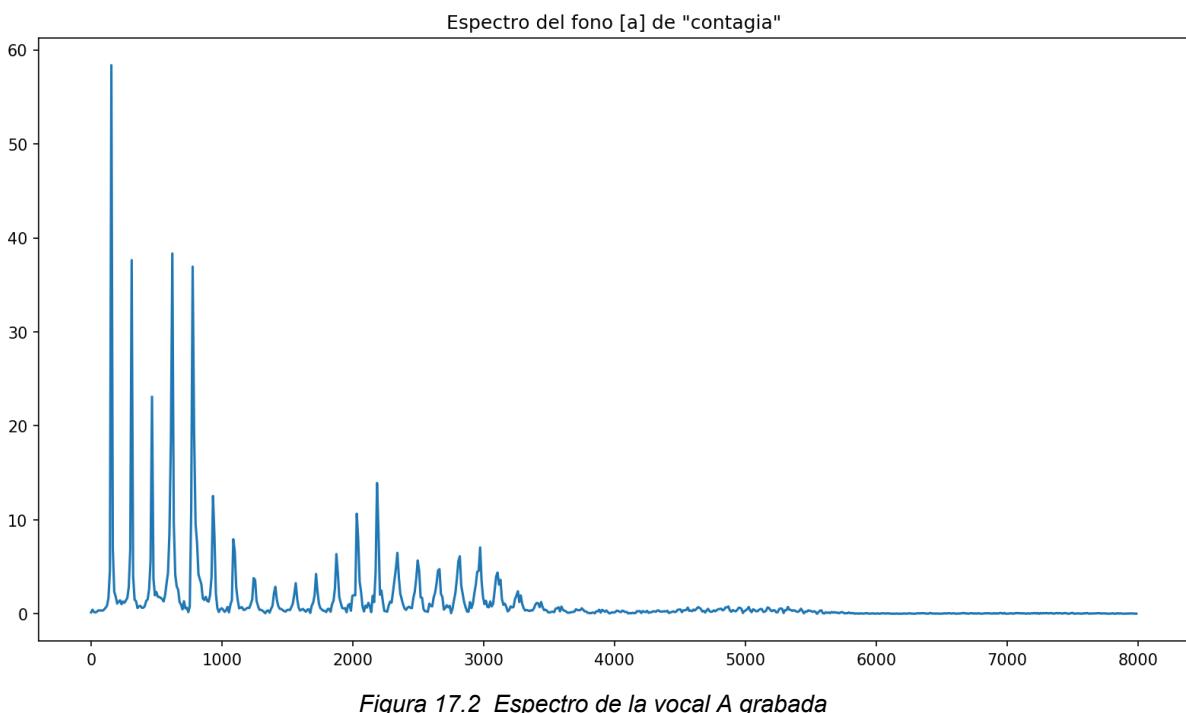


Figura 17.2 Espectro de la vocal A grabada

La salida wav para cada vocal se puede encontrar dentro de la carpeta “out”. Al oírse puede notarse una voz robótica pronunciando la vocal. Para las restantes vocales se hizo el mismo análisis, obteniendo el mismo resultado, el audio “robótico”.

Vocal “E”

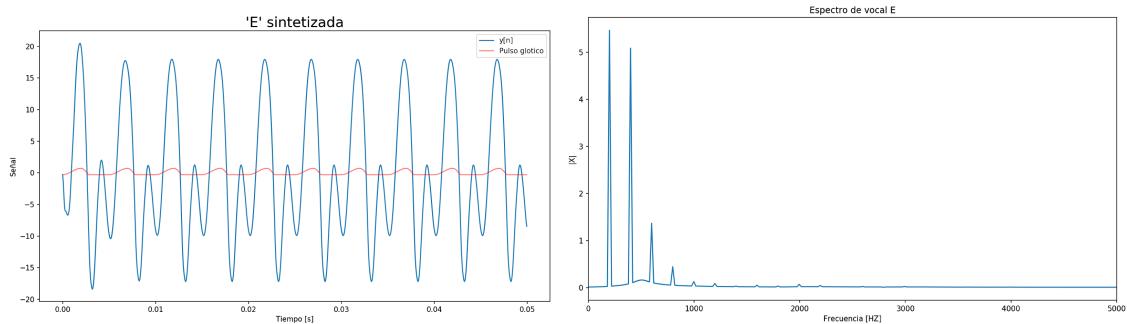


Figura 18.1. Señal de la vocal E sintetizada (izq). Espectro de la señal de la vocal E (der)

Vocal “I”

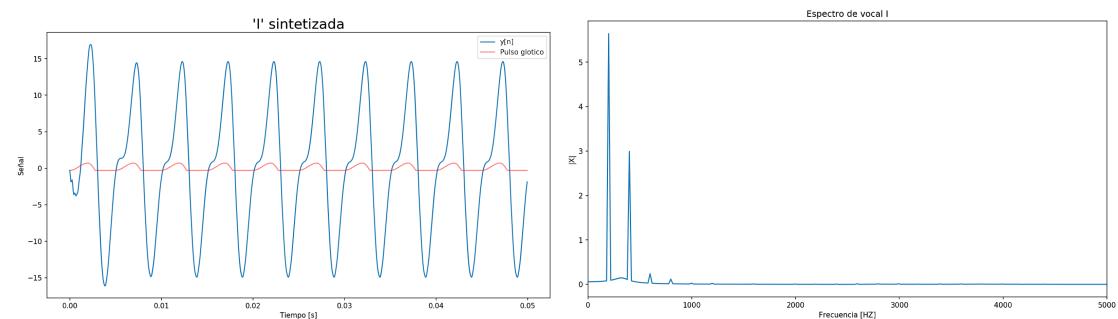


Figura 18.2 Señal de la vocal I sintetizada (izq). Espectro de la señal de la vocal I (der)

Vocal “O”

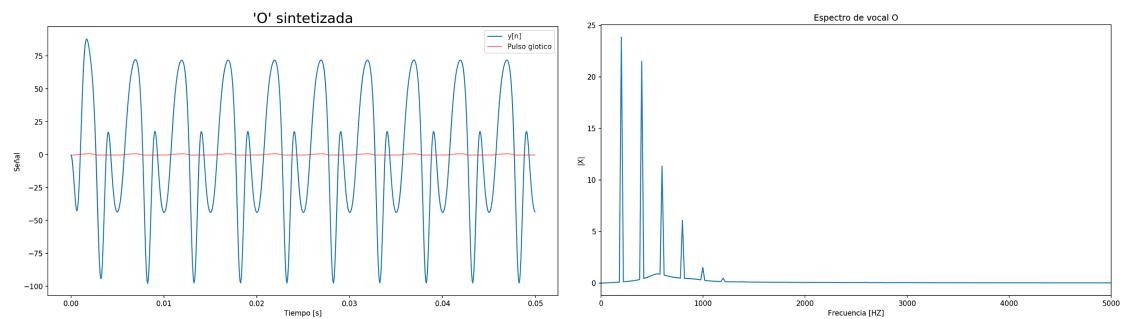


Figura 18.3 Señal de la vocal O sintetizada (izq). Espectro de la señal de la vocal O (der)

Vocal “U”

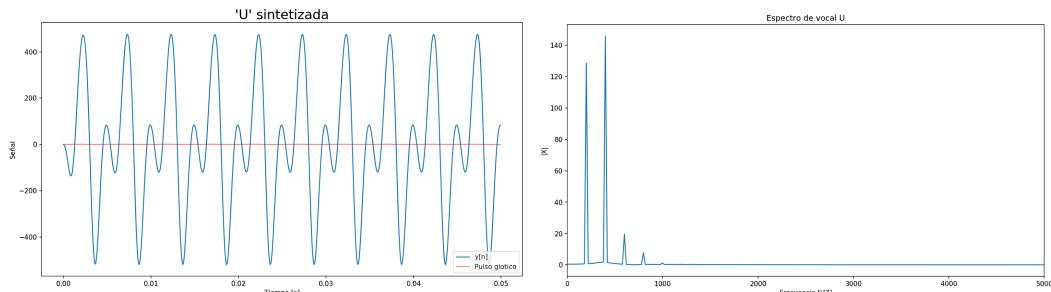


Figura 18.4 Señal de la vocal U sintetizada (izq). Espectro de la señal de la vocal U (der)

9 - A partir de las vocales sintetizadas del ejercicio anterior, estime la respuesta en frecuencia correspondiente al tracto vocal y el contorno de la frecuencia fundamental mediante la transformada cepstrum.

Estimación de la respuesta en frecuencia

Teniendo la señal sintetizada de una vocal y la función Cepstrum de la misma. Realizó los siguientes cálculos para estimar la respuesta en frecuencia.

$$1) y^*[n] = x^*[n] + h^*[n]$$

$y^*[n]$ es Cepstrum de señal de una vocal sintetizada

Eliminando frecuencias bajas descarto $x^*[n]$. Como Quefrecia es inverso a frecuencia. Se tiene que eliminar las Quefrecias bajas. El umbral se define manualmente. Ejemplo de cómo se eliminan las Quefrecias Bajas para la vocal “A” utilizando un umbral de 2 milisegundos.

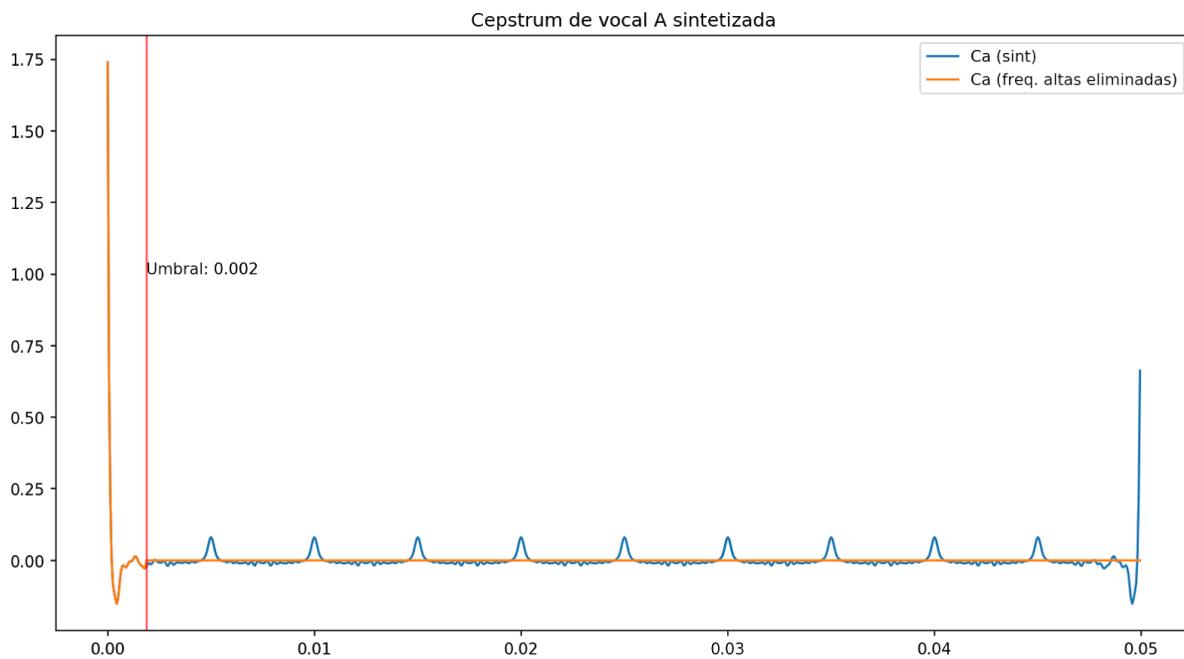


Figura 19. Cepstrum de la vocal A sintetizada y eliminación de Quefrecias altas

$$2) y[n] = h[n]$$

$$F\{h[n]\} = F^{-1}\{\log|F\{h[n]\}|\}$$

$$F\{h[n]\} = \log|F\{h[n]\}|$$

: Donde $F\{h[n]\} = H(w)$ <--- Respuesta en frecuencia

$$F\{h[n]\} = \log|H(w)|$$

$$\exp^{\{h[n]\}} = H(w) \text{ aproximada.}$$

Comparó con la H de la figura 14. Con módulo y logaritmo aplicado. El resultado obtenido para la vocal A es el siguiente:

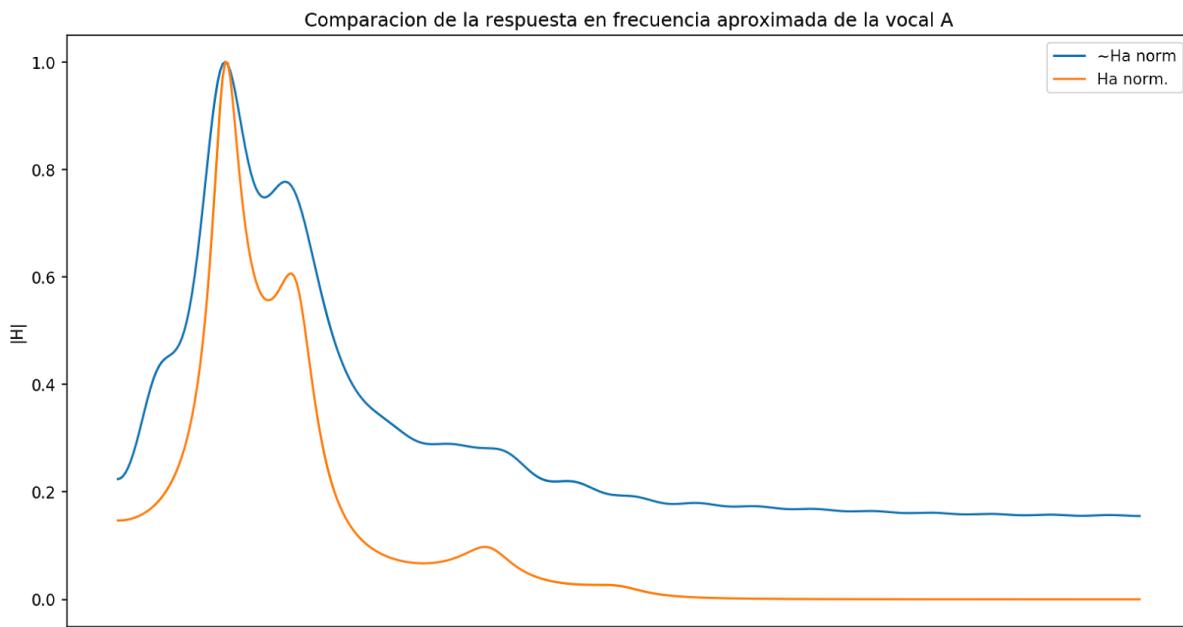


Figura 20. Aproximación de la respuesta en frecuencia de la vocal A a partir de la transformada Cepstrum

Puede observarse una pequeña similitud en las curvas y sus picos.

El resultado para las restantes vocales utilizando el mismo umbral son los siguientes:

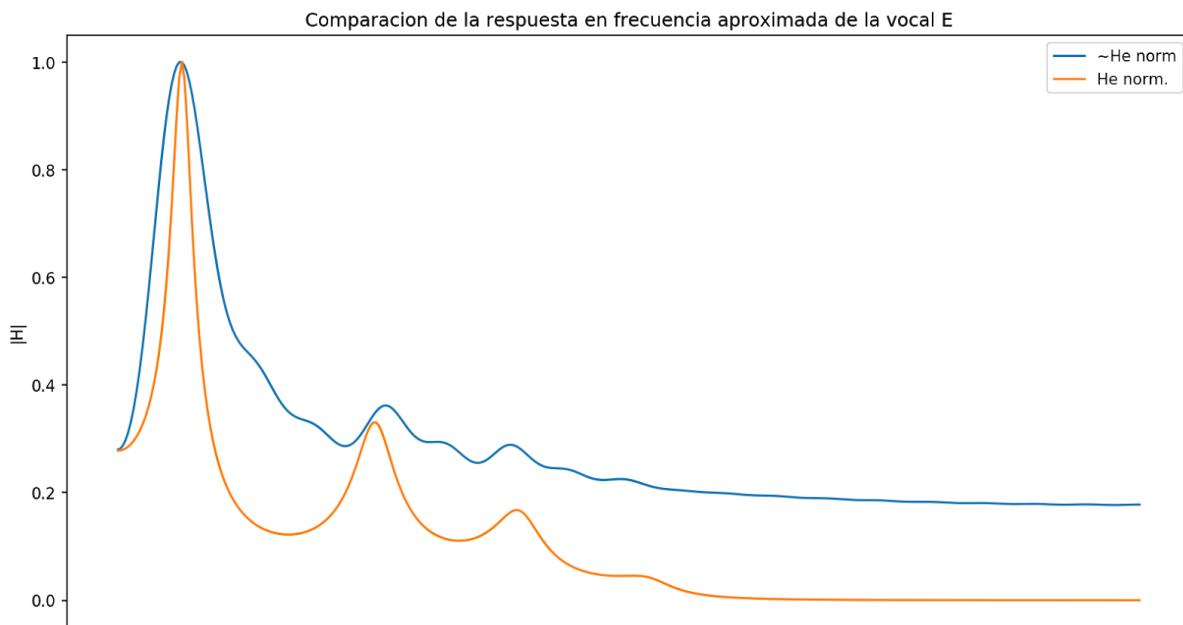


Figura 21. Aproximación de la respuesta en frecuencia de la vocal E a partir de la transformada Cepstrum

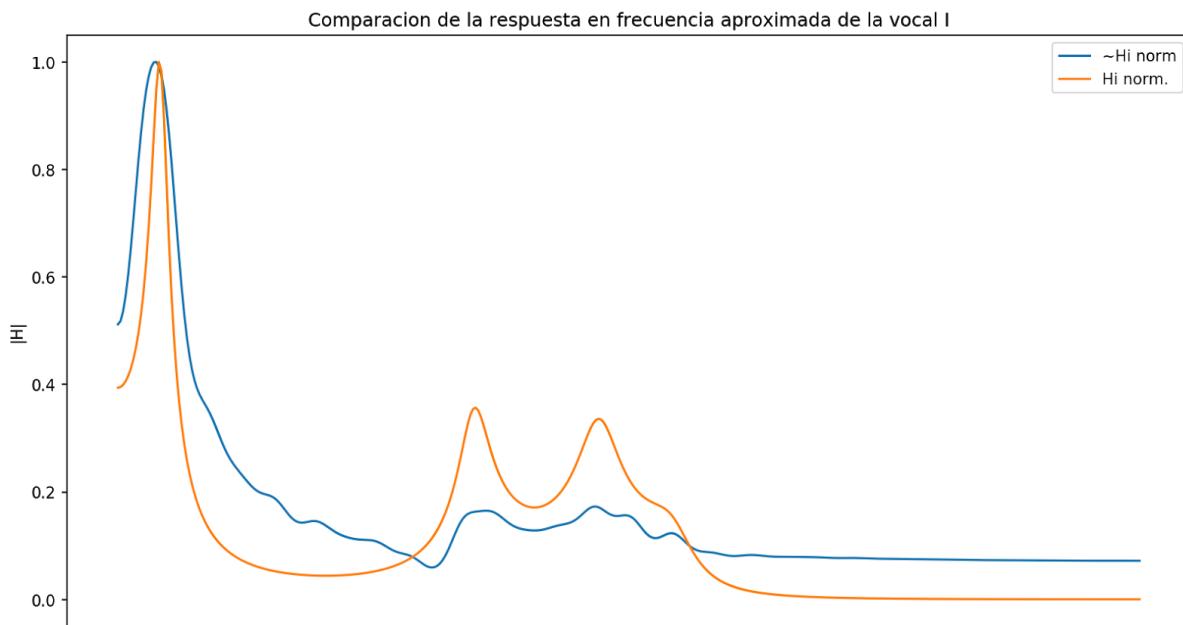


Figura 22. Aproximación de la respuesta en frecuencia de la vocal I a partir de la transformada Cepstrum

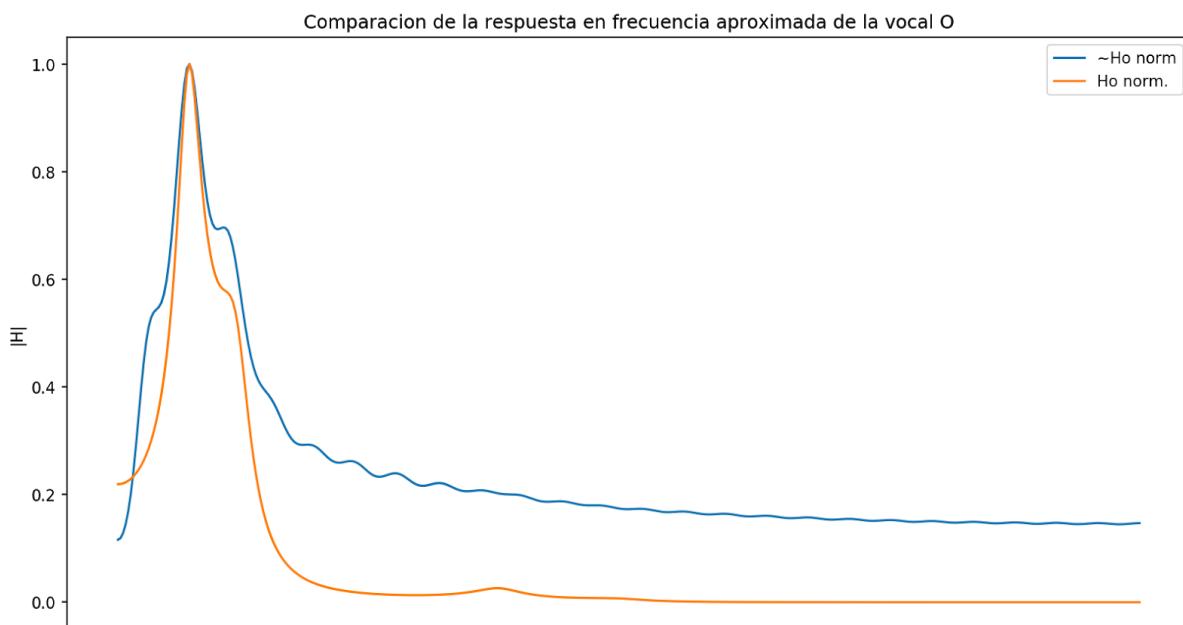


Figura 23 Aproximación de la respuesta en frecuencia de la vocal O a partir de la transformada Cepstrum

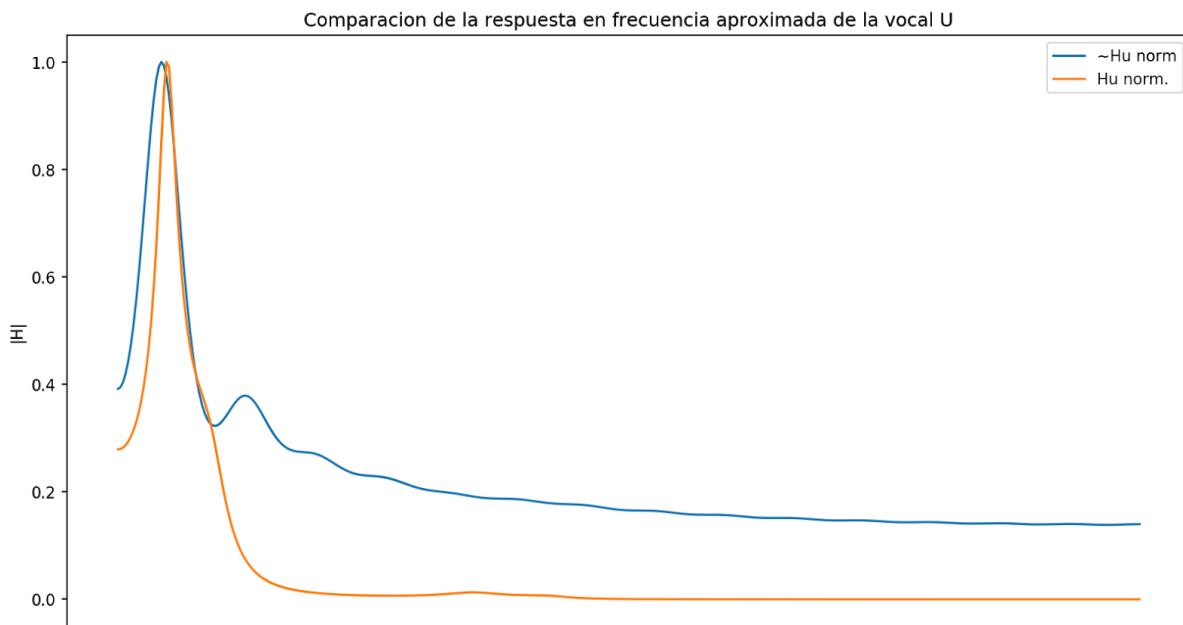


Figura 24. Aproximación de la respuesta en frecuencia de la vocal U a partir de la transformada Cepstrum

Frecuencia fundamental

Para el cálculo de la frecuencia fundamental de cada una de las vocales. Se toma la vocal, se le aplica la transformada Cepstrum y dentro del rango de los 0.002 segundos a los 0.02 segundos se busca la posición (quefrecuencia) del pico máximo. A partir de la quefrecuencia obtenida se calcula la frecuencia invirtiendo. El análisis para el fono [a] es el siguiente:

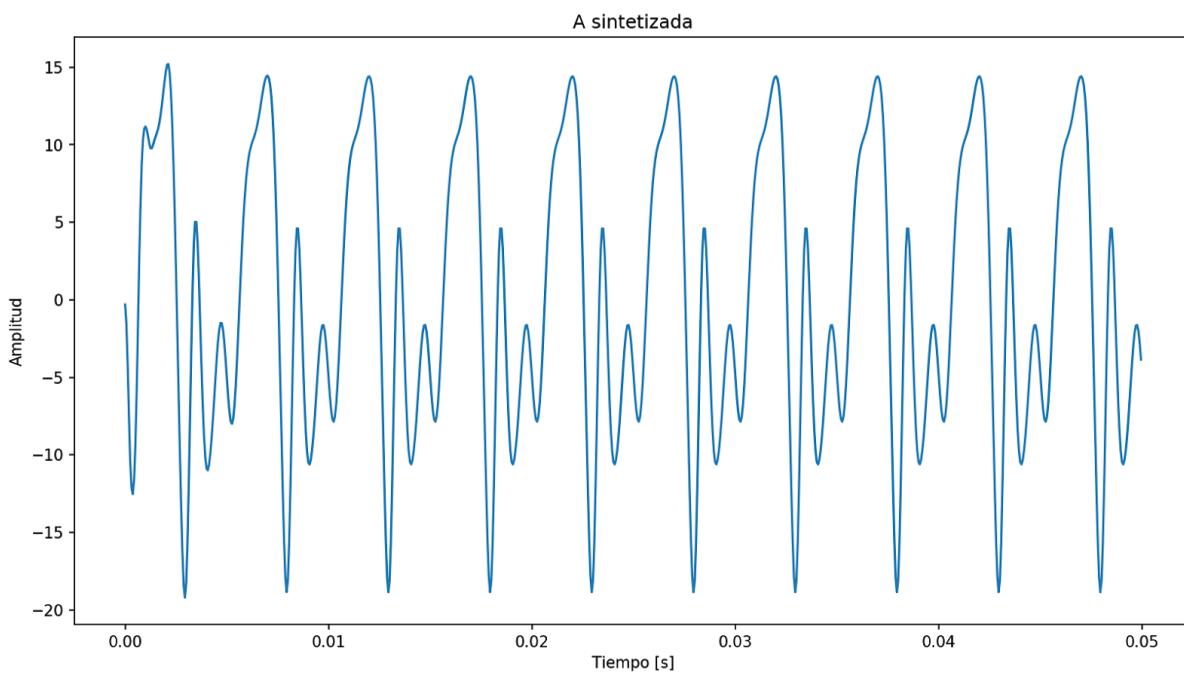


Figura 25. Señal del fono [a] sintetizado

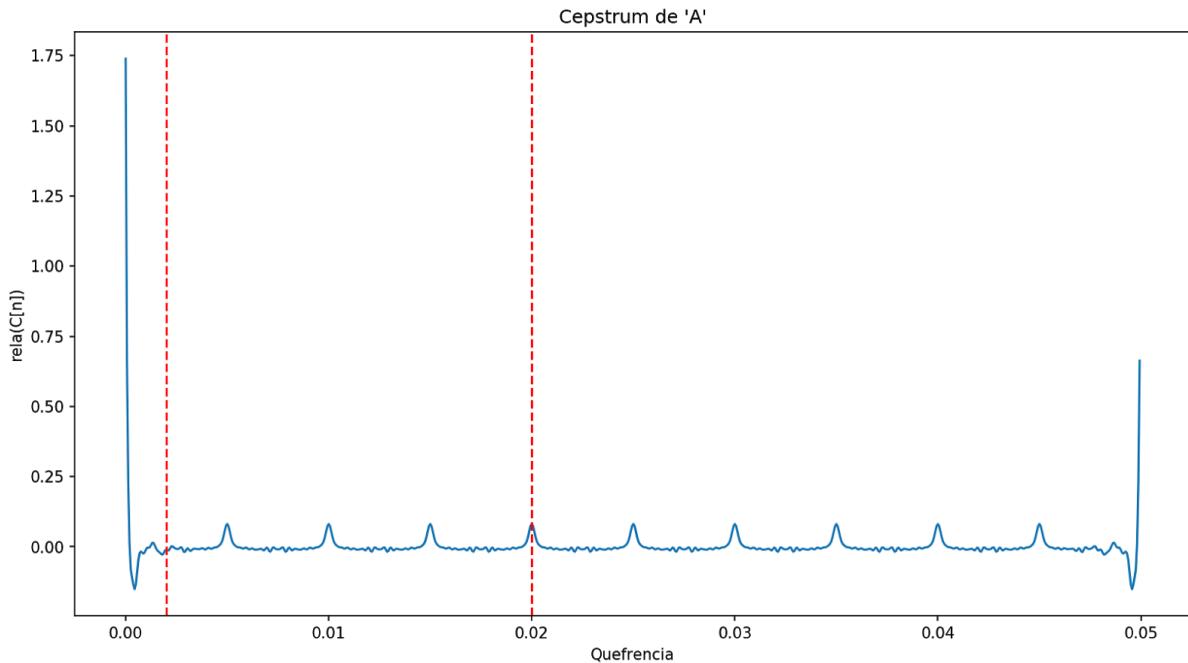


Figura 26. Cepstrum de la señal del fono [a] sintetizada

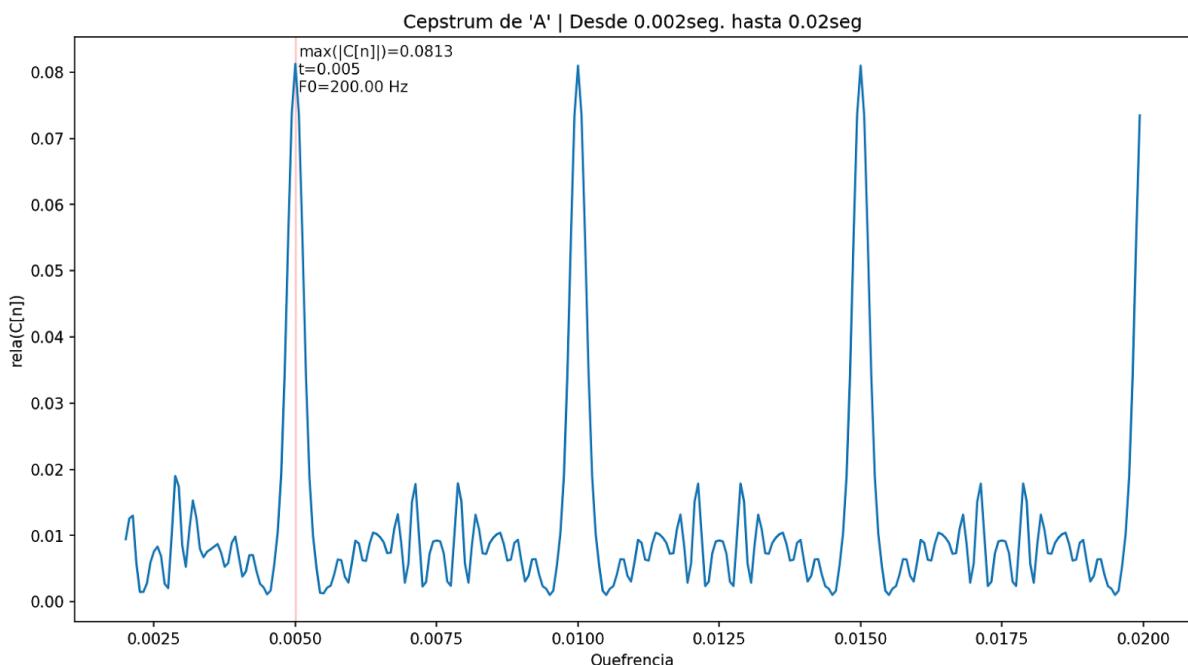


Figura 27. Zoom de Cepstrum en el rango 0.002 a 0.02 segundos. Se aprecia el pico donde se obtiene F0

Donde se encuentra que la frecuencia fundamental F0 es de **200 Hz**. Ya que el primer pico se con amplitud 0.0813 se encuentra en la Quefrecuencia 0.005.

La frecuencia fundamental (F0) calculada, realizando el análisis análogo a la vocal A, para las demás vocales son las siguientes:

- Vocal E

$$\max(|C[n]|) = 0.0547$$

$$q=0.05 \text{ [s]} \\ F0=200.0 \text{ Hz}$$

- Vocal I

$$\max(|C[n]|)=0.0740 \\ q=0.005 \text{ [s]} \\ F0=200.0 \text{ Hz}$$

- Vocal O

$$\max(|C[n]|)=0.0435 \\ q=0.05 \\ F0=200.0 \text{ Hz}$$

- Vocal U

$$\max(|C[n]|)=0.0534 \\ q=0.005 \\ F0=200.00 \text{ Hz}$$

10 - Utilizando nuevamente la transformada cepstrum, estime el contorno de la frecuencia fundamental de la voz en el archivo hh15.wav. Grafique en forma sincrónica con la onda.

A partir de una señal y su tiempo (t, X) se itera en ventanas de tamaño parametrizable. En cada una de las ventanas se hace el análisis que se hizo en el ejercicio 9. Es decir, se realiza la transformada cepstrum, se toma el intervalo entre $1/500 - 1/50$ y luego tomó la quefrencia del máximo valor en el rango, invierto la misma y obtener la frecuencia fundamental para esa ventana.

El resultado obtenido después de procesar la señal de audio completa es el siguiente:

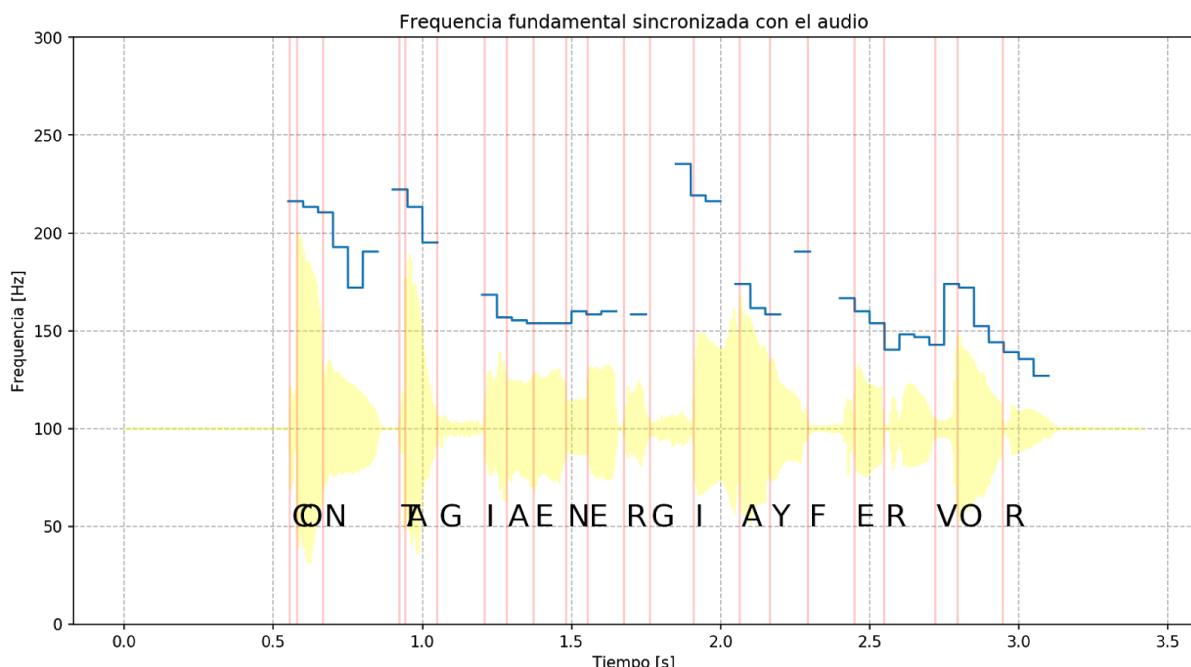


Figura 28. Contorno de frecuencia en sincronía con la señal del audio "hh15.WAV"

11 - Aplique el método PSOLA para aumentar y disminuir un 10%, 20% y 30% la frecuencia fundamental de la voz en el archivo hh15.wav. Para ello utilice la curva de frecuencia fundamental obtenida en el ejercicio anterior

Ejercicio realizado en código.

12 - Aplique el método PSOLA para aumentar y disminuir un 10%, 20% y 30% la duración de la voz en el archivo hh15.wav. Para ello utilice la curva de frecuencia fundamental obtenida en el ejercicio anterior.

Ejercicio realizado en código.

13 - Modifique la frecuencia fundamental de las vocales sintetizadas del ejercicio 8 desde 200 a 300 Hz en forma lineal. Escuche la onda resultante, ¿cómo se percibe el cambio en la frecuencia fundamental? Estime el F0 resultante y compárelo con el teórico.

Ejercicio realizado en código.

14 - Repita el ejercicio anterior pero esta vez variando la frecuencia fundamental desde 200 a 100 Hz.

Ejercicio realizado en el código.

15 - Aplique un filtro a las vocales sintetizadas del ejercicio 8 para eliminar la frecuencia fundamental. Puede utilizar la herramienta fdatool para diseñar el filtro. Justifique el filtro implementado. Grafique ambas señales, haga un análisis en frecuencia y compare. ¿Perceptualmente se nota alguna diferencia? ¿Porqué?

Para eliminar la frecuencia fundamental de las vocales sintetizadas a partir de un filtro debemos analizar los conjuntos de frecuencias fundamentales calculados en el ejercicio 9.

VOCAL	F0 [Hz]
A	200
E	200
I	200

O	200
U	200

Se puede observar que las frecuencias fundamentales están por debajo de los 200 Hz. Para ello se propone diseñar un filtro pasa altos en la que ignore frecuencias por debajo de los 200Hz.

Las características del filtro son:

- **Fs:** 16000
- **Fh:** 300 (Frecuencia de corte)
- **Bh:** 200 (Ancho de banda de transición)
- **Tipo Ventana:** Hamming

El filtro en cuestión es el siguiente:

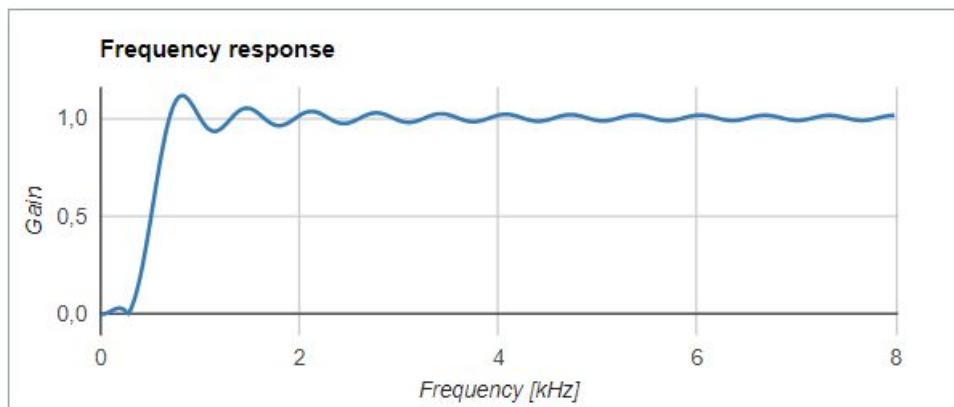


Figura 29. Respuesta en frecuencia del filtro propuesto

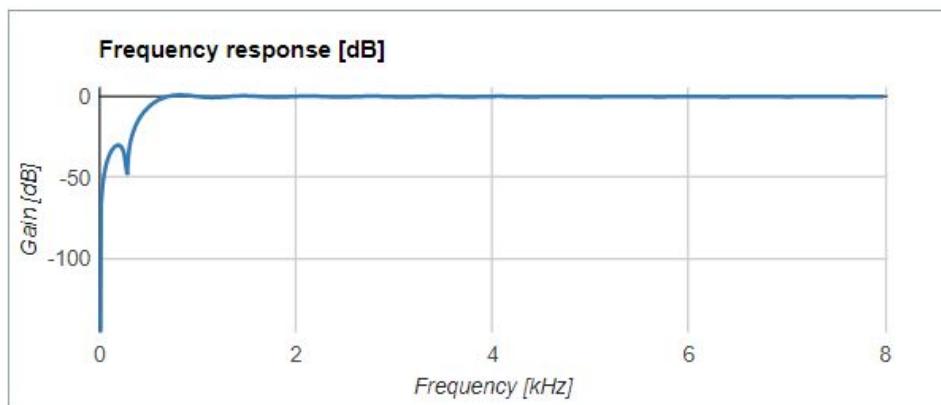


Figura 30. Respuesta en frecuencia del filtro propuesto

El señal resultante al pasar la señal correspondiente a el fono [a] sintetizado por el filtro (convolución) es la siguiente:

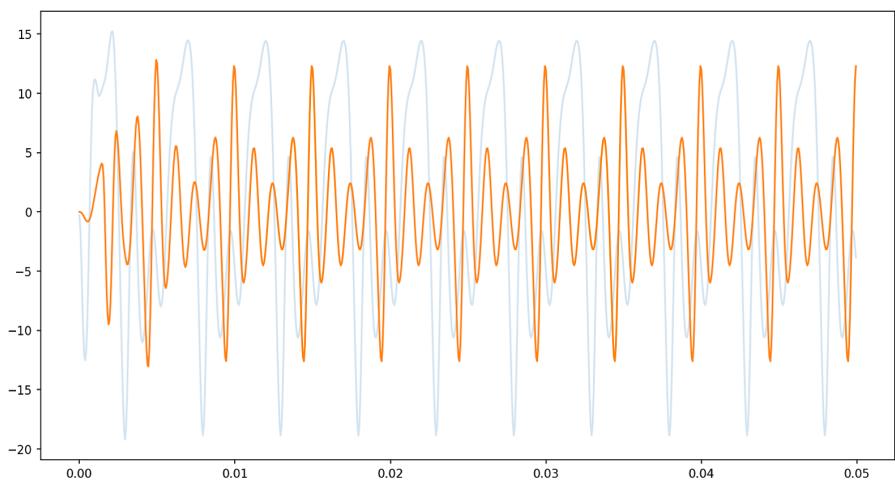


Figura 31. Señal resultante

Y su espectro de frecuencias es el siguiente. Puede notarse que las frecuencias menores a los 200 Hz están suprimidas. Mientras que de un color más claro esta el valor previo a realizar la convolución con el filtro.

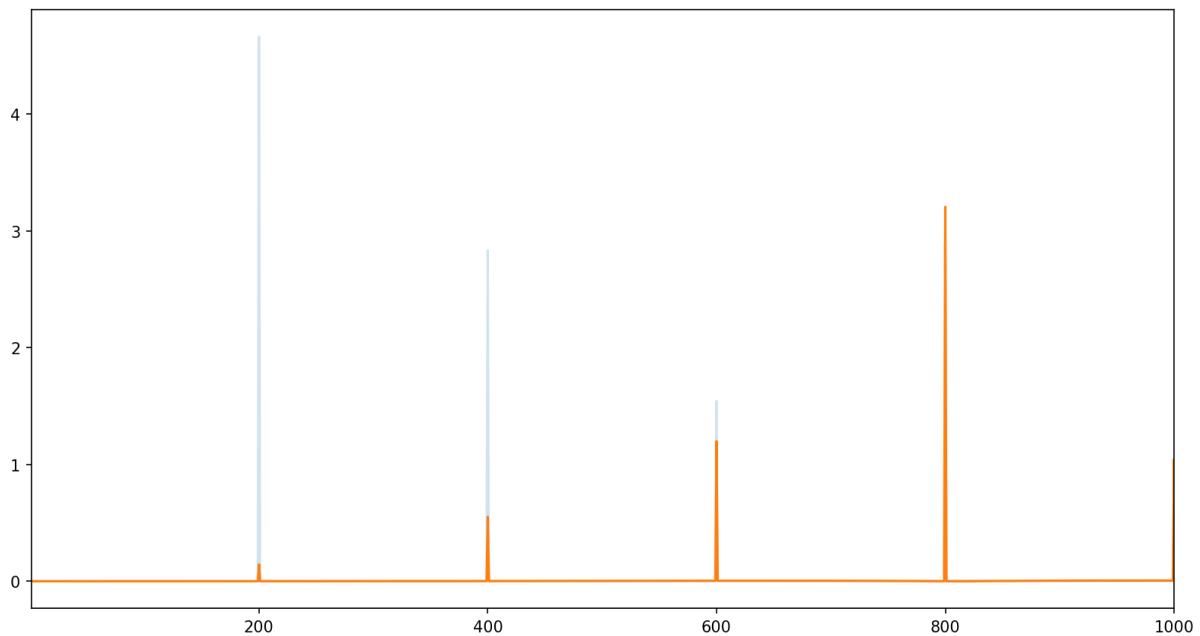


Figura 31. Espectro de frecuencia de Señal resultante