

Facultad de Ingeniería – Universidad de Buenos Aires
Señales y Sistemas TB065
Primer Cuatrimestre del 2025

Curso 1: Rey Vega - Caldentey - Marques Rojo - Florentín

Trabajo Práctico Especial - Parte 2

Integrantes

María de las Mercedes Gamberale	108834
Carolina Gonzales Peralta	110804
Ezequiel Sanchez Aliva	103799

Índice

Introducción.....	3
Realización del Espectrograma de Banda Angosta de la señal completa.....	3
Realización del Espectrograma de Banda Ancha de la señal completa.....	5
Realización de los Espectrogramas de Banda Angosta y Ancha de las distintas vocales.....	7
Letra ‘i’:.....	7
Letra ‘a’:.....	8
Letra ‘o’:.....	10
Modificación de la frecuencia fundamental de la señal.....	12
Conclusión.....	15

Introducción

En esta segunda parte del Trabajo Práctico Especial (TPE), se trabajará con la generación de espectrogramas de la señal para analizar la muestra en el dominio frecuencial. Primero se trabajará con la señal completa, de la cual se realizará un espectrograma de banda angosta, para observar la frecuencia fundamental de la señal y los armónicos de la misma, y de banda ancha, para observar los formantes de los sonidos vocálicos. Luego se trabajará con las vocales de la muestra, analizándolas mediante espectrogramas para comparar sus frecuencias fundamentales, armónicos y formantes. Finalmente se modificará la frecuencia fundamental de la señal para que ésta suene más aguda por medio del método TD - PSOLA.

Para la realización de esta parte del TPE se utilizó la muestra grabada en la primera parte a una velocidad más lenta, esperando que la mayor extensión de la señal permitiese realizar un análisis más detallado de la misma.

Realización del Espectrograma de Banda Angosta de la señal completa

Un espectrograma de banda angosta es una representación de la variación del contenido de frecuencia de una señal a lo largo del tiempo. Particularmente, se divide la señal en ventanas largas para tener una buena resolución en frecuencia (lo que, por consiguiente, implica una peor resolución temporal). Se utilizó un solapamiento de ventanas del 50% con el objetivo de reducir los efectos de borde y obtener una gráfica más continua en frecuencia.

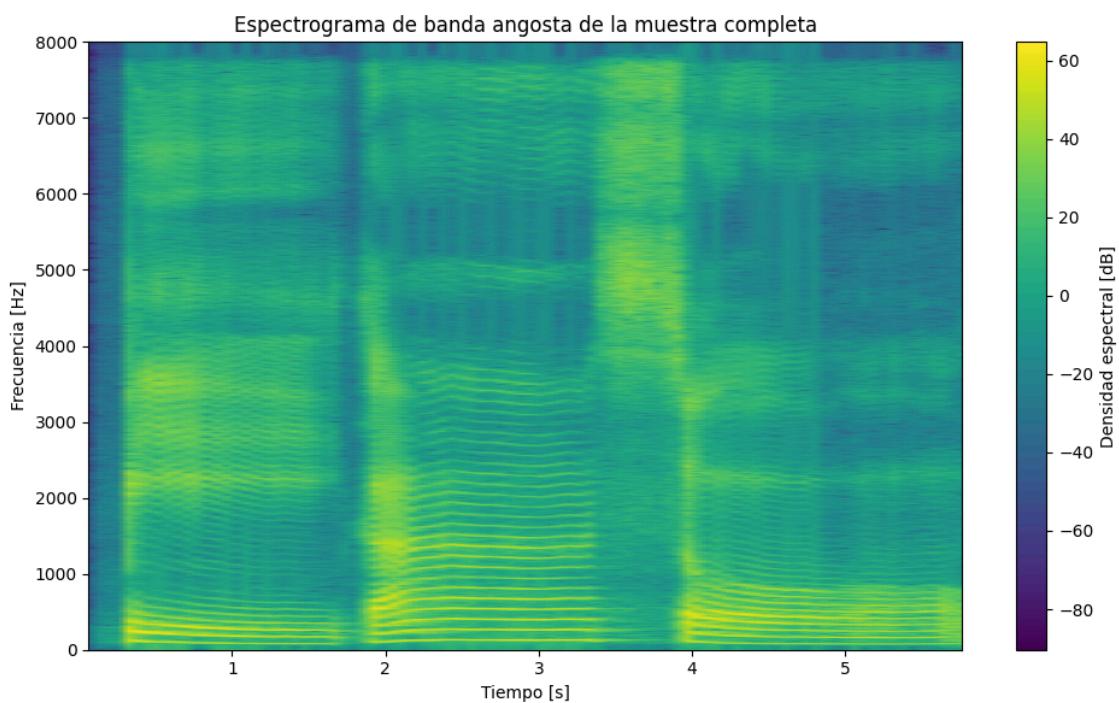


Figura 1: Espectrograma de banda angosta utilizando la señal completa.

En la Figura 1 se pueden observar los distintos períodos pseudo-estables de la señal en porciones de la misma, donde la frecuencia parece mantenerse o reducirse ligeramente en intervalos de tiempo considerables, destacándose los intervalos (1s; 1,5s), (2,5s, 3,1s) y (4,2s, 5s). Estos intervalos son parte, respectivamente, de cada una de las vocales presentes en la muestra; y se pueden discernir sus distintas frecuencias fundamentales además de verse sus armónicos claramente. En adición a ésto, se puede apreciar la presencia de la frecuencia fundamental por debajo de los 1000Hz, y de los armónicos de la señal equidistantes entre sí, pero no se puede apreciar con claridad cuál es la frecuencia fundamental de la señal con la escala provista. Por ende, se realizó una ampliación de la señal para ver únicamente el espectrograma entre 0 Hz y 2 kHz, presente en la Figura 2.

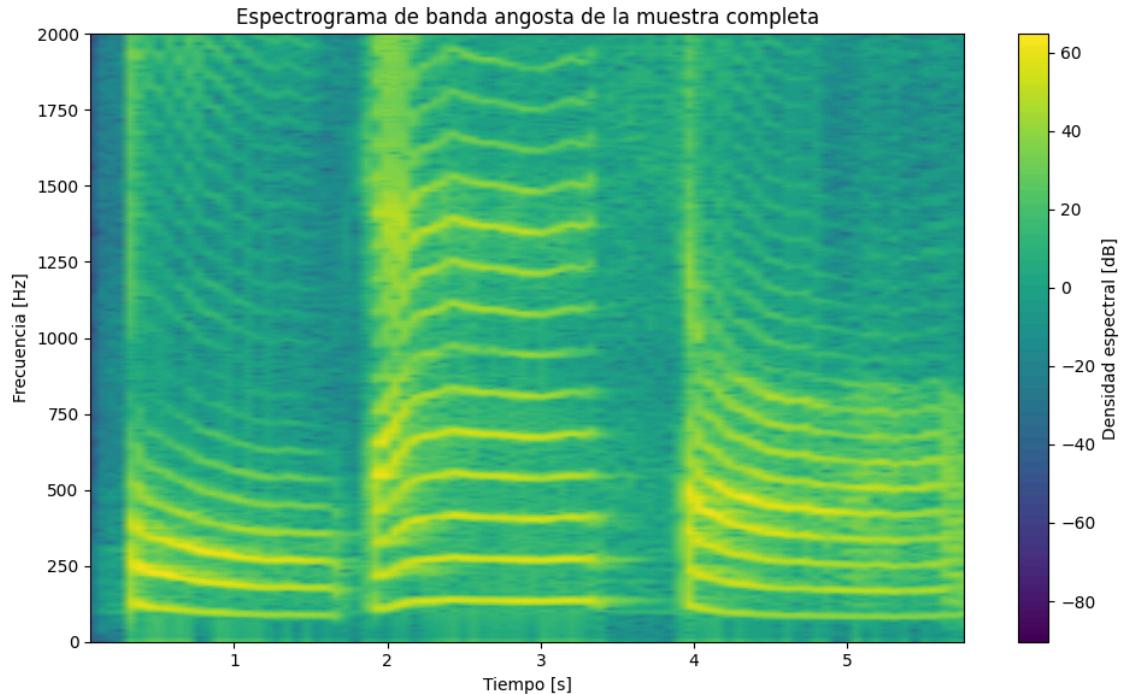


Figura 2: Espectrograma de banda angosta utilizando la señal completa, viéndose frecuencias de hasta 2 kHz.

En la Figura 2 se observa claramente que, si bien hay una ligera variación dependiendo de la sección de la muestra considerada, para las partes armónicas la frecuencia fundamental se encuentra entre los 114 Hz y los 140 Hz. Se observan también sus armónicos equidistantes al aumentar la frecuencia (el primero, por ejemplo, se encuentra cerca de los 250 Hz).

Para verificar ésto por software, se buscó la frecuencia fundamental de cada muestra de la señal, obteniéndose lo marcado en la línea roja en la Figura 3, en la cual se puede apreciar esta ligera variación en las secciones armónicas (y la falta de una clara frecuencia fundamental en los sonidos percusivos presentes en ciertos sectores de la señal).

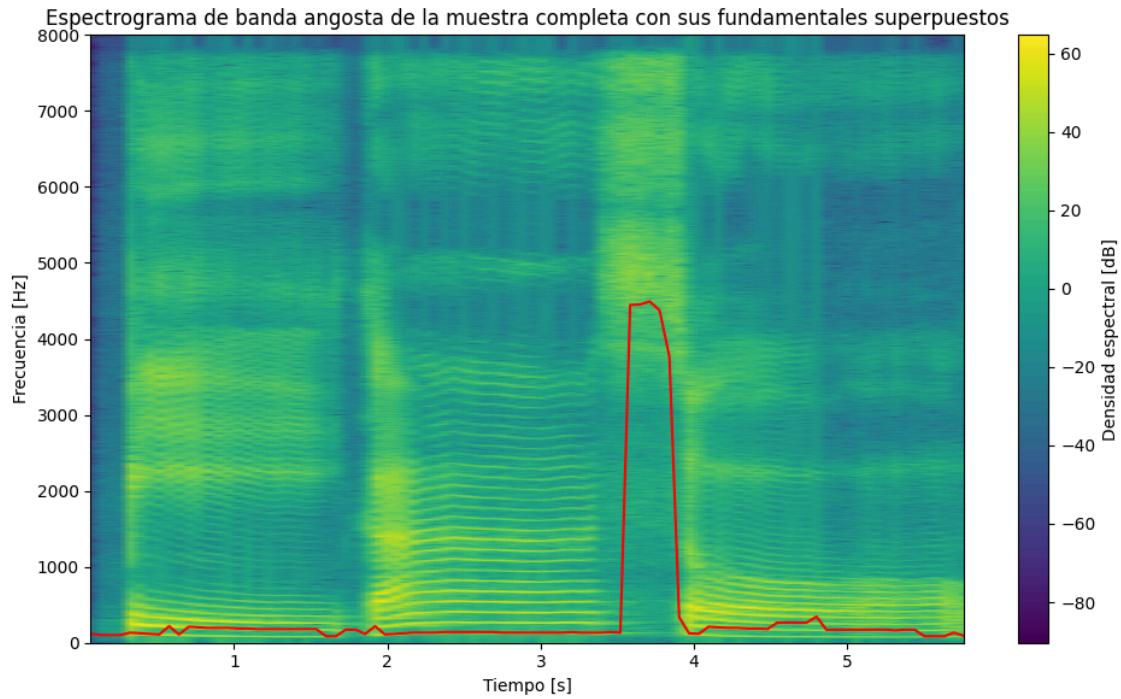


Figura 3: Espectrograma de banda angosta utilizando la señal completa, con sus frecuencias fundamentales superpuestas (indicadas en rojo).

Realización del Espectrograma de Banda Ancha de la señal completa

Un espectrograma de banda ancha es una representación de la variación del contenido de frecuencia de una señal a lo largo del tiempo. Particularmente, se divide la señal en ventanas cortas para tener una buena resolución temporal (lo que, por consiguiente, implica una peor resolución en frecuencia). Es útil para ver cuándo ocurren los cambios bruscos en la señal, pero no permite distinguir las frecuencias cercanas de manera precisa. Para mantener una buena resolución temporal sin perder tantos detalles y poder detectar cambios en el tiempo, se usó un 60% de solapamiento.

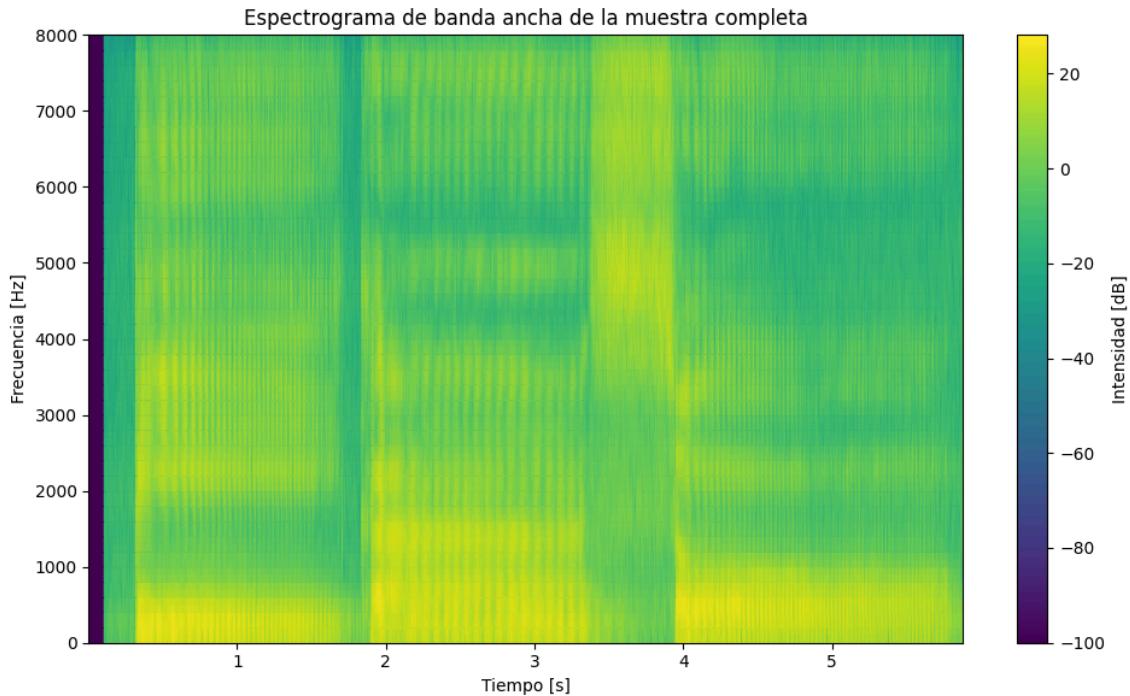


Figura 4: Espectrograma de banda ancha utilizando la señal completa.

En la Figura 4 se puede observar con buena resolución temporal los formantes de los sonidos vocálicos en los intervalos de tiempo presentados en la Tabla 1.

	Tiempo de inicio (start time) [s]	Tiempo de finalización (end time) [s]
“i”	0,32	1,69
“a”	1,89	3,36
“o”	3,95	5,79

Tabla 1: Tiempos de inicio y final de cada vocal tomados del espectrograma de banda ancha de la Figura 4.

Como la ventana es de 5 ms, no se puede discernir el valor de la frecuencia fundamental y sus armónicos (sólo una noción entre qué intervalos puede estar observando su intensidad). Pero se puede interpretar los formantes y la forma en que se dijo la palabra (rasgo vocal).

Para las sílabas ‘pi’ y ‘ca’ se observa un breve silencio (muy poca energía) en el comienzo de la sílaba seguido de una gran intensidad de energía al pasar de la consonante a la vocal, y luego el formante es casi estable. Para la sílaba ‘so’, al pronunciar la ‘s’ se observan curvas en las frecuencias altas mostrando la transición desde la vocal ‘a’ hasta la ‘o’ que mantiene una intensidad casi constante por un tiempo, pero va perdiendo energía (por la naturaleza de terminar de pronunciar la palabra).

Se realiza una ampliación de la Figura 4 para poder observar mejor el primer y segundo formante.

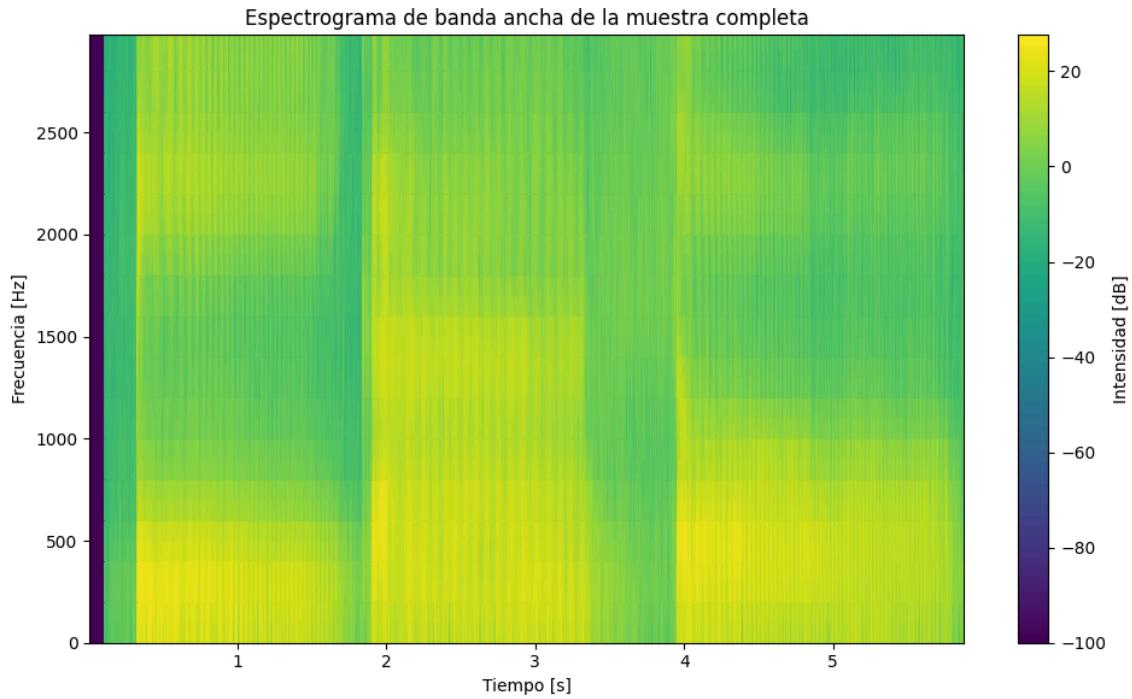


Figura 5: Espectrograma de banda ancha utilizando la señal completa con frecuencias hasta 3 kHz.

Por la posición y la intensidad de cada formante en la Figura 5, para la primera sílaba la lengua está alta y adelantada, con la boca casi cerrada (primer formante bajo y segundo formante alto). Para la segunda sílaba la lengua es baja y central, con la boca abierta (primer formante alto y segundo formante medio). Para la tercera sílaba la lengua está al medio y retraída, con la boca parcialmente abierta (primer formante medio y segundo formante bajo, en el gráfico se muestran en una gran proximidad).

Realización de los Espectrogramas de Banda Angosta y Ancha de las distintas vocales

Letra ‘i’:

A partir del espectrograma de banda angosta del fragmento [0.32, 1.69] que contiene la vocal I (Fig. 6), es posible señalar que la frecuencia fundamental de la letra se encuentra aproximadamente en $f_{0(i)} = 110 \text{ Hz}$. En dicho espectrograma también destacan sus armónicos, ubicados en $f_{1(i)} = 220 \text{ Hz}$, $f_{2(i)} = 330 \text{ Hz}$, $f_{3(i)} = 440 \text{ Hz}$, y en los subsiguientes múltiplos de la frecuencia fundamental.

En un espectrograma de banda ancha, aún habiendo perdido resolución en el dominio de la frecuencia, es posible identificar zonas horizontales con más energía que otras a ciertas frecuencias, estos son los formantes. Estas regiones son difusas como puede verse en la Figura 7, por lo que es complicado señalar una frecuencia particular. Por este motivo, los valores presentados deben tomarse como

estimaciones. En el caso de la I se estima que el primer formante, $F_{1(i)}$, se encuentra alrededor de los 300 Hz. Luego, $F_{2(i)} \approx 2250$ Hz, $F_{3(i)} \approx 2850$ Hz y $F_{4(i)} \approx 3500$ Hz.

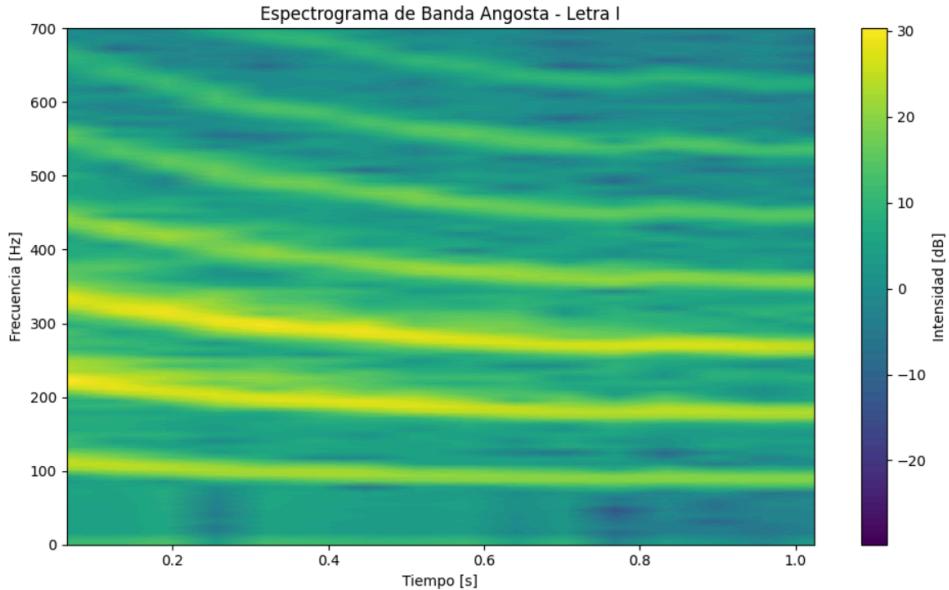


Figura 6: Espectrograma de banda angosta de un fragmento que contiene la vocal 'i.'

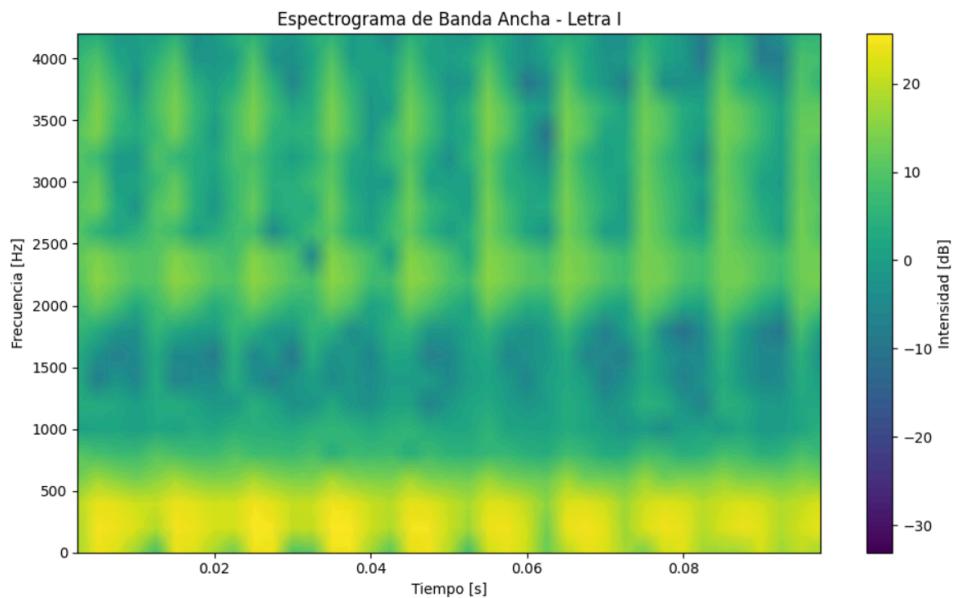


Figura 7: Espectrograma de banda ancha de un fragmento que contiene la vocal 'i.'

Letra 'a':

En función del espectrograma de banda angosta del fragmento [2, 3.2] que contiene la vocal A (Fig. 8), es posible señalar que la frecuencia fundamental de la letra se encuentra aproximadamente en

$f_{0(a)} = 125 \text{ Hz}$. Sus armónicos están ubicados en este caso en $f_{1(a)} = 250 \text{ Hz}$, $f_{2(a)} = 375 \text{ Hz}$, $f_{3(a)} = 500 \text{ Hz}$, y en los subsiguientes múltiplos de la frecuencia fundamental.

Para esta vocal, los formantes se enumeran a continuación: $F_{1(a)} \approx 500 \text{ Hz}$, $F_{2(a)} \approx 1400 \text{ Hz}$, $F_{3(a)} \approx 2200 \text{ Hz}$ y $F_{4(a)} \approx 3500 \text{ Hz}$, como puede verse en la Figura 9.

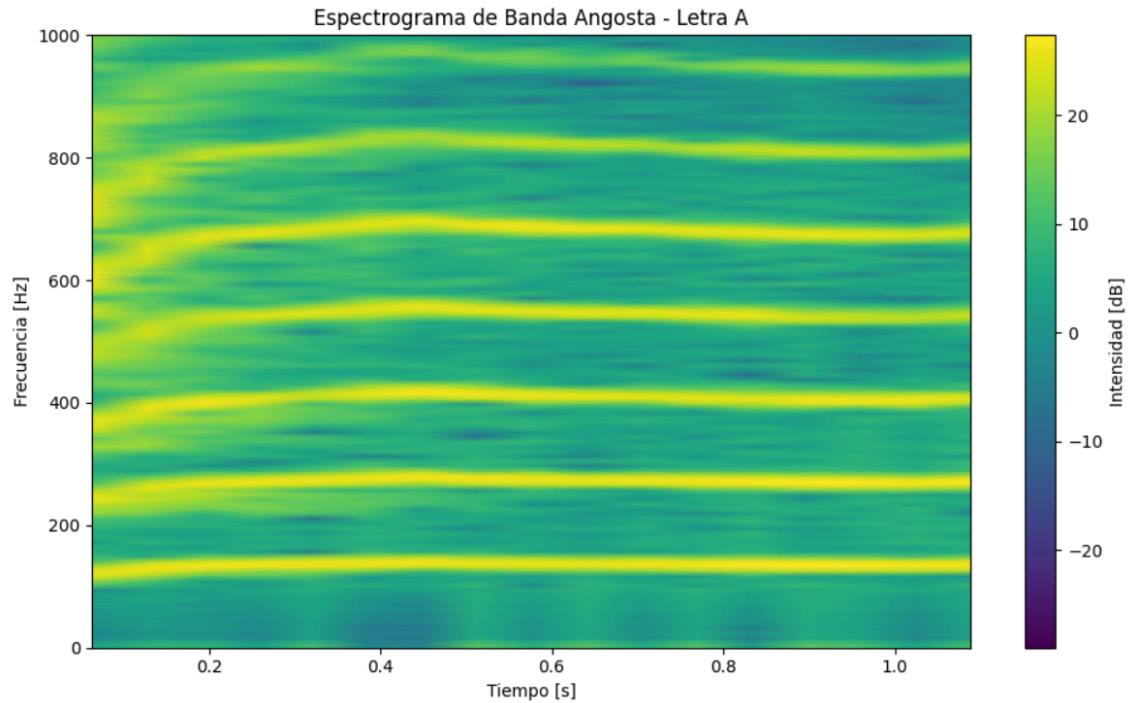


Figura 8: Espectrograma de banda angosta de un fragmento que contiene la vocal ‘a.’

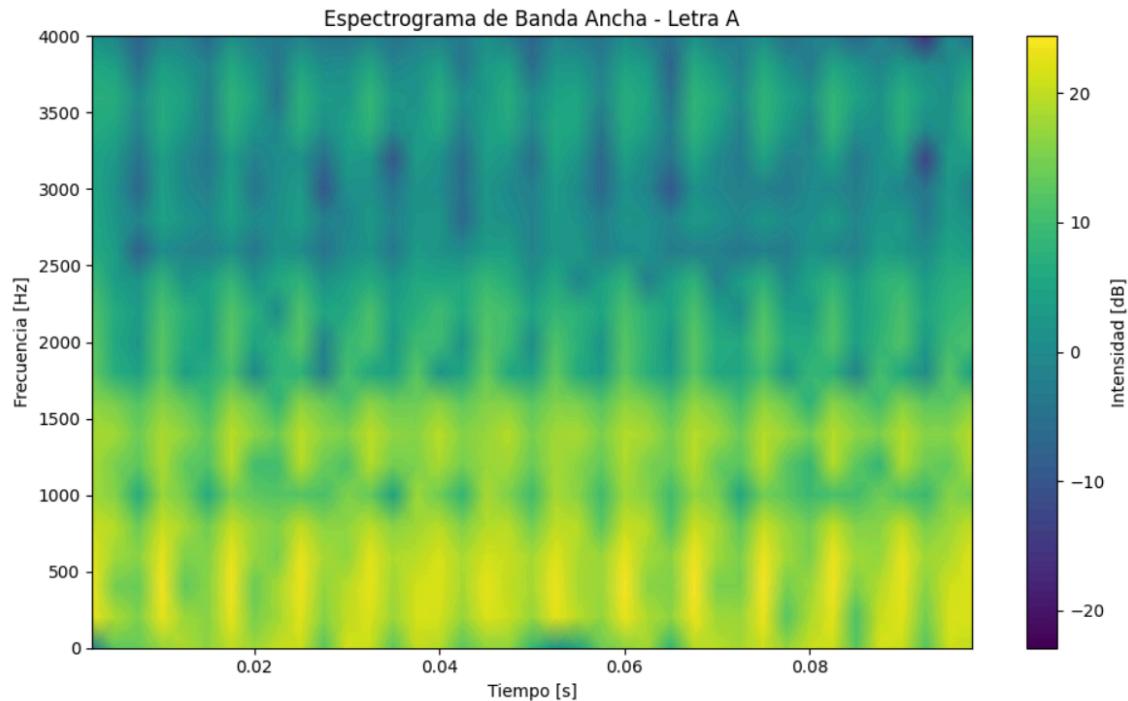


Figura 9: Espectrograma de banda ancha de un fragmento que contiene la vocal ‘a.’

Letra ‘o’:

De acuerdo con el espectrograma de banda angosta del fragmento [4. 5, 5. 5] que contiene la vocal O (Fig. 10), es posible señalar que la frecuencia fundamental de la letra se encuentra aproximadamente en $f_{0(o)} = 90 \text{ Hz}$. Sus armónicos se encuentran en $f_{1(o)} = 180 \text{ Hz}$, $f_{2(o)} = 270 \text{ Hz}$, $f_{3(o)} = 360 \text{ Hz}$, y en los subsiguientes múltiplos de la frecuencia fundamental.

Los formantes de la ‘O’ se ubican en $F_{1(o)} \approx 500 \text{ Hz}$, $F_{2(o)} \approx 2250 \text{ Hz}$ y $F_{3(o)} \approx 3250 \text{ Hz}$, como se indica en la Figura 11.

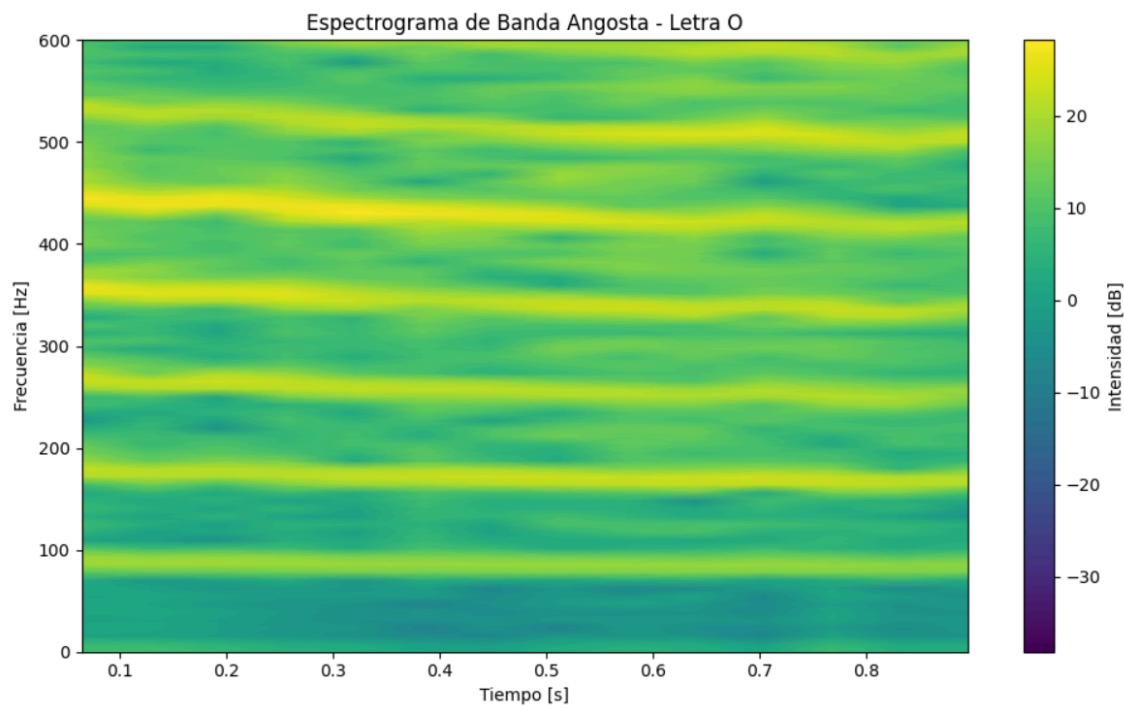


Figura 10: Espectrograma de banda angosta de un fragmento que contiene la vocal ‘o’.

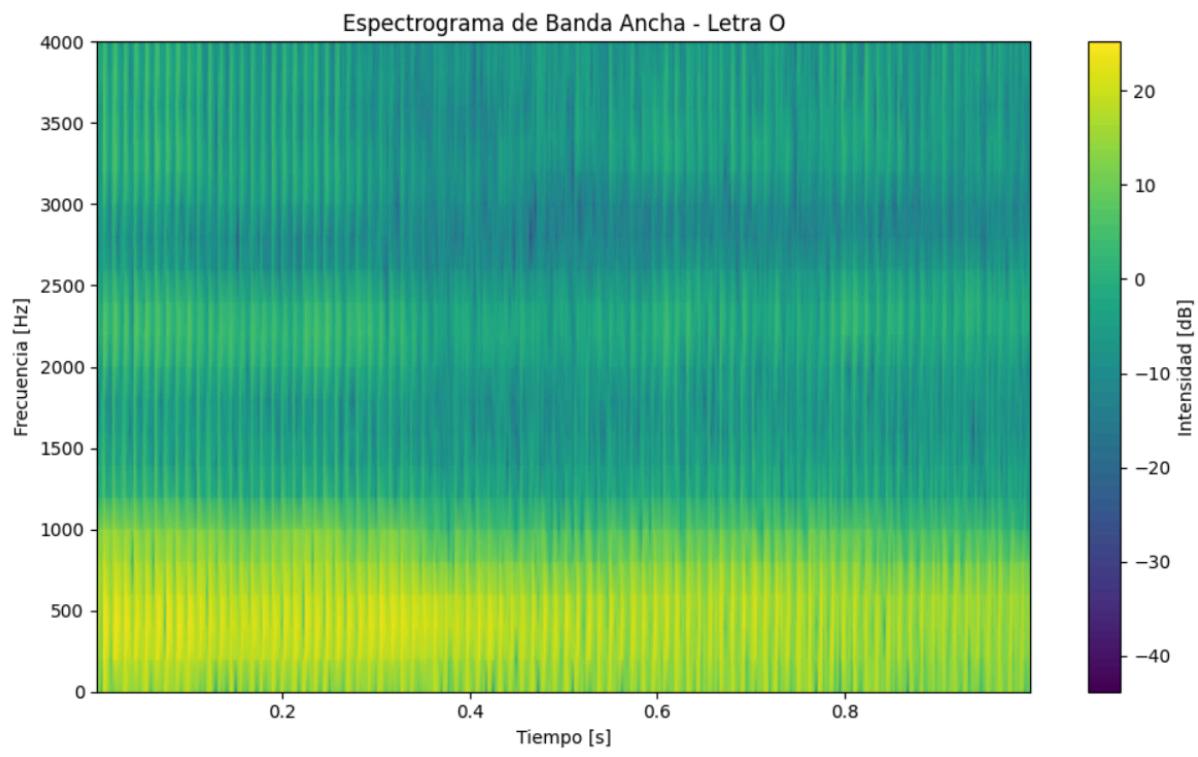


Figura 11: Espectrograma de banda ancha de un fragmento que contiene la vocal ‘o’.

Modificación de la frecuencia fundamental de la señal

Como se tiene la grabación de la señal realizada por un hombre, se buscará multiplicar la frecuencia fundamental de la señal por el factor 1.4, a modo de que ésta tenga un pitch más agudo. Se busca que la señal se agudice en un 40%. Se utilizará el método de Time Domain Pitch Synchronous Overlap and Add (TD-PSOLA) para poder variar la frecuencia fundamental y agudizar la señal sin modificar la velocidad ni la duración de la muestra; por ende, se espera tener al final una señal con la misma longitud temporal pero cuya frecuencia fundamental se encuentre más elevada.

Para ello primero se seleccionaron los fragmentos con sonidos armónicos a los cuales se les pudiese variar la frecuencia fundamental. Estos corresponden a los de las vocales “i,” “a” y “o.” Los inicios y finales de cada vocal que se utilizaron se muestran en la Tabla 1, y se obtuvieron a partir del espectrograma de banda ancha de la Figura 4.

	Tiempo de inicio (start time) [s]	Tiempo de finalización (end time) [s]
“i”	0,32	1,69
“a”	1,89	3,36
“o”	3,95	5,79

Tabla 1: Tiempos de inicio y final de cada vocal tomados para la realización del método TD-PSOLA.

Para cada fragmento se buscaron los picos de la señal entre los instantes de inicio y fin. Luego se buscaron los distintos períodos entre picos consecutivos para luego buscar el promedio de estos valores y así obtener el periodo medio de la señal. Luego, para cada pico hallado, se fija el inicio y el fin del mismo de modo tal que entre éstos haya dos períodos medios. Si esas ventanas no entran en los tiempos positivos o requieren de tiempos más allá de todo lo comprendido en la señal, se descarta ese pico. Lo presente originalmente centrado en ese pico se multiplica por una ventana de tamaño de dos períodos medios (para suavizar la señal y reducir el ruido). Esto se agrega a una ventana centrada en el pico por el factor de modificación de la antena, de modo que éste ahora se encuentra a un 40% de su posición original. Se agrega en esas posiciones lo obtenido de multiplicar la ventana por lo presente en la señal centrada en el pico original, para realizar el corrimiento temporal necesario para generar una señal de un periodo menor y, por ende, de una frecuencia mayor.

Finalmente, se desatura levemente la señal y se agregan las muestras faltantes para completar la muestra. Para evitar agregar muestras discretas en espacios nulos dentro de lo modificado mediante el método TD-PSOLA, se agregan únicamente muestras luego de que al menos 100 muestras consecutivas no se encontraron en la señal original. Esto también evita posibles ruidos que se puedan sumar a la señal por una cantidad reducida de muestras no incluidas en ninguna ventana. Finalmente se desnormaliza la señal y se escribe un archivo .wav con la muestra modificada.

Un aspecto relevante del resultado obtenido es que, por el método usado, las ventanas solo se están copiando en una porción de la señal (al acelerarse los períodos). Es por esto que quedan dos

“secciones” distinguidas en el spectrograma entre la parte en la cual se pudo efectivamente reducir el periodo fundamental (y por ende aumentarse el periodo fundamental) y la parte que, por cómo se realizó la operatoria, nunca tuvo picos (que llegan a centrarse hasta un periodo medio más allá del 40% del último pico) multiplicados por la ventana y desplazados temporalmente para reducir su frecuencia. Como se copiaron los pedazos de la señal original en esas secciones, su frecuencia fundamental es la previamente obtenida.

El spectrograma obtenido a partir de ésto es el presentado en la Figura 12.

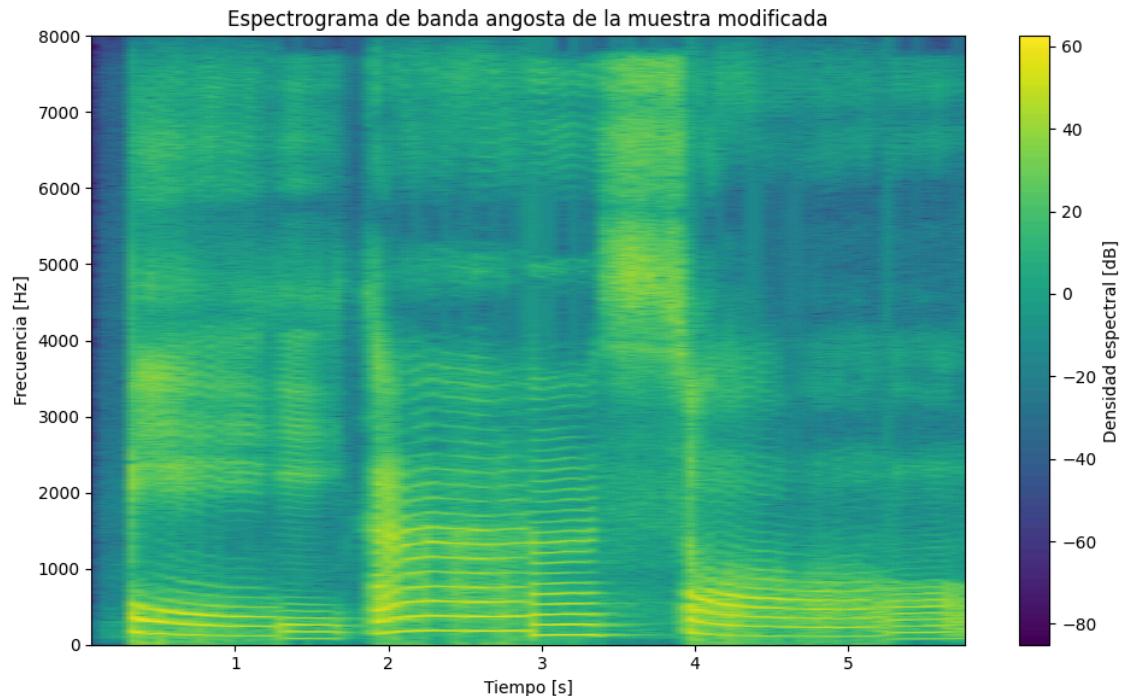


Figura 12: Espectrograma de la señal modificada completa.

Viendo el spectrograma con una escala tan amplia en frecuencia, resulta difícil notar las diferencias en frecuencia. Sí se puede observar una diferencia entre las franjas izquierdas y derechas dentro de los armónicos de cada vocal, con las frecuencias de la derecha encontrándose ligeramente por debajo de las de la izquierda. Para verlo con más claridad, en la Figura 13 se reduce la visibilidad del eje de frecuencias a 2 kHz.

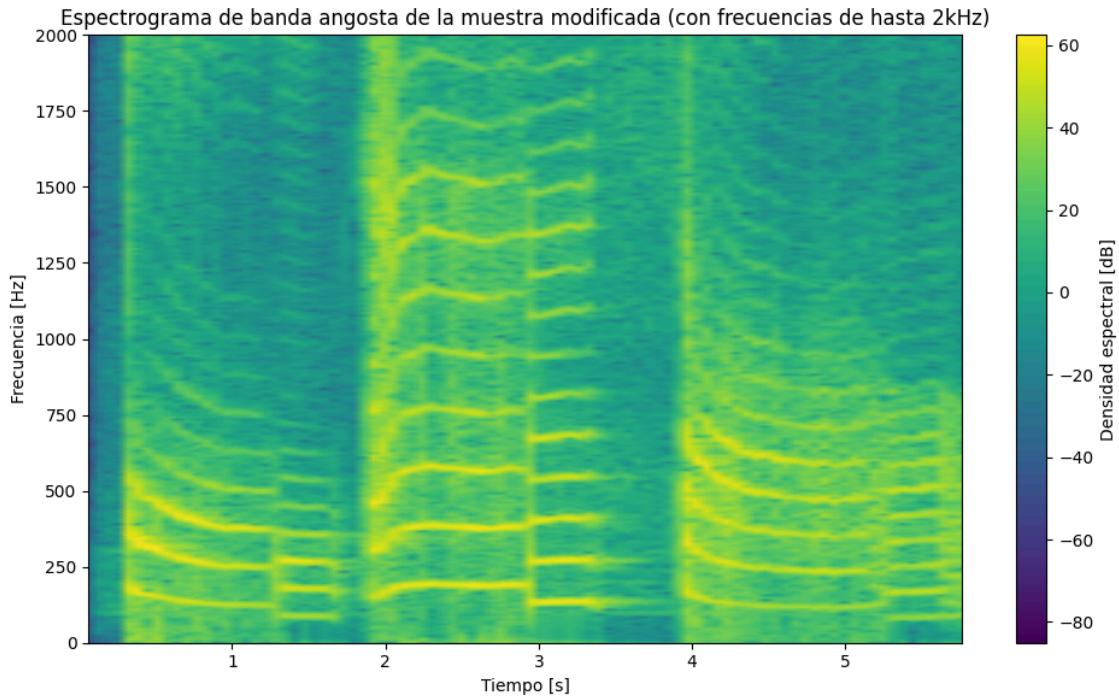


Figura 13: Espectrograma de la señal modificada completa, viéndose las frecuencias hasta 2kHz.

En esta Figura 13 se aprecia el aumento de la frecuencia fundamental, en especial al compararlo con los fragmentos de la derecha que no parecen haberse levantado de su frecuencia original. Esto se debe al algoritmo utilizado para hallar los nuevos picos, como se explicó previamente.

Se realizó también un gráfico temporal para comparar las señales originales y modificadas, obteniéndose lo mostrado en la Figura 14.

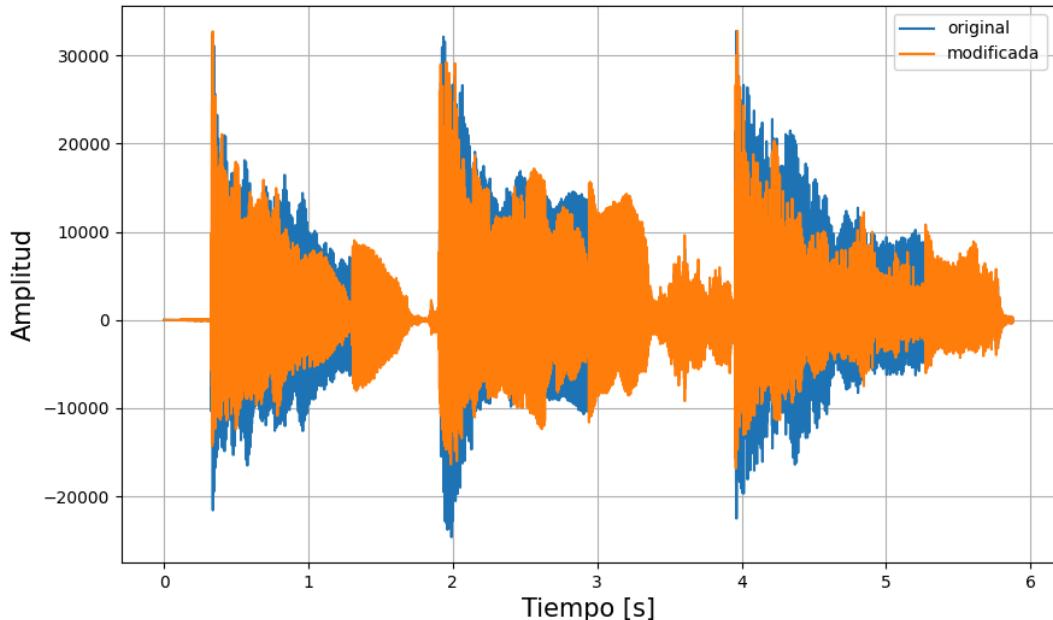


Figura 14: Comparación de la señal temporal modificada y la original.

Se observa que la energía máxima de las señales es la misma, y se notan las diferencias del “emparchado” de las secciones en las cuales no se realizó correctamente el TD-PSOLA (como entre los 3,00 segundos y los 3,36 segundos para el sonido “/a/”).

Conclusión

A lo largo de esta segunda parte del Trabajo Práctico Especial se utilizaron los espectrogramas de banda angosta y ancha para hallar la frecuencia fundamental de los sonidos armónicos producidos con la voz, y posteriormente se buscó aumentar la frecuencia fundamental de la voz para agudizarla, verificando que no pudo realizarse por completo en todos los armónicos de la señal mediante el uso de espectrogramas. También se pudo visualizar en el tiempo cómo afectaba a la señal la variación de la frecuencia fundamental, reduciendo las amplitudes de la misma y viendo claramente las secciones en las que se realizó el overlap and add en el proceso con aquello que no pertenecía a los sonidos armónicos o en lo que no se podía incluir una ventana completa.