

Análisis de una señal musical en frecuencia

Duque Karla A.
kduque@fi.uba.ar
108406

Rodriguez Guido E.
guerodriguez@fi.uba.ar
108723

Vargas Joaquín
jvargas@fi.uba.ar
104323

Dietrich Pablo
pdietrich@fi.uba.ar
83383

Abstract—Se analizan señales musicales en los dominios del tiempo y la frecuencia mediante la Transformada de Fourier y la FFT. Se generan espectrogramas utilizando diferentes ventanas de muestreo para comparar resolución temporal y frecuencial, y se implementan efectos musicales interpretados como sistemas lineales y no lineales.

I. INTRODUCCIÓN

En este informe se presenta el análisis de dos señales musicales en las que a diferencia del informe anterior se analizarán en el dominio de la frecuencia mediante la Transformada de Fourier, además se observarán algunos efectos utilizados en la producción musical.

A. Transformada de Fourier

La transformada de Fourier, en particular la transformada rápida de Fourier (de ahora en adelante **FFT**), es un algoritmo que en forma sintética nos permite descomponer una señal en el dominio del tiempo (no necesariamente periódica) como la suma ponderada de señales periódicas armónicas. La descomposición en frecuencia nos presentará el denominado espectro de la señal, mediante este podremos analizar y hallar información contenida en la señal que de hacerlo en el dominio del tiempo sería difícil o incluso imposible de identificar.

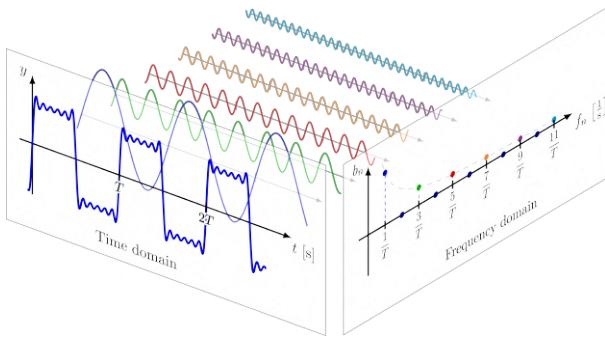


Figura 1. Ilustración de la descomposición de una señal en el tiempo en sus componentes armónicas

En este sentido, resulta sencillo identificar por simple inspección del espectro las notas musicales que componen una melodía o distinguir diferentes instrumentos dentro de una mezcla, ya que cada uno posee un patrón espectral característico.

B. Espectrograma

Una herramienta muy útil para analizar la evolución del espectro de la señal en el tiempo es el espectrograma, que nos permite visualizar la energía de la señal para diferentes ventanas temporales. Habitualmente se habla de espectrogramas de ventana angosta/banda ancha o ventana ancha/banda angosta, lo que muestra la importancia de la elección del tamaño de la ventana. En caso de elegir una ventana ancha, es decir contar con un espectrograma de banda angosta, se gana resolución en la frecuencia, pudiendo identificar de forma inequívoca las componentes armónicas presentes en

el espectro, pero a costa de perder la capacidad de observar variaciones rápidas en la energía a lo largo del tiempo. Por el contrario, si se escoge una ventana angosta, es decir un espectrograma de banda ancha, ganaremos resolución en el tiempo permitiendo observar variaciones rápidas en la energía de la señal; sin embargo, se dificulta la identificación de componentes de frecuencia próximas entre sí, ya que la imagen se verá menos definida en el eje de frecuencia

Este comportamiento nos muestra cómo el principio de incertidumbre se manifiesta al momento de elegir una ventana u otra, ya que, a medida que obtenemos mayor detalle o resolución en el dominio de la frecuencia, perdemos precisión temporal, y viceversa.

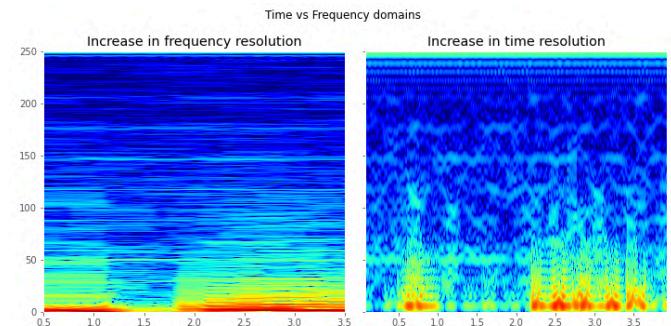


Figura 2. Ejemplo de un espectrograma de la misma señal con una ventana angosta (izq.) y una ventana ancha (der.)

En la Figura 2 se presenta un ejemplo de espectrogramas sobre una misma señal evidenciando en cada uno de ellos las características antes mencionadas en relación al tamaño de la ventana utilizada. En el espectrograma de la izquierda se distinguen “líneas horizontales” bien definidas, correspondientes a una gran definición en la frecuencia, pero incapaz de identificar variaciones rápidas de la señal como si se logra evidenciar en el espectrograma de la derecha donde se observan con mayor detalle los cambios o “saltos” de energía en el tiempo, pero con una menor definición en las frecuencias entre las que se producen dichos cambios.

II. DESARROLLO

A continuación se presenta el desarrollo involucrado para obtener la representación espectral de las señales de audio y su posterior análisis.

A. Representación de las señales en el dominio de la frecuencia

Haciendo uso del script en python realizado para el trabajo práctico anterior, extendiendo las funcionalidades de la clase desarrollada para el análisis de señales se obtiene la magnitud de la FFT de ambas canciones presentadas en

la Figura 3 y Figura 4, en ellas por simple inspección se escogieron tres tramos ordenados de mayor a menor en función de la densidad de información observada. En el caso de la canción 1 (Figura 3), la cual se ejecuta únicamente con instrumentos musicales, se nota que la mayor cantidad de información se encuentra en el rango de 0 a 2kHz y en menor medida desde 2kHz hasta 8kHz debido a los armónicos de las notas presentes en el primer tramo mencionado.

En lo que respecta a la canción 2 (Figura 4), el segmento de frecuencias con mayor predominancia de información es el que se encuentra entre los 0 y 500Hz, en particular entre los 40 y 200 Hz lo cual se explica, como fue mencionado en

el trabajo práctico anterior, en que esta pieza musical se compone principalmente de una percusión marcada por aplausos y golpes como así también la voz de humana masculina de Freddy Mercury la cual se clasifica como la de un barítono. Según el análisis realizado en el *paper* [1] los límites extremos de su voz se encuentran entre 87 y 253Hz.

Un aspecto a destacar es la mayor “pulcritud” del espectro de la canción 1 en comparación con el de la canción 2. Como se mencionó anteriormente, la canción 1 se compone únicamente por instrumentos musicales que a diferencia de la canción 2, compuesta por percusión y la voz humana generan notas definidas con menos componente armónica.

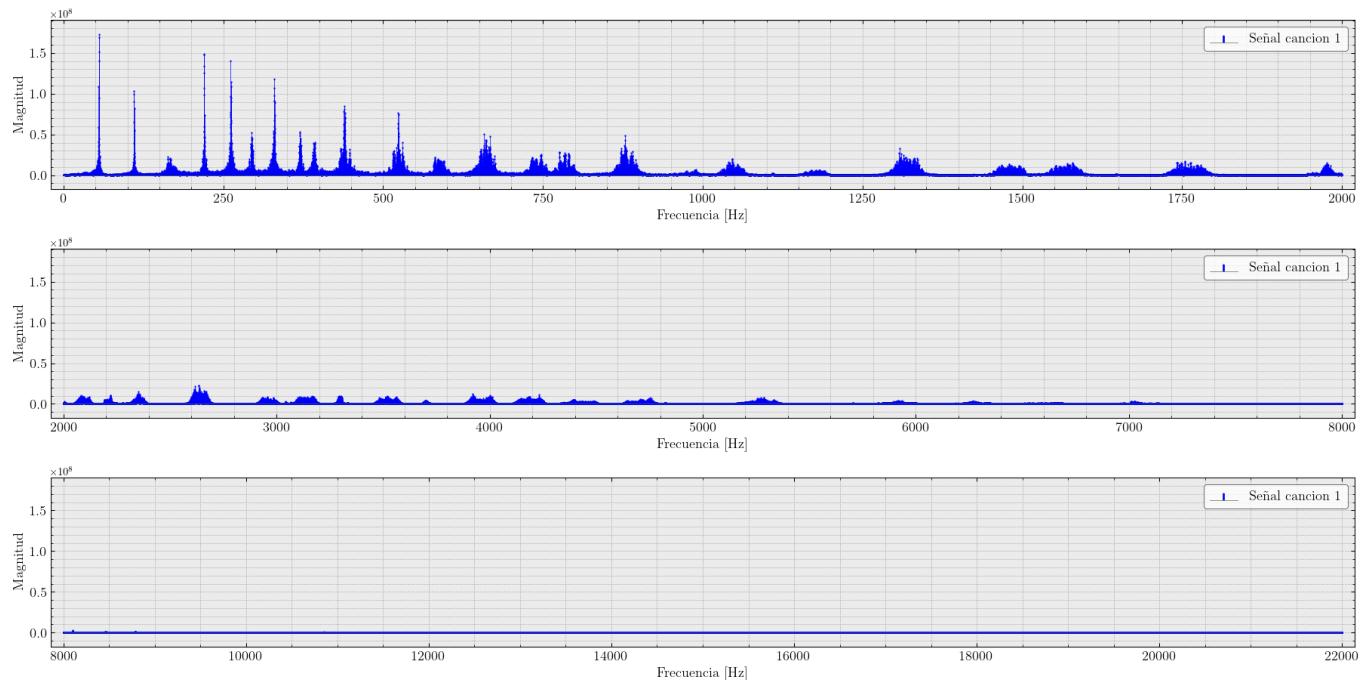


Figura 3. Espectro de la canción 1

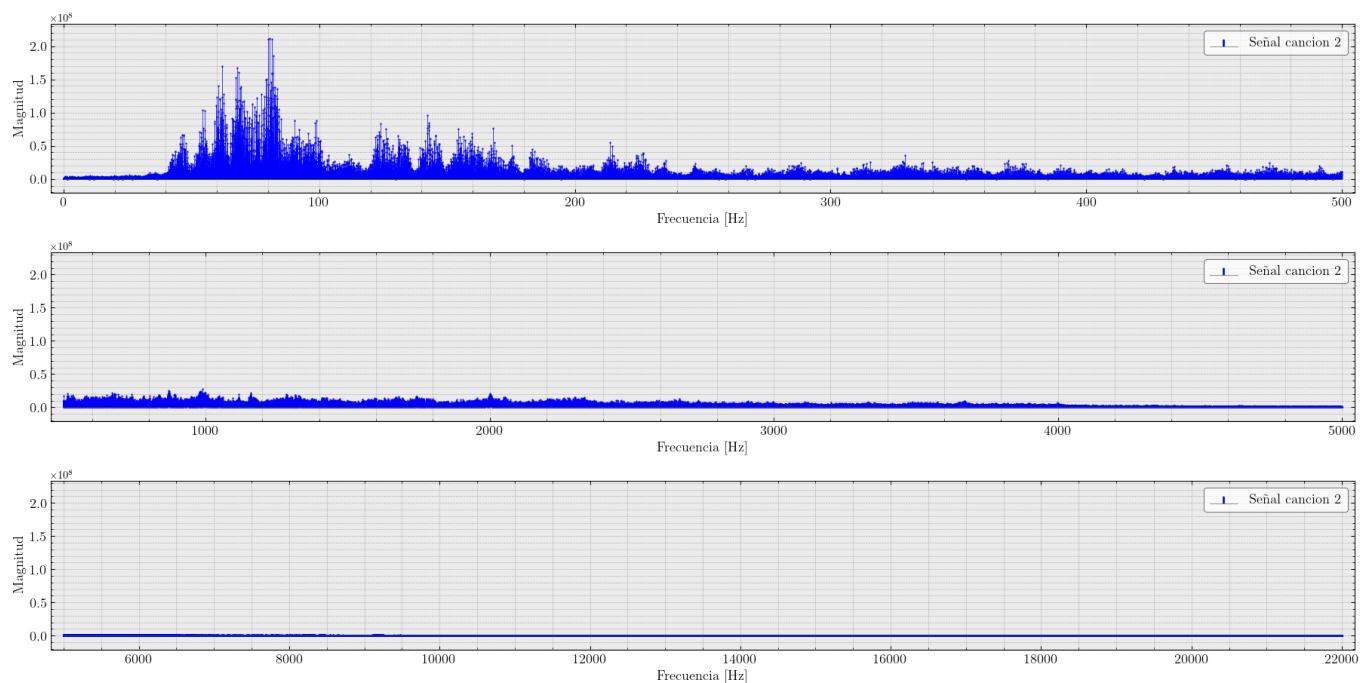


Figura 4. Espectro de la canción 2

Asimismo, con el objetivo de analizar la información contenida en el espectro de la canción 1, en la Figura 5 se pueden identificar regiones representativas que abarcan desde la octava 0 hasta la 8, donde se observan las notas musicales [2] junto con sus respectivos armónicos. El espectro muestra frecuencias fundamentales y armónicos bien definidos, destacándose una fundamental en La-1 con armónicos que se extienden hasta La-8, este último con la menor energía. De manera similar, se identifican fundamentales en Mi-3, Re-4, Do-4 y Si-6.

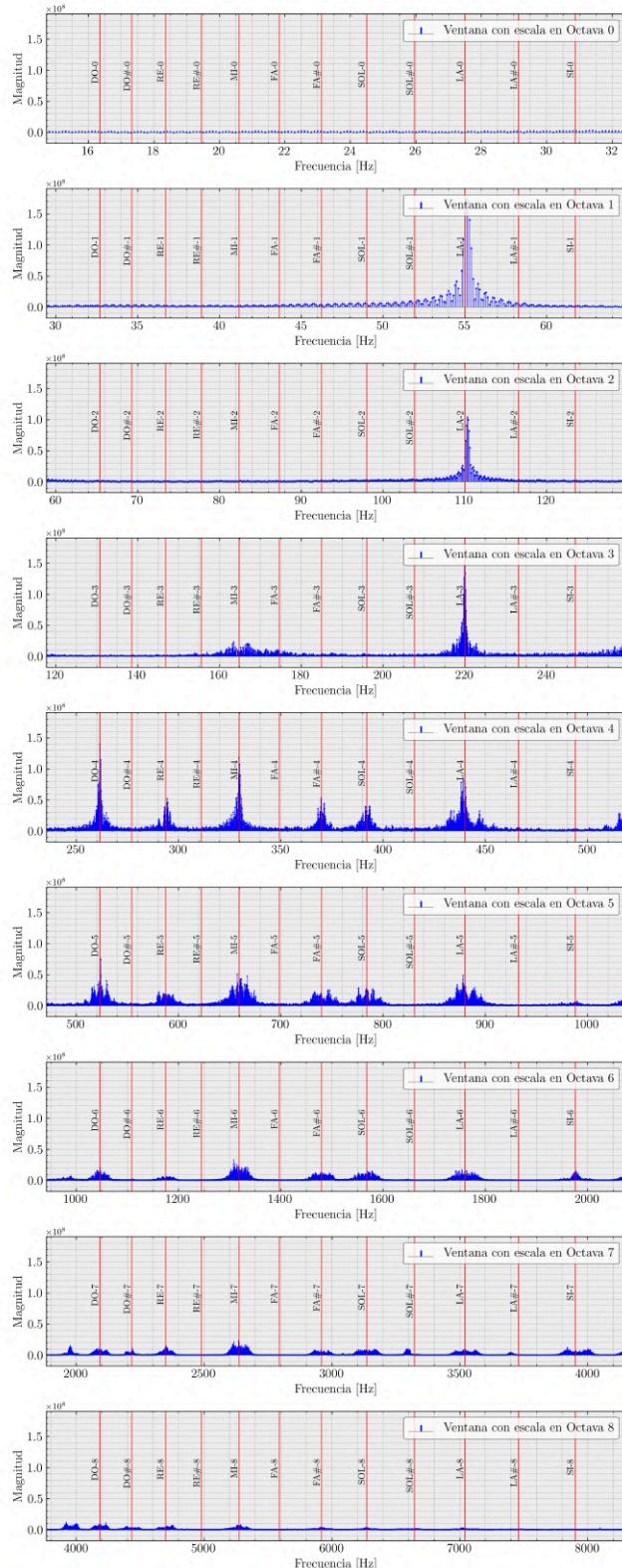


Figura 5. Espectro de la canción 1 con ventanas para cada octava

En lo que respecta a la canción 2, en la Figura 6 se observa un espectro con la energía distribuida, sin la presencia de notas inequívocamente definidas. Esto se debe a que la canción está compuesta principalmente por percusión y voz humana. La percusión se dice que tiene un comportamiento transitorio lo que genera un contenido armónico “amplio”, mientras que la voz humana, al variar continuamente en frecuencia, produce un espectro más denso e irregular en comparación con el de la canción 1 lo que dificulta la identificación de notas musicales

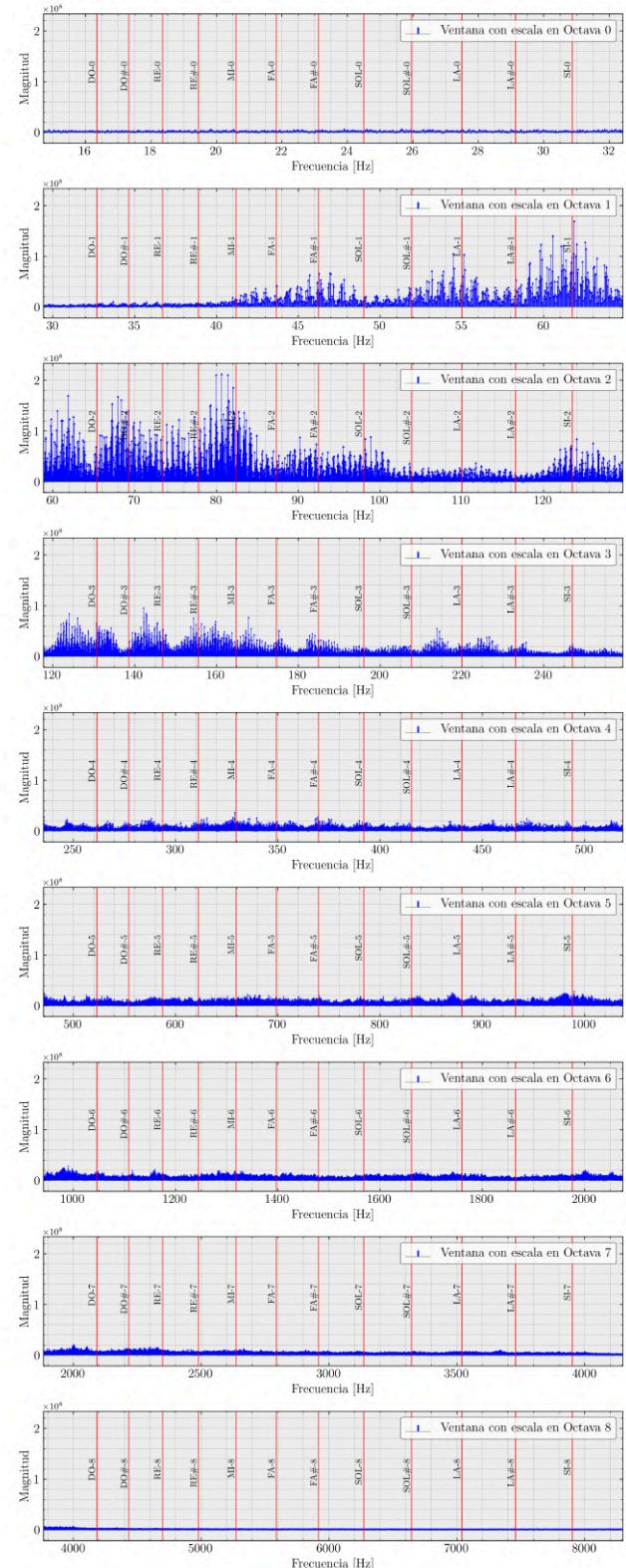


Figura 6. Espectro de la canción 2 con ventanas para cada octava

Es importante mencionar que si bien este análisis nos permite conocer las notas presentes en la canción, no nos permite saber el tiempo en el que estas ocurren como si podríamos hacerlo con el spectrograma.

B. Señales filtradas

En la Figura 8 y Figura 9 se muestran los espectros de la señal correspondiente a la canción 1 y 2 luego de ser convolucionadas con los filtros 1 y 2 presentados a través de su respuesta impulsiva en la Figura 7.

Tal como se mencionó en el trabajo práctico anterior, luego de observar el espectro de ambas respuestas confirmamos que el filtro 1 corresponde a un filtro pasa altos, mientras que el filtro 2 se comporta como un pasa bajos

pues en ambos espectros se observa la atenuación de las componentes armónicas por debajo y por arriba de una frecuencia de corte respectivamente. Analizando gráficamente, en el caso del filtro 1 la frecuencia de corte se encuentra aproximadamente entre 850 Hz y 900 Hz, lo que implica que atenúa los armónicos por debajo de esa frecuencia, eliminando las componentes graves y dejando pasar las frecuencias medias y altas. Por otro lado, el filtro 2 presenta una frecuencia de corte entre 4250 Hz y 4500 Hz, atenuando las componentes agudas y conservando el contenido de baja frecuencia. Asimismo, se observa que elimina los armónicos de alta frecuencia con una pendiente de atenuación más pronunciada que la del filtro 1, lo que nos da a entender que este es de una mayor selectividad y de orden superior.

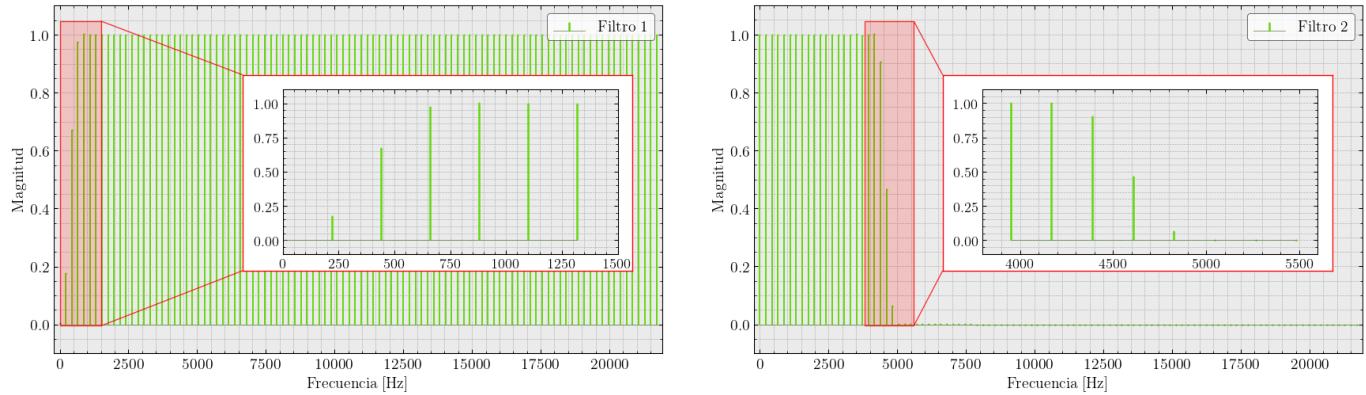


Figura 7. Espectro de la respuesta al impulso del filtro 1 y 2

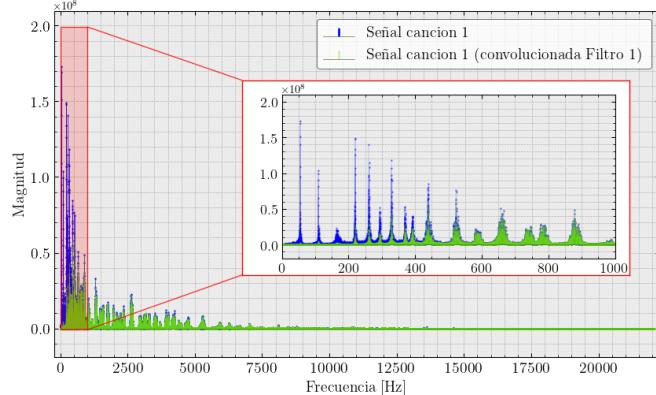


Figura 8. Espectro de la canción 1 filtrada con el filtro 1 y 2

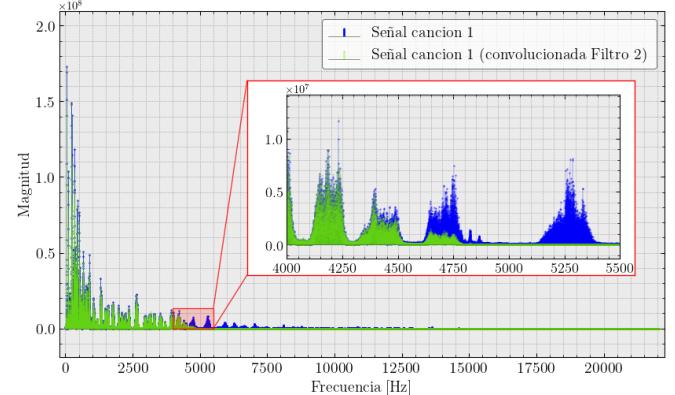
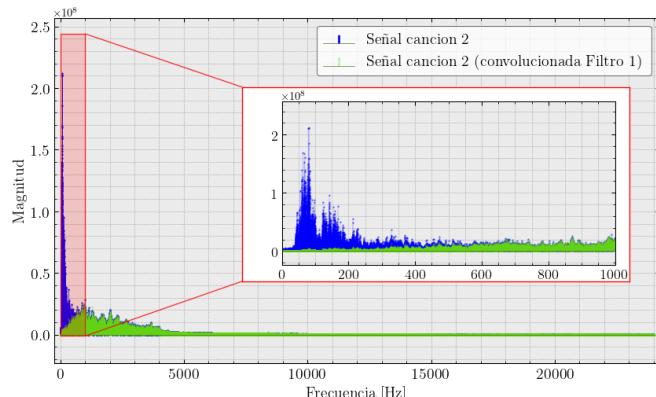


Figura 9. Espectro de la canción 2 convolucionada con el filtro 1 y 2

Teniendo en cuenta que el sistema 1 es un filtro pasa altos al pasar la canción 1 y la canción 2 por dicho sistema se evidencia que disminuyen las amplitudes que se corresponden con la entrada. Otro detalle, el sistema no crea armónicos nuevos de los ya existentes por lo que se puede decir que es un sistema LTI.

Con respecto al sistema 2 se puede observar un efecto contrario al anterior. Es decir, elimina o atenua los armónicos de alta frecuencia. Ademas, no crea armónicos nuevos. Por lo tanto, el filtro 2 tambien es un sistema LTI.

C. Análisis mediante espectrogramas

Como se mencionó anteriormente, en un espectrograma es posible determinar las notas que componen una melodía mediante el uso de una ventana ancha/banda angosta. Este tipo de ventana ofrece una mayor resolución en frecuencia, lo que permite observar con claridad las frecuencias correspondientes a cada nota. Sin embargo, presenta una menor resolución temporal, por lo que no es posible identificar con precisión el momento exacto en que cada nota se produce dentro de la melodía.

Al momento de generar el espectrograma se debió escoger una ventana de muestreo, de las cuales decidimos utilizar Hann, Hamming y rectangular siguiendo las recomendaciones de [3] y [4]. Asimismo, se evaluaron distintos tamaños de ventana y valores de *overlap* entre los *chunks* de datos. Todos estos parámetros se eligieron como potencias de 2, de acuerdo con las recomendaciones de la biblioteca matplotlib, por razones de eficiencia computacional.

La ventana Hann [5] suaviza los bordes, mostrando bandas más definidas, mientras que la rectangular mantiene mayor resolución temporal pues es plana en la región de ventaneo pero el salto abrupto en los extremos genera discontinuidades en la señal ventaneada lo que provoca dispersión en las frecuencias al introducir espurias que pueden solaparse y dificultar la identificación de las componentes armónicas originales como puede observarse en la Figura 10 obtenida de [6] donde el piso del espectro se ve aumentado producto de la dispersión en frecuencia.

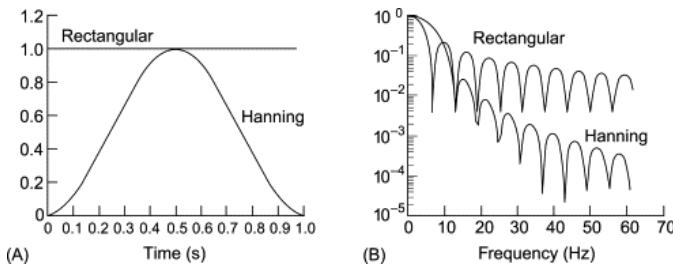


Figura 10. Comparativa entre una señal muestreada con ventana Hann y Rectangular

En la jerga, este efecto se lo conoce como *spectrum smeared* o “espectro untado” por la forma visual en la que se presenta en un espectrograma como el de la Figura 11 donde la figura superior en las frecuencias más altas que originalmente como hemos visto en la Figura 3 eran casi nulas, ahora se ven aumentadas producto de la dispersión provocada por la ventana rectangular, exhibiendo cierta amplitud no despreciable. Sin embargo, en el caso de ventana Hann se logra diferenciar con un salto casi abrupto a mínimos ya que conserva de una mejor forma las amplitudes originales para los armónicos de alta frecuencia.

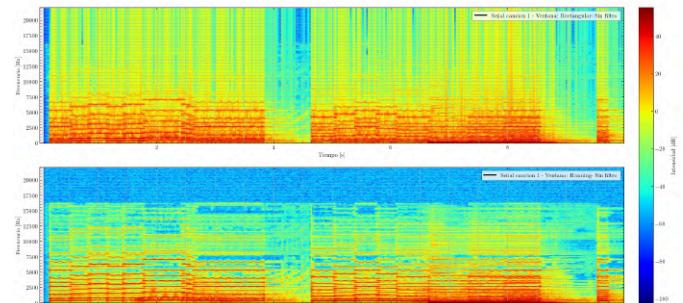


Figura 11. Espectrograma de la canción 1 con ventana rectangular y Hann

Por esta razón, se prefieren ventanas como la Hann o Hamming, que introducen un suavizado progresivo en los bordes de la ventana, reduciendo así las discontinuidades y mejorando la resolución espectral. En particular, la ventana Hann tiende a llevar a cero los valores de la señal en los lóbulos extremos, lo que elimina las discontinuidades en los bordes. En cambio, la ventana Hamming no lleva a cero los bordes, pero suaviza de manera más gradual la transición entre los valores de la señal como se observa en la Figura 12 donde de forma esquemática se presenta las mencionadas transiciones en los lóbulos de ambas ventanas.

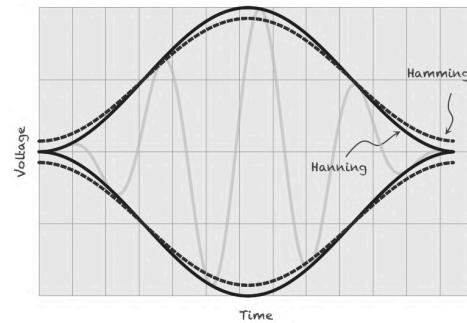


Figura 12. Comparativa entre una ventana Hann y Hamming en el tiempo

Al seleccionar distintos valores de ventana, se observó que a medida que el tamaño de esta aumentaba, la definición en frecuencia mejoraba visualmente, hasta distinguir únicamente líneas horizontales correspondientes a los armónicos a costa de difuminarse en el tiempo sin lograr apreciar cualquier variación. Por el contrario, al utilizar ventanas más pequeñas, como se muestra en la Figura 13 para las ventanas Hann y Hamming.

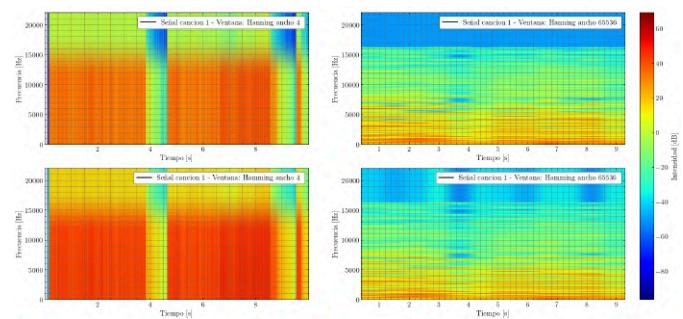


Figura 13. Espectrograma de la canción 1 con una ventana Hann y Hamming para tamaños chicos y grandes

A su vez notamos que la ventana Hann, para valores cercanos a cero, proporcionaba una mejor representación que Hamming. Esto es especialmente evidente para frecuencias superiores a 16 kHz, donde, como se mencionó previamente, la canción 1 es prácticamente nula.

Luego de realizar los espectrogramas de cada canción y los resultantes de ser estas filtradas se puede observar en la Figura 14 y en la Figura 15 se confirma que el filtro 2 es un filtro pasa bajos, pues este elimina la energía de la señal que se encuentra en frecuencias superiores a la de corte. Análogamente, el filtro 1 es un filtro pasa altos, pues elimina las frecuencias inferiores a su frecuencia de corte. Además se debe mencionar que los espectrogramas de la canción 1 no presentan una gran dispersión en frecuencia como si lo hacen los espectrogramas de la canción 2 esto es debido a que en la canción 1 esta compuesta únicamente por instrumentos y la canción 2 esta compuesta por percusión (aplausos) y voz humana la cual genera transitorios por lo tanto una dispersión en frecuencia.

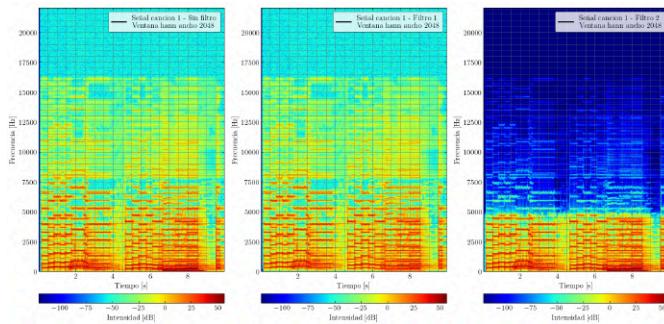


Figura 14. Espectrograma de la canción 1 y el resultado de ser convolucionada con el filtro 1 y 2

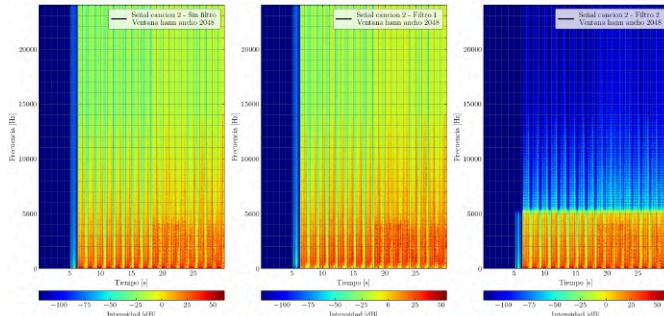


Figura 15. Espectrograma de la canción 1 y el resultado de ser convolucionada con el filtro 1 y 2

D. Detección de notas en el espectrograma

De forma suplementaria al análisis anterior, concluimos que si bien podemos identificar las notas existentes, no podemos determinar el momento exacto en que ocurrieron, imposibilitando discernir si son armónicos de una misma nota en octavas menores o notas efectivamente ejecutadas. Por ello, con el espectrograma confirmamos si las notas relevadas son armónicos o notas reales observando los tonos más intensos del gráfico en torno a sus frecuencias y duración. Se dividió el espectro en tramos por octavas, como se muestra en la Figura 16.

Al dividir el espectrograma en octavas (0 a 8) identificamos la duración y momento de ejecución de las notas. En la canción 1 predominan notas desde la cuarta octava en adelante, destacando MI, FA#, SOL y LA. En la primera octava la energía es intensa entre SI y SOL#-1, aunque difícil de precisar su origen; en octavas superiores la energía persiste acercándose a un LA-#, por lo que, al surgir en la primera octava como armónico, suponemos que proviene de un LA-#1.

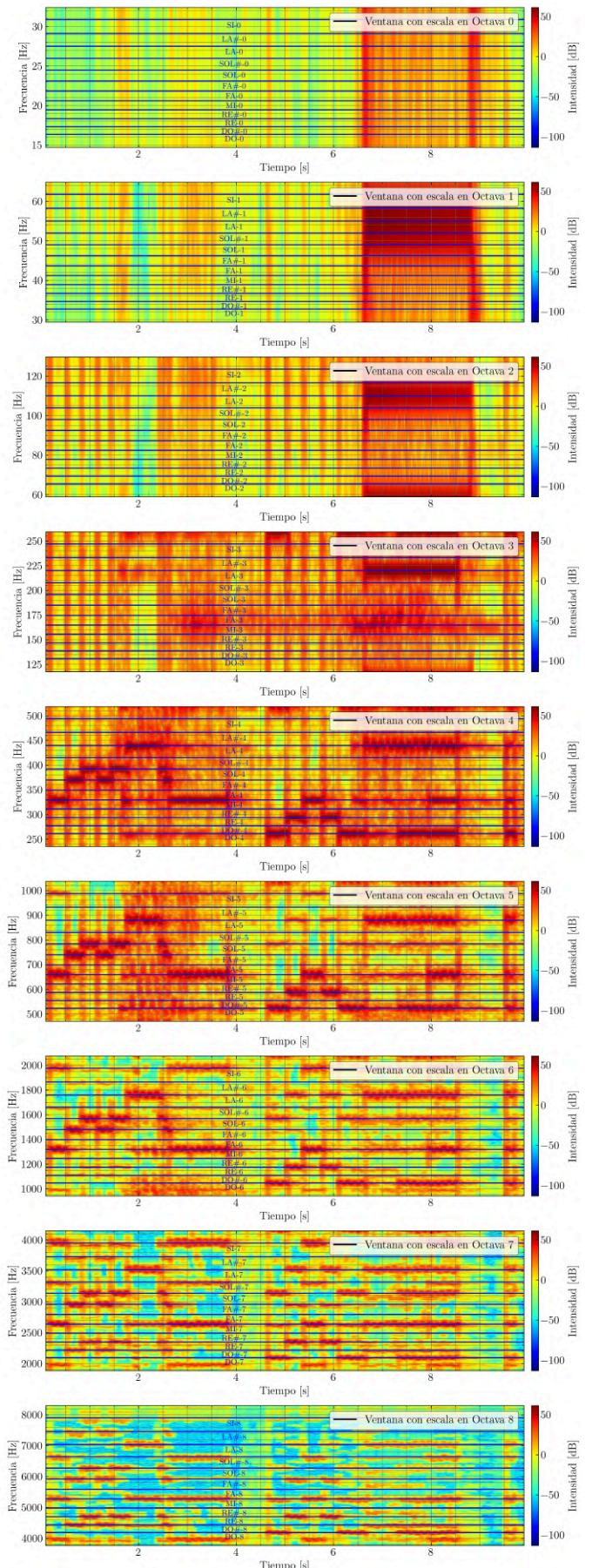


Figura 16. Espectrograma de Canción 1 dividido por octavas

E. Efectos Musicales

A continuación se explica como se producen algunos efectos musicales y que propiedades tienen:

a) *Delay*: El efecto delay consiste en sumar a la señal original una o varias copias con retardo en el tiempo:

$$y(t) = x(t) + \alpha x(t - \tau) \quad (1)$$

Es un sistema lineal, invariante en el tiempo y causal, este efecto también es estable solo si $|\alpha| \leq 1$. Produce ecos o repeticiones que generan sensación de espacio.

b) *Distorsión*: La distorsión aplica una función (f) no lineal de saturación o recorte sobre la amplitud:

$$y(t) = f(x(t)) \quad (2)$$

Es un sistema no lineal, invariante en el tiempo, causal y estable. Genera un aumento de armónicos y un sonido más intenso o agresivo. Debido a la primera característica el sistema que realiza este efecto en la señal de entrada no es un sistema LTI.

Esto genera armónicos adicionales, cambiando el timbre del sonido sin alterar su duración ni su frecuencia fundamental. En el dominio temporal se observa un recorte de picos, y en el espectro un aumento de contenido en altas frecuencias.

c) *Tremolo*: El tremolo modula la amplitud de la señal mediante una onda de baja frecuencia:

$$y(t) = [1 + m \cdot \sin(2\pi ft)]x(t) \quad (3)$$

Sistema lineal, no invariante en el tiempo, causal y estable. Produce una variación periódica del volumen.

d) *Vibrato*: El vibrato consiste en una modulación periódica de la frecuencia de una señal de audio. En términos de un sistema se puede decir que es un retardo variable en el tiempo

$$y(t) = x(t - T_d(t)) \quad (4)$$

siendo $T_d(t) = T_0 + A \sin(2\pi f_m t)$ siendo f_m la frecuencia de modulación típicamente entre los 4 y 7 Hz y A es la amplitud con la que oscila y esta entre los 1ms y los 5ms y T_0 es el tiempo promedio de delay tipicamente esta entre 5ms y 10ms En términos de sistema no es lineal, no invariante en el tiempo, causal y es estable. e) *Chorus*: El chorus es un sistema lineal, causal, estable y variante en el tiempo que combina la señal original con una o más copias retardadas y moduladas.

$$y(t) = x(t) + g \cdot x(t - \tau(t)) \quad (5)$$

En donde la función de retardo suele ser una senoidal. Esto hace que el sistema deje de ser invariante en el tiempo, por lo tanto no es un sistema LTI. En términos coloquiales, el sistema produce: ligeras deferencias de tono (asociado a efecto Doppler) y pequeñas interferencias constructivas/destructivas en las frecuencias.

Luego, en la Figura 17 se muestran los gráficos en el tiempo con su correspondiente spectrograma para cada uno de los efectos mencionados anteriormente.

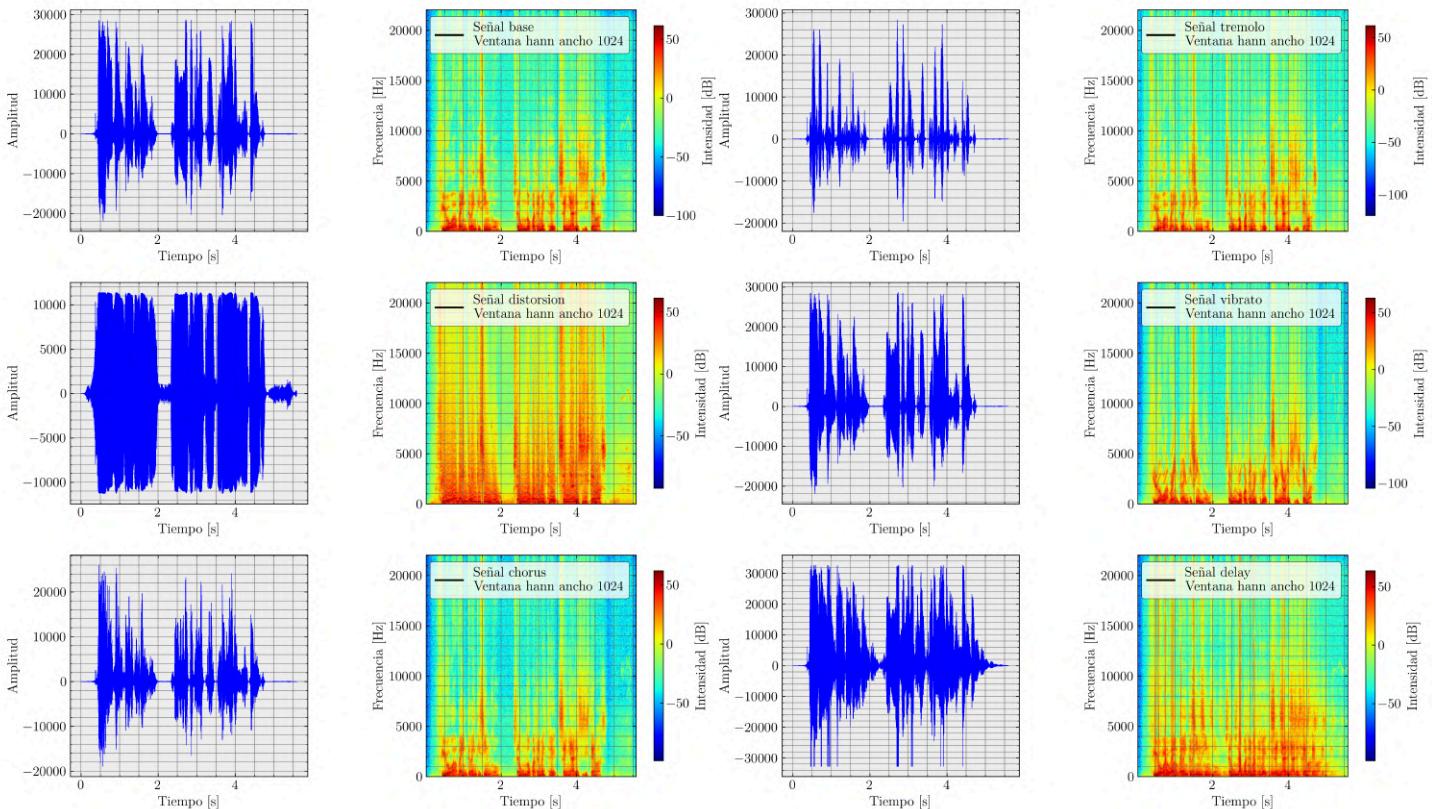


Figura 17. Frase “FIUBA 2025 señales y sistemas segundo cuatrimestre” con distintos efectos musicales

F. Implementación de Efectos Musicales

Se generó una melodía musical, sobre la cual se aplicaron dos efectos: delay y distorsión. A continuación se presentan los gráficos temporales y espectrogramas correspondientes a la señal original y a las señales procesadas con cada efecto.

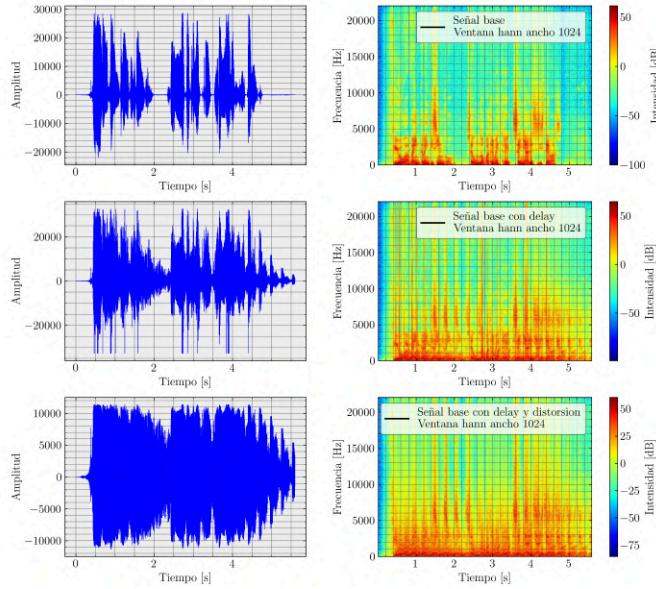


Figura 18. Melodía con la progresión de la señal en el tiempo y su espectrograma asociado para cada efecto aplicado

G. Descomposición en Serie de Fourier mediante la Transformada de Fourier

Se analiza la nota A4 obtenida con distintos instrumentos. Se selecciona un tramo cuasiperiódico de cada señal y, dentro de él, se tomaron intervalos que abarcan uno o varios períodos para calcular la FFT y obtener los coeficientes de la Serie de Fourier de la zona cuasiperiódica. Luego, mediante la IFFT, se reconstruyeron las señales limitando la cantidad de armónicos con el fin de eliminar componentes de alta frecuencia y reducir la información sin afectar perceptiblemente la señal original.

Al analizar la reconstrucción temporal se observa que la señal obtenida a partir de un solo período se asemeja a la original únicamente dentro de ese intervalo, pero luego aparecen diferencias y desfases. Esto se debe a que la nota A4 real es cuasiperiódica: cada ciclo presenta leves variaciones en amplitud, fase y contenido armónico.

Cuando se toma un solo período para la FFT, se asume que esa forma de onda se repite indefinidamente, generando una versión idealizada y provocando discontinuidades en los extremos del tramo, lo que introduce energía artificial en frecuencias no armónicas.

En cambio, al considerar varios picos (o varios períodos cuasiperiódicos), la FFT refleja mejor las variaciones naturales del sonido y el timbre del instrumento, reduciendo los efectos de borde. Sin embargo, esto implica una menor precisión temporal, ya que la señal deja de ser estrictamente periódica.

Seguidamente, se muestran las imágenes que ilustran el análisis realizado en Figura 19, Figura 20, Figura 21 y Figura 22.

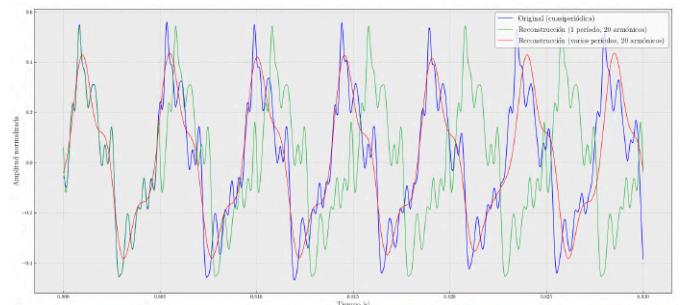


Figura 19. Señal original y reconstruida de un Saxo en La4 en el tiempo

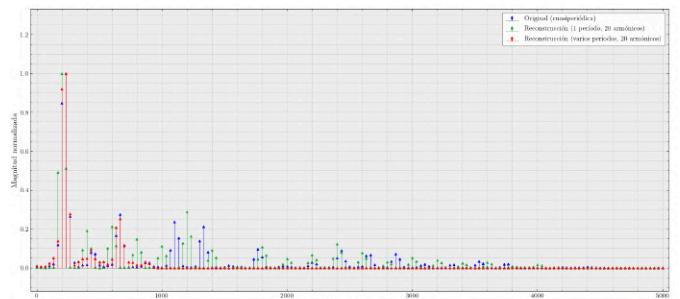


Figura 20. Señal original y reconstruida de un Saxo en La4 en la frecuencia

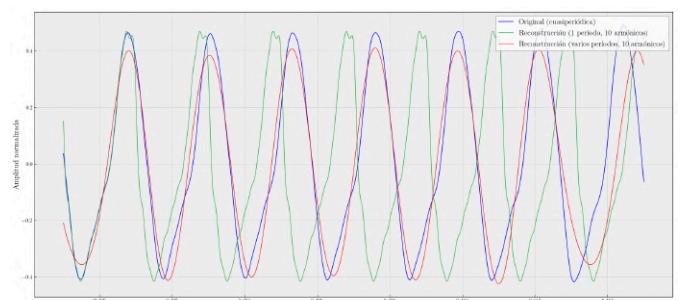


Figura 21. Señal original y reconstruida de un Piano en La4 en el tiempo

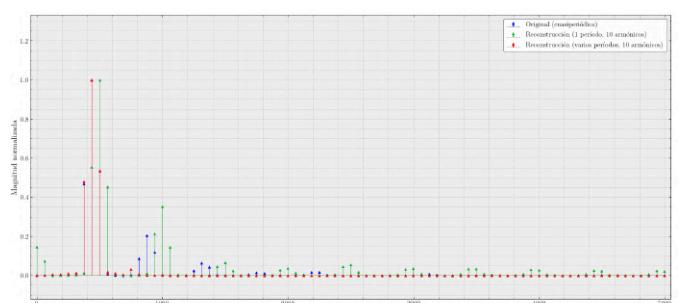


Figura 22. Señal original y reconstruida de un Piano en La4 en la frecuencia

III. CONCLUSIONES

- Pudimos observar que un espectrograma de una ventana ancha/banda angosta tiene mas resolución en frecuencia a costa de perder resolución en el tiempo, caso contrario con uno de ventana angosta/banda ancha el cual gana resolución en el tiempo pero la pierde en frecuencia.
- El efecto de *spectrum smeared* en un espectrograma es la dispersión de la energía en un rango de frecuencias, causada por cómo se corta la señal en la ventana para la FFT. Su magnitud depende del tamaño y tipo de ventana: ventanas rectangulares aumentan la dispersión debido al corte abrupto de la señal, mientras que ventanas suaves (Ventana de Hann) lo reducen debido a que lleva progresivamente a 0 la señal en los lóbulos extremos, lo que elimina discontinuidades en los bordes.
- Logramos obtener las notas musicales de una canción utilizando el espectrograma y la FFT de la canción. El análisis conjunto en tiempo y frecuencia nos mostró que ningún dominio por sí solo es suficiente para caracterizar completamente una señal musical sino que ambos son complementarios.
- La reconstrucción de señales cuasiperiódicas mediante FFT e IFFT muestra que usar un solo período genera una versión idealizada y limitada, mientras que considerar varios períodos captura mejor las variaciones naturales del sonido y el timbre, pero con menor precisión temporal.
- A pesar de la limitación debido al uso de uno o varios períodos para obtener los coeficientes de Fourier de la señal original cuasiperiodica, se confirma que el aumento de armónicos mejora la aproximación y que aquellos de alta frecuencia en un momento dejan de tener un aporte significativo. Esto último se evidenció al volver a oír las notas, las cuales preservaron en gran parte la estructura original.
- Los efectos musicales modifican la señal original de distintas formas, ya sea alterando la amplitud, frecuencia o combinando retardos. Cada efecto tiene propiedades de linealidad, invarianza y estabilidad específicas, lo que determina cómo afecta el timbre, la dinámica y la percepción espacial del sonido.

REFERENCIAS

- [1] C. T. Herbst, S. Hertegard, D. Zanger-Borch, y P.-Å. Lindstrand, «Freddie Mercury—acoustic analysis of speaking fundamental frequency, vibrato, and subharmonics», *Logopedics Phoniatrics Vocology*, vol. 42, n.º 1, pp. 29-38, 2017, doi: 10.3109/14015439.2016.1156737.
- [2] «Notas musicales y su correspondiente frecuencia». [En línea]. Disponible en: <https://mixbutton.com/music-tools/frequency-and-pitch/music-note-to-frequency-chart>
- [3] «Understanding FFTs and Windowing». [En línea]. Disponible en: <https://www.ni.com/en/shop/data-acquisition/measurement-fundamentals/analog-fundamentals/understanding-ffts-and-windowing.html>
- [4] «Window Functions in Spectrum Analyzers». [En línea]. Disponible en: <https://www.tek.com/en/blog/window-functions-spectrum-analyzers>
- [5] «Spectral Audio Signal Processing». [En línea]. Disponible en: https://ccrma.stanford.edu/~jos/sasp/Hann_Hanning_Raised_Cosine.html
- [6] S. Braun, «Windows», *Encyclopedia of Vibration*. Elsevier, Oxford, pp. 1587-1595, 2001. doi: <https://doi.org/10.1006/rwvb.2001.0052>.

IV. ANEXO

- Link al repositorio donde se encuentra el código desarrollado: [Repositorio \(click acá\)](#) o escaneando el QR

