

Situación Problema: Análisis de Audio usando Fourier

F1009: Análisis de métodos matemáticos para la física

November 12, 2024

Juan Pablo Guerrero Escudero ITESM, Querétaro A01706810@tec.mx	Romina Nájera Fuentes ITESM, Querétaro A01424411@tec.mx	Juan Braulio Olivares Rodríguez ITESM, Querétaro A01706880@tec.mx
--	---	---

1 Introducción

En los últimos años, se han hecho populares aplicaciones y programas de reconocimiento de canciones a partir de un fragmento de ellas, y aunque estos algoritmos son patentados, varios de ellos utilizan los principios del análisis espectral, utilizando Análisis de Fourier. Este análisis resulta una herramienta muy útil para esto, ya que, por medio de la Transformada de Fourier, provee frecuencias de la canción, y el uso de espectogramas permite sacar conclusiones de este sonido. Por medio de este concepto matemático, se puede analizar el fenómeno físico de las ondas de audio, que son ondas longitudinales en un medio, principalmente aire, el cuál el ser humano es capaz de escuchar a distintas frecuencias. En este reporte, se hará una investigación sobre los conceptos físicos y matemáticos necesarios, y se analizarán 10 canciones, para tratar de identificar el género de cada una, entre música instrumental y reggaetón, demostrando así una de las múltiples aplicaciones real de las matemáticas en la física.

2 Teoría

2.1 Conceptos físicos

2.1.1 Ondas de sonido

De acuerdo a Young y Freedman [1], El sonido se define como una onda longitudinal en un medio, principalmente aire, pero puede ser otros como otros gases, líquidos o sólidos. Las ondas de sonido más simple son ondas sinusoidales con una frecuencia, amplitud y longitud definida. [1]. El ser humano es capaz de escuchar ondas en el rango de 20 a 20,000 Hz, con frecuencias por encima del rango (ultrasónicas) o debajo del rango (infrasónicas) fuera del rango de escucha humano. De acuerdo a Hwaitat [2], las ondas de sonido son perturbaciones propagadas por un medio el cuál no se ve afectado, y éstas ondas pueden ser ya sea longitudinales, o transversales. Para una onda de tipo longitudinal, el medio vibra

en ángulos rectos al movimiento de la onda, y en el caso de ondas longitudinales, el medio vibra en la misma dirección que el movimiento. En la Figura 1 se observa gráficamente lo discutido.

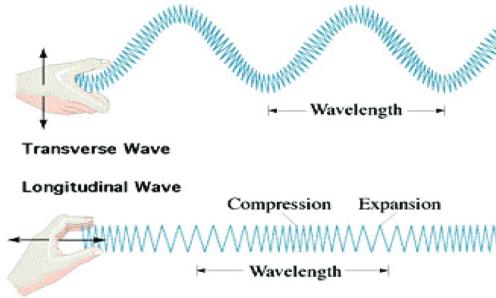


Figura 1: Ondas longitudinales y transversales

Los parámetros de cualquier onda constituyen la amplitud, la frecuencia y la longitud, mencionados anteriormente. La amplitud puede ser definida como la "altura" de la onda, la frecuencia se define como los ciclos por segundo, y la longitud se define como la distancia entre un pico de onda y otro. Para una vista gráfica, vea la Figura 2. Generalmente, sucede que cuando dos partículas están en movimiento en el mismo medio, ocurre interferencia. Ésto significa que las amplitudes de onda son sumadas algebraicamente, y se siguen moviendo por el medio sin distracciones. En el mundo real, la interferencia de ondas crea patrones complejo, y puede ser muy difícil de analizar.

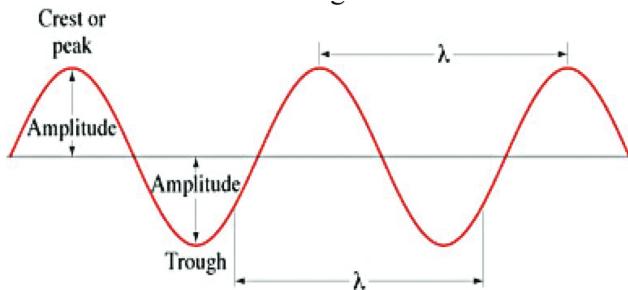


Figura 2: Parámetros de las ondas de sonido

2.1.2 Frecuencias de audio/sonido

Como se definió anteriormente, la frecuencia de una onda es, de acuerdo a [1], como el número de repeticiones de una función periódica durante una unidad de variación en la variable independiente. En otras palabras, es el número de ocurrencias de un evento repetitivo por unidad de tiempo. Matemáticamente, se dice que en medios no dispersivos (medios donde la velocidad de la onda es independiente de la frecuencia), la frecuencia tiene una relación inversa con la longitud de onda, en la forma de la ecuación 1, donde λ es la longitud de la onda, y v es la velocidad de la onda.

$$f = \frac{v}{\lambda} \quad (1)$$

Analizando la ecuación 1, observamos que ondas de longitud más corta (λ), tienen mayores variaciones en la frecuencia debido a que los máximos y mínimos están más cerca uno del otro, y viceversa. Una vista gráfica se observa en la Figura 3.

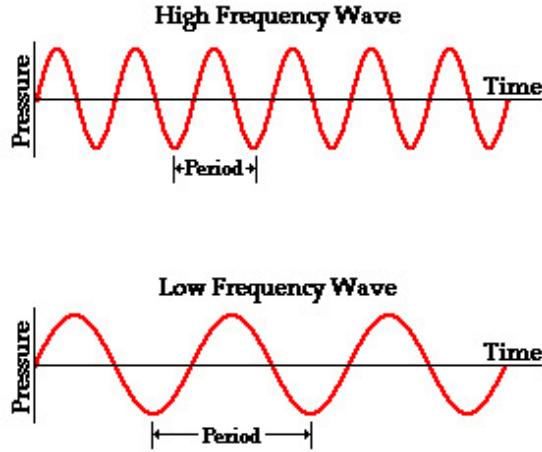


Figura 3: Diferentes frecuencias de onda

Además, la frecuencia se mide en Hz o Hertz, equivalente a un evento de repetición por segundo. De acuerdo a [1], la frecuencia de una onda de sonido es el factor primario al determinar el tono de un sonido. Frecuencias más altas emiten sonidos con mayor tono que frecuencias más bajas. En conjunto, cuando se juntan diferentes frecuencias al mismo tiempo, se crean patrones más complejos que ondas simples sinusoidales, debido a que físicamente, las variaciones en presión del medio que se generan son más complejas.

2.1.3 Sonidos armónicos

De acuerdo a [1], puede suceder que dos tonos producidos por diferentes instrumentos tengan la misma frecuencia pero suenan diferente, y ésto es debido a la diferencia en contenido armónico, que está definido como la colección de diferentes frecuencias que componen un sonido complejo formado por muchas frecuencias fundamentales. Además, de acuerdo a [1], otro factor que determina los sonidos armónicos de un sonido es el comportamiento al inicio (attack) y al final (decay) de cada tono. Cada instrumento posee una diferente dinámica de sonidos armónicos, lo cuál entrega ondas en una mezcla diferente de frecuencias, que da su sonido característico.

Para dar algunos ejemplos [3], la nota de afinación en una orquesta sinfónica es A4, que tiene una frecuencia de 440Hz. Frecuencias más complejas se generan cuando se utiliza el concepto de octavas en música, que de acuerdo a [4], es una serie de 8 notas musicales que ocupan el intervalo entre 2 notas, una teniendo el doble o la mitad de la frecuencia que la otra. Es decir, un salto de una octava corresponde a duplicar la frecuencia de la onda de sonido.

2.1.4 Beats

En la física, sucede que cuando se tienen dos ondas con igual amplitud pero ligeramente diferentes frecuencias, si ambas ondas van hacia la misma dirección, pasa que en ciertos momentos, las dos ondas están en la misma fase, es decir, sus máximos coinciden y sus amplitudes se suman. Sin embargo, debido a que las ondas están en ligeramente distintas frecuencias, llega a haber momentos en los que estas ondas están exactamente fuera de fase, y se cancelan una con otra. La variación de amplitud causa variaciones en el volumen, llamadas *beats*, y la frecuencia con la que éstos varían se le llama frecuencia de beat. En la Figura 4 se muestra éste proceso de manera gráfica.

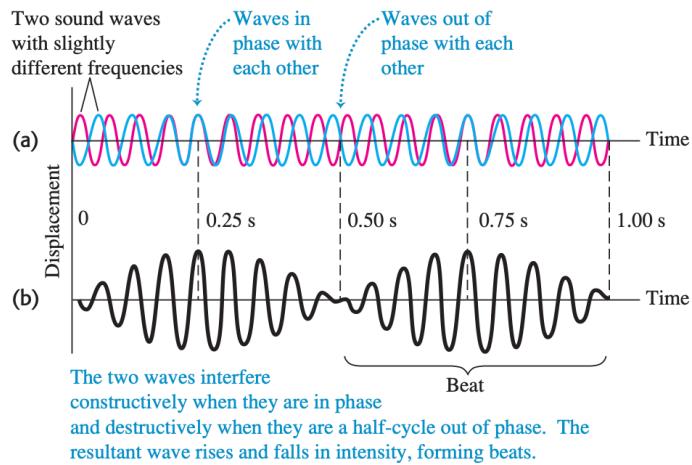


Figura 4: Gráfica del proceso de beats

Los beats entre dos tonos, de acuerdo a [1], pueden ser escuchados hasta una frecuencia de beat de 6 o 7Hz. En la práctica, una técnica importante es escuchar los beats para afinar los instrumentos musicales. Además, de acuerdo a [1], cuando las diferencias de frecuencia son mayores a 7Hz, se dejan de escuchar beats individuales, y las frecuencias se mezclan en una de consonancia o disonancia, dependiendo de la naturaleza de las frecuencias.

2.2 Análisis de las canciones

Para la clasificación de las canciones entre los géneros de música instrumental o reggaetón, realizaremos un análisis espectral, el cual busca descomponer una serie de tiempo en las ondas senoidales que la conforman [5]. Este análisis permitirá obtener las diferentes frecuencias que conforman al pedazo de canción a analizar, y poder sacar conclusiones sobre el género de la canción.

Para ello, utilizaremos la transformada de Fourier, utilizada comúnmente en el campo científico, como en la acústica y el procesamiento de señales. Esta herramienta transforma el dominio de una señal, pasando del tiempo a la frecuencia, sin alterar su contenido [6]. Al perderse la noción del tiempo, analizaremos los rangos de frecuencias en los que se encuen-

tran magnitudes más grandes, para así identificar si la canción presentada es instrumental o reggaetón.

2.2.1 Transformada de Fourier

Por definición, la transformada de Fourier de una función $f(x)$ es dada por la ecuación 2

$$\phi_f(\alpha) = \int_{-\infty}^{\infty} e^{i\alpha x} f(x) dx \quad (2)$$

Esta transformada fue desarrollada por Jean-Baptiste Joseph Fourier en el siglo XIX, quien inició proponiendo que cualquier función arbitraria de una variable podía ser expresada como una combinación lineal de funciones de senos y cosenos, que son las series de Fourier [7].

A través de esas series, logró sintetizar la transformada, la cual es distinguida por:

- Determinar qué frecuencias están presentes en una señal.
- Transformar una señal del dominio temporal al dominio de frecuencia y viceversa.

Así como las series de Fourier descomponen una función en senos y cosenos, la transformada descompone una señal en sus frecuencias, y aquellas que tengan una mayor amplitud, se verán representadas como picos más altos, así como se puede observar en la figura 5.

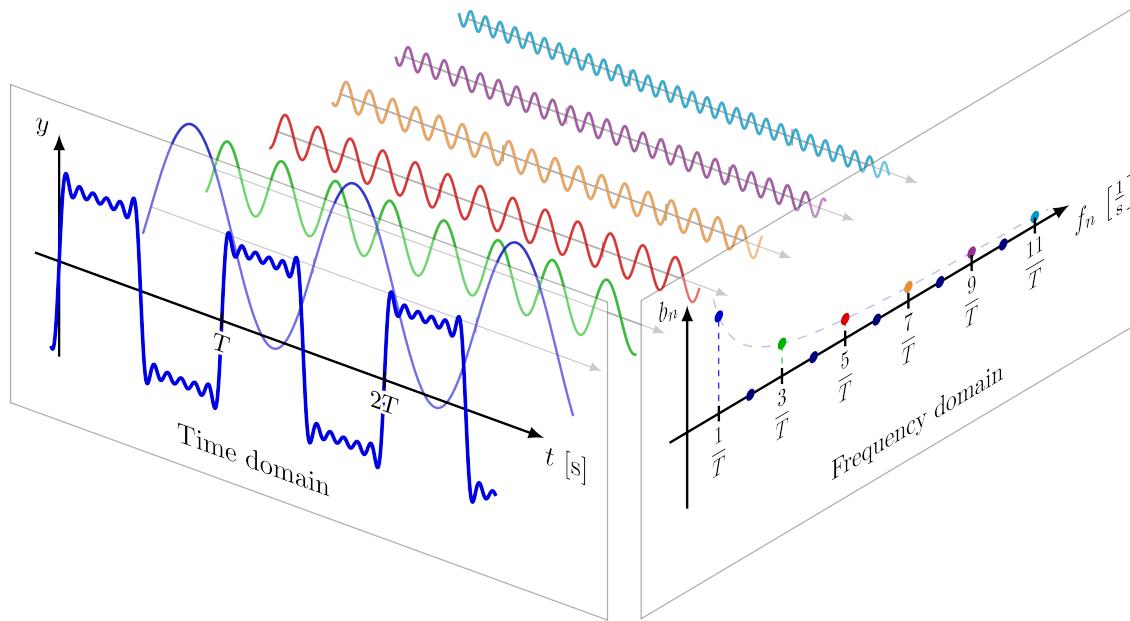


Figura 5: Representación visual de la transformada de Fourier [7]

2.2.2 Espectrogramas

La transformada de Fourier nos provee con las frecuencias de la canción, pero la forma en la que estas pueden ser analizadas puede ser muy variada. Una de las formas que existen para analizar estas frecuencias es a través de los espectrogramas, los cuales son una representación

visual de la cual se pueden sacar distintas conclusiones del sonido que se tiene. Tratar la representación visual de frecuencias como una imagen con textura permite reconocer la suavidad, la regularidad, el contraste, entre otros [8].

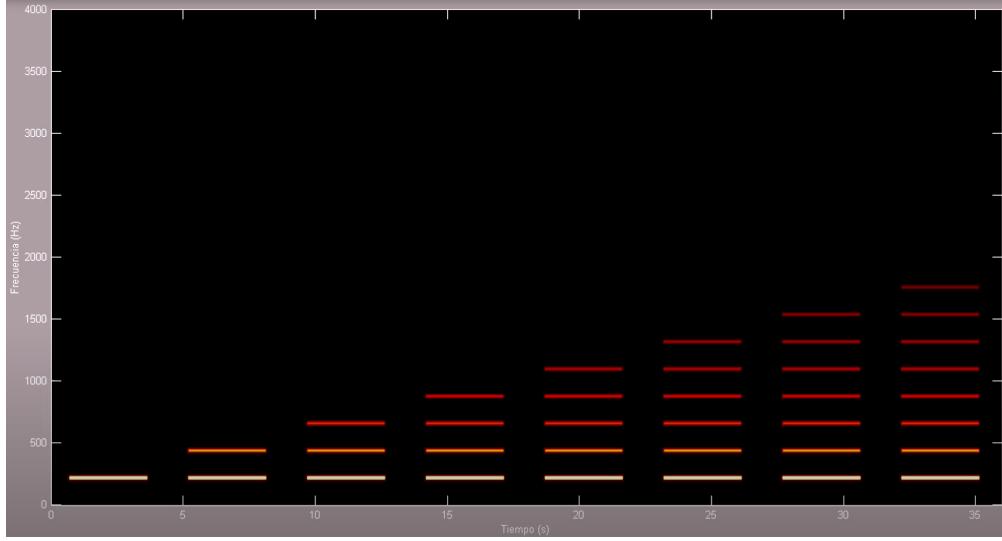


Figura 6: Espectrograma de sonidos armónicos estables [9]

La figura 6 representa un espectrograma con los componentes de la serie armónica. Se observan líneas bien definidas, lo cual se asemejaría a lo que se busca en una canción instrumental.

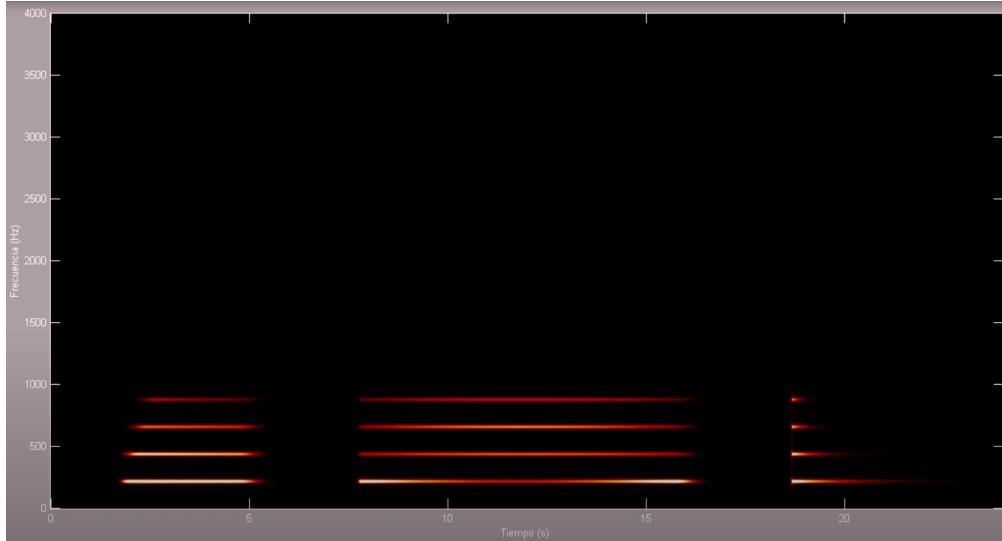


Figura 7: Espectrograma de tres sonidos armónicos formados por componentes cuya amplitud evoluciona de diferentes formas [10]

En la figura 7 se tienen también sonidos armónicos, solo que sus amplitudes cambian. A pesar de ello, se observan aún líneas bien definidas, que también se podría esperar de las canciones instrumentales.

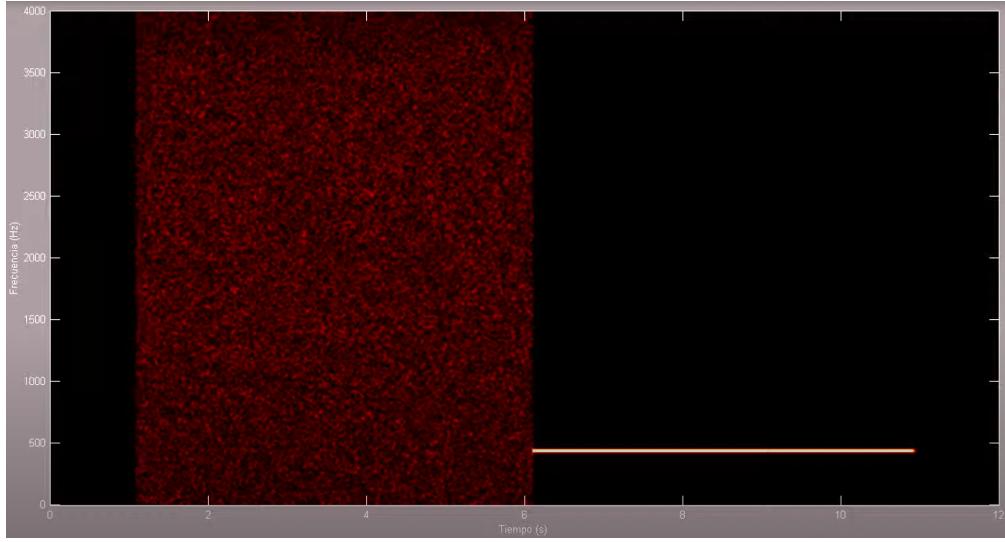


Figura 8: Espectrograma de ruido blanco y sonido simple [11]

En la figura 8 se tiene ahora una comparación entre el ruido blanco, el cual contiene a todas las frecuencias, y el sonido simple. Se puede notar que cuando existe una gran combinación de frecuencias, el espectrograma se encuentra muy saturado, mientras que cuando tenemos sonidos más puros, en el espectrograma se tiene una línea definida.

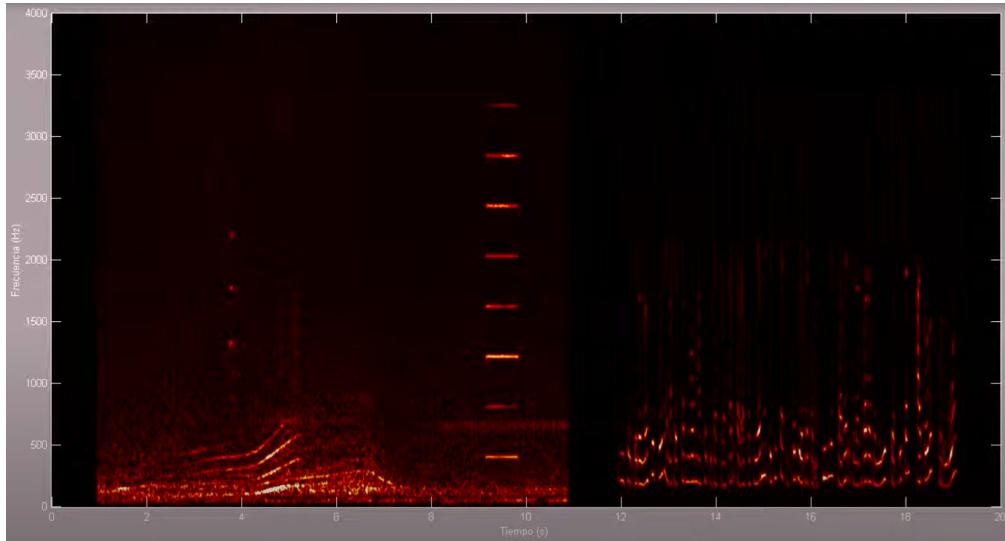


Figura 9: Espectrograma de ruido de tráfico y de habla [12]

En la figura 9, se tiene un espectrograma que representa al ruido del tráfico en la primera parte, y a una voz de un programa de radio en la segunda mitad. Así como el ruido blanco, al tener una gran combinación de sonidos, la parte del espectrograma con el ruido de tráfico está saturado en frecuencias bajas, y en la parte derecha, hay varias líneas fragmentadas y que siguen distintas frecuencias. Aquí ya no se llegan a ver líneas tan bien definidas como en las figuras 6 y 7, mas que a la mitad, con el sonido de un claxon de auto, que se podría decir que es un sonido más definido que el tráfico o la voz.

Utilizando estos ejemplos, se puede tener una intuición de qué esperar en el espectrograma de cada género musical. Por un lado, con la música instrumental, al no tener voces, y tener sonidos mezclados entre beats y los armónicos producidos por los mismos instrumentos. Como no existe una gran mezcla de sonidos, el espectrograma tendería a verse con líneas horizontales mejor definidas, como en las figuras 6 y 7. Por otro lado, en el género musical del reggaetón, se tiene una gran combinación de sonidos, como el beat y los armónicos, además de tener una voz que se podría asemejar a la segunda parte de la figura 9. Entonces el reggaetón tendría una mayor combinación de ruidos, y su espectrograma tendría mucha textura.

2.2.3 Espectro de frecuencias

A pesar de que la voz sea una forma para identificar entre el reggaetón y la música instrumental, no se tiene la seguridad de que esta sea incluida en los fragmentos de canciones, y las formas de identificar la canción con el espectrograma sigue siendo algo ambigua. Es por ello que, a la par, se pueden identificar ciertos rangos de frecuencias que caracterizan a cada tipo de música.

El oído humano es capaz de escuchar frecuencias entre 20 y 20000 Hz, dentro de las cuales se suelen hacer 3 categorías: bajos, medios y altos. A continuación se presentan las subcategorías de estas, explicadas por Gleeson [13]:

- Sub-bass (de los bajos): 20-60 Hz, es una frecuencia que suele sentirse más que escucharse.
- Mid-bass (o bass, de los bajos): 60-250 Hz, frecuencia baja que empieza a tener tonos más reconocibles.
- Low mids (de los medios): 250-500 Hz, suele representar una transición de las frecuencias bajas al rango medio.
- Center mids (o midrange, de los medios): 500-2000 Hz, en este rango de frecuencias se suelen encontrar los armónicos y fundamentales de las partes más importantes de las canciones.
- Upper mids (de los medios): 2000-4000 Hz, un rango de frecuencia que suele consistir en armónicos, detalles del timbre y sonidos transitorios.
- Presence (de los altos): 4000-6000 Hz, que consiste principalmente por frecuencias armónicas.
- Brilliance (de los altos): 6000-20000 Hz, que suele indicar el timbre general del sonido, mas no su tonalidad.

Estilos de música como el reggae, dub y dancehall Jamaiquinos, el hip-hop, la cumbia, el reggaetón, el Miami bass, entre otros, utilizan una gran cantidad de sonidos de baja frecuencia. Las bandas de frecuencias por debajo de los 100 Hz, en los límites de lo audible, alteran la realidad física al vibrar [14].

De acuerdo con García[15], el reggaetón es uno de los géneros orientados al bass, cuyos beats utilizan frecuencias alrededor de los 20 Hz y producen una experiencia energizante, opresiva, impulsiva, desorientadora, entre otros. Es a través de estas frecuencias se involucran los sentidos hápticos al escuchar y bailar este tipo de música.

Tanto García[15] como Fink[14] coinciden en que el reggaetón utiliza muchas frecuencias bajas para generar el efecto querido en el oyente. La música de reggaetón se presta mucho para bailar, mientras que la música instrumental no presenta este tipo de beats que hacen vibrar al cuerpo, no generando así el mismo efecto que el reggaetón en el oyente.

3 Resultados

Resumen de clasificación

Canción	Clasificación
1	Instrumental
2	Reggaetón
3	Instrumental
4	Instrumental
5	Reggaetón
6	Reggaetón
7	Instrumental
8	Instrumental
9	Reggaetón
10	Instrumental

Canción 1

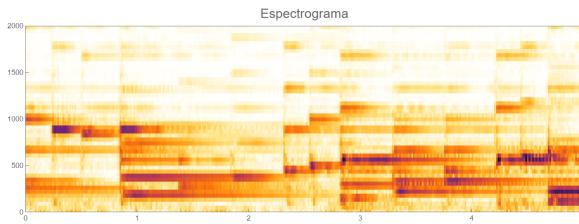
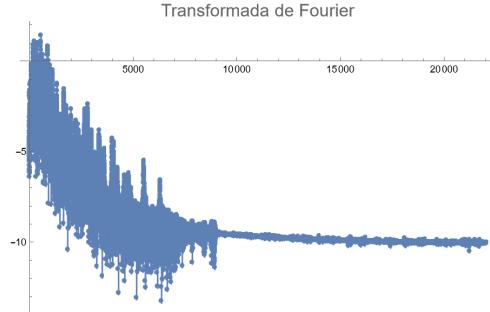


Figura 11: Espectrograma de la canción 1

Figura 10: Logaritmo de la transformada de Fourier de la canción 1

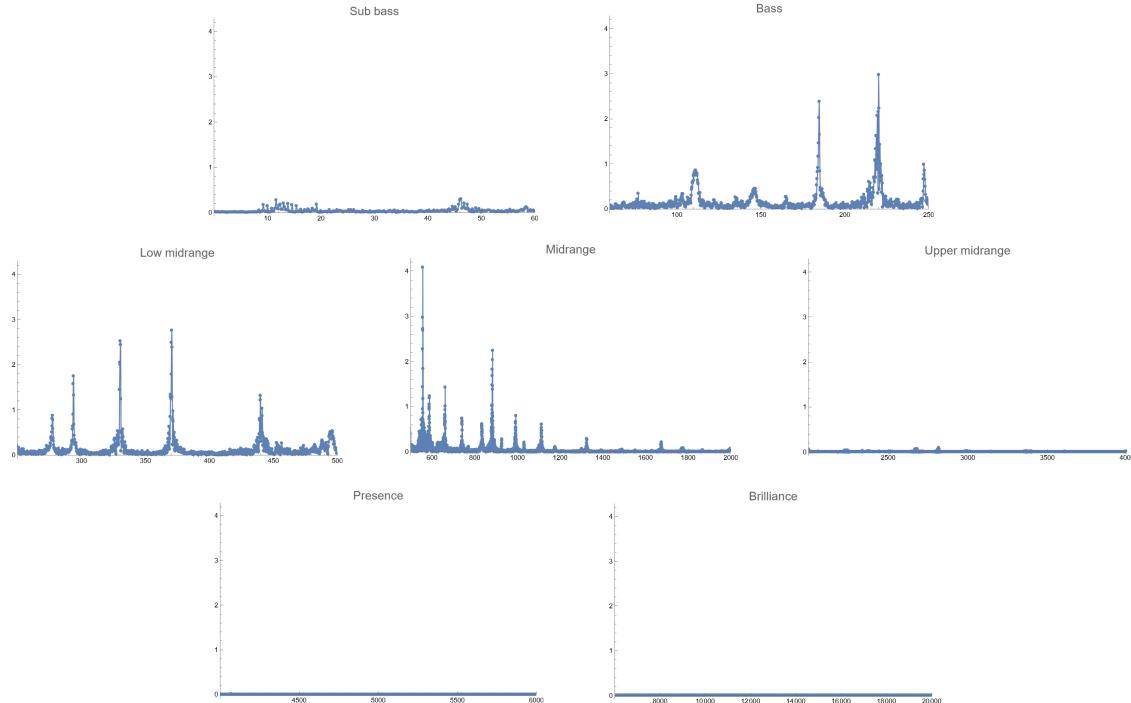


Figura 12: Espectro de frecuencias de la canción 1

Clasificación: Instrumental

Justificación: Observando el espectro de frecuencias de la canción, se ve una mayor predominancia en frecuencias que se encuentran en bass, low midrange y en midrange. Al no incluir un subbass tan marcado, clasificamos a esta canción como instrumental. Asimismo, observamos que el espectrograma no parece presentar mucho ruido, y se parece más a los espectrogramas de ejemplo de sonidos armónicos (como en la Figura 7), confirmando nuestra clasificación.

Canción 2

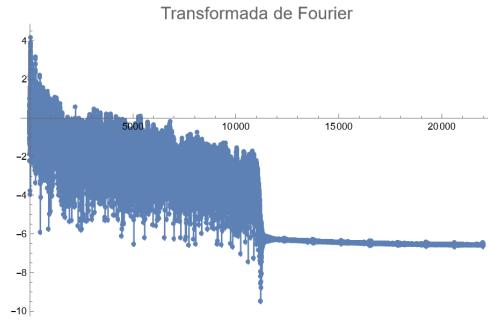


Figura 13: Logaritmo de la transformada de Fourier de la canción 2

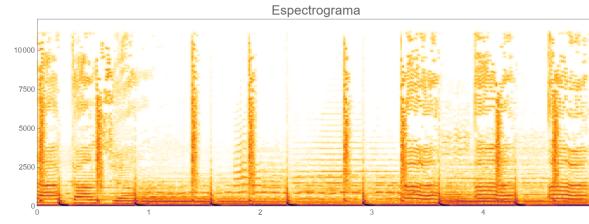


Figura 14: Espectrograma de la canción 2

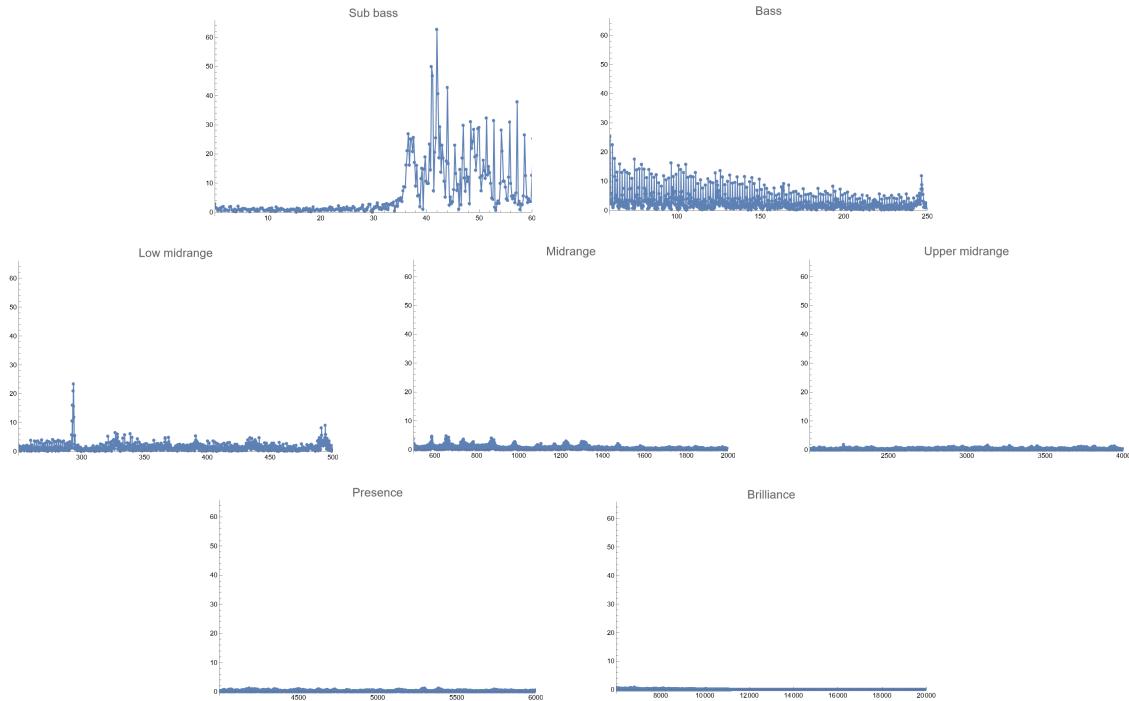


Figura 15: Espectro de frecuencias de la canción 2

Clasificación: Reggaetón

Justificación: A diferencia de la canción anterior, en esta observamos que las frecuencias que dominan en esta canción son subbass y bass. Es por ello que clasificamos a esta canción como reggaetón, por su baja frecuencia. Asimismo, observando el espectrograma, existe mucho ruido que no parece ser tan armónico, que sería más probable a ser reggaetón que música instrumental.

Canción 3

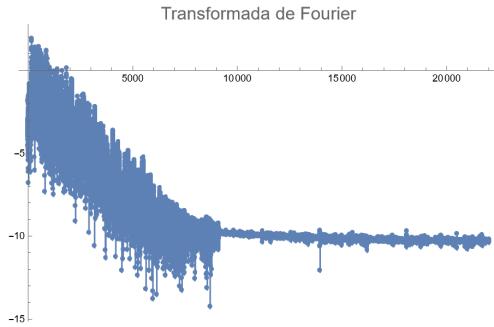


Figura 16: Logaritmo de la transformada de Fourier de la canción 3

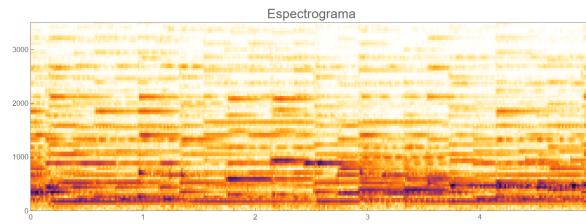


Figura 17: Espectrograma de la canción 3

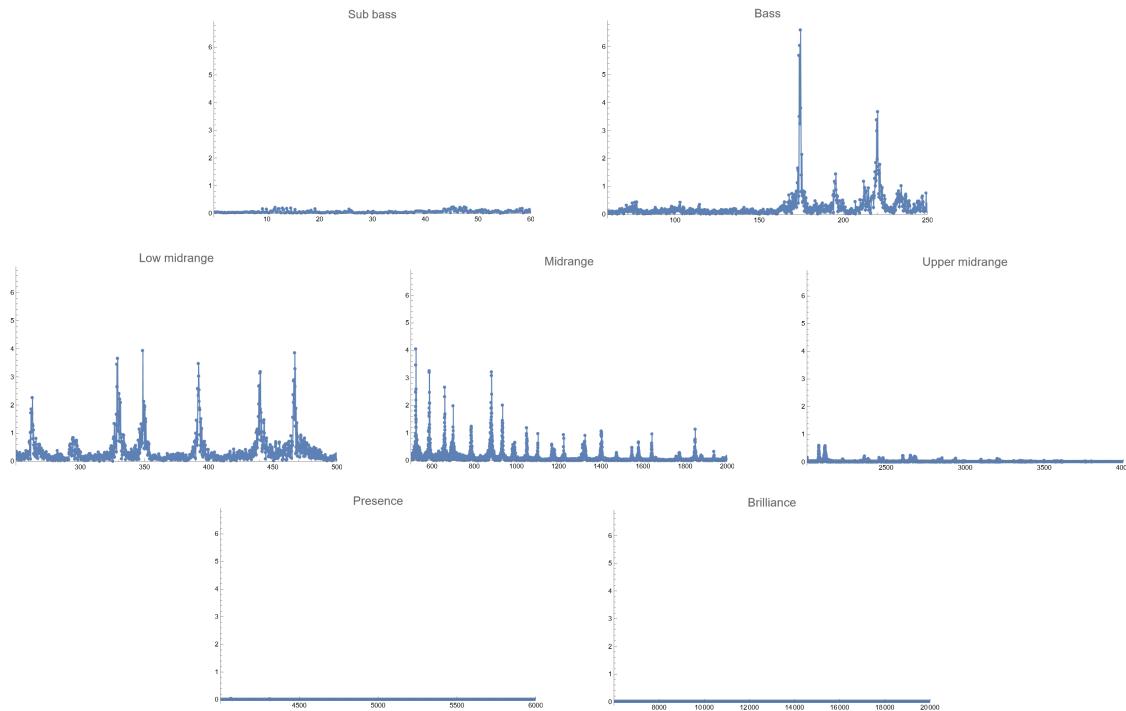


Figura 18: Espectro de frecuencias de la canción 3

Clasificación: Instrumental

Justificación: Esta canción presenta más frecuencias entre bass y midrange dentro de su espectro de frecuencias. Al no presentar tanto subbass, la clasificamos como instrumental. Observando el espectrograma, también observamos que las frecuencias se ven más armónicas.

Canción 4

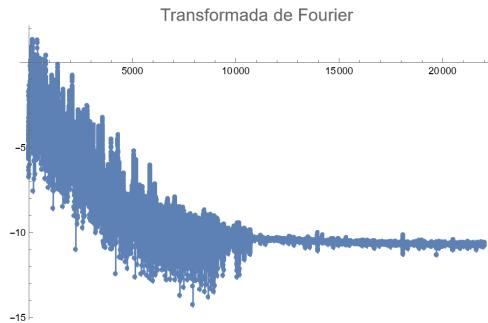


Figura 19: Logaritmo de la transformada de Fourier de la canción 4

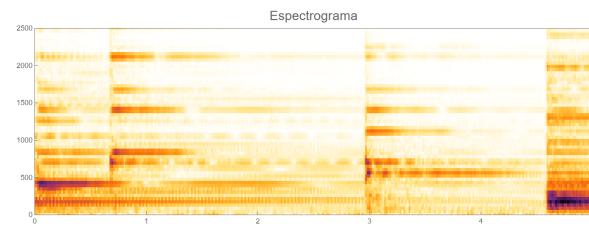


Figura 20: Espectrograma de la canción 4

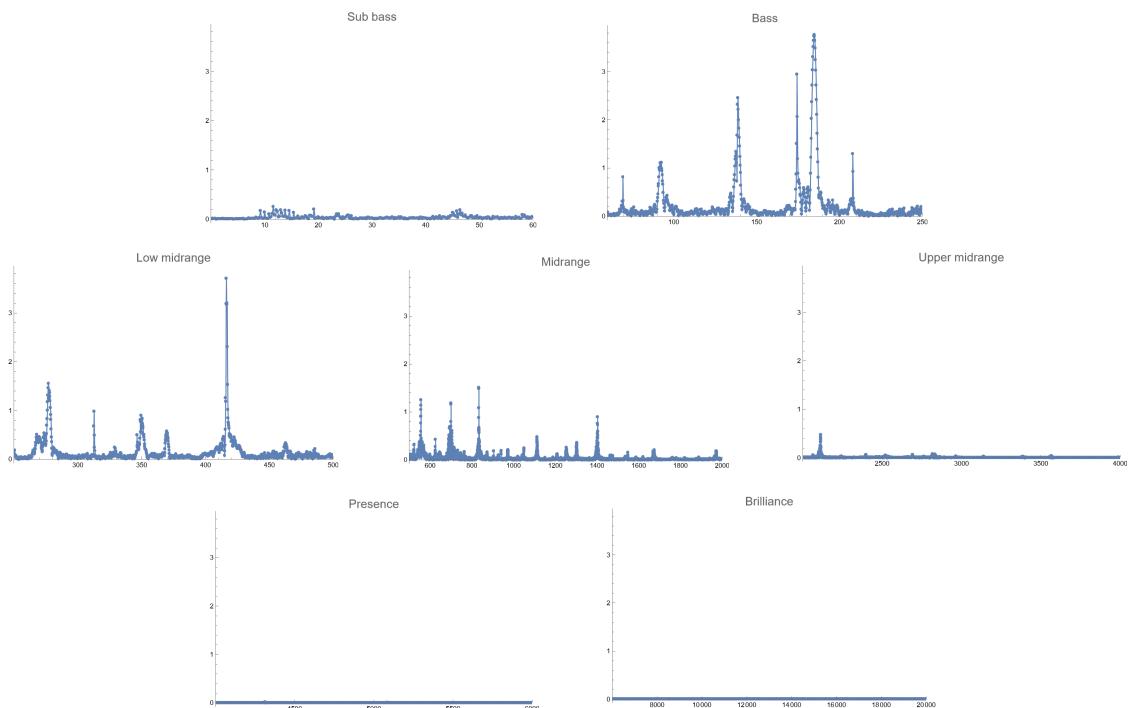


Figura 21: Espectro de frecuencias de la canción 4

Clasificación: Instrumental

Justificación: El espectro de frecuencias de esta canción no muestra subbass, por lo que la clasificamos como instrumental. Igualmente, el espectrograma tiene sonidos que se ven armónicos, así como en la música instrumental.

Canción 5

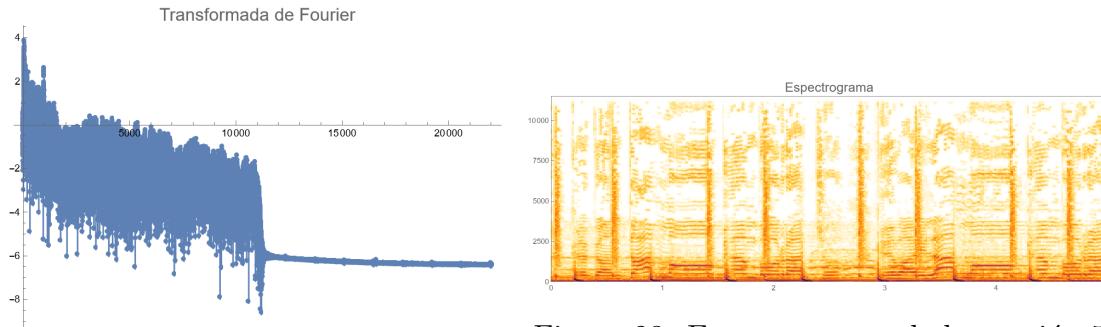


Figura 22: Logaritmo de la transformada de Fourier de la canción 5

Figura 23: Espectrograma de la canción 5

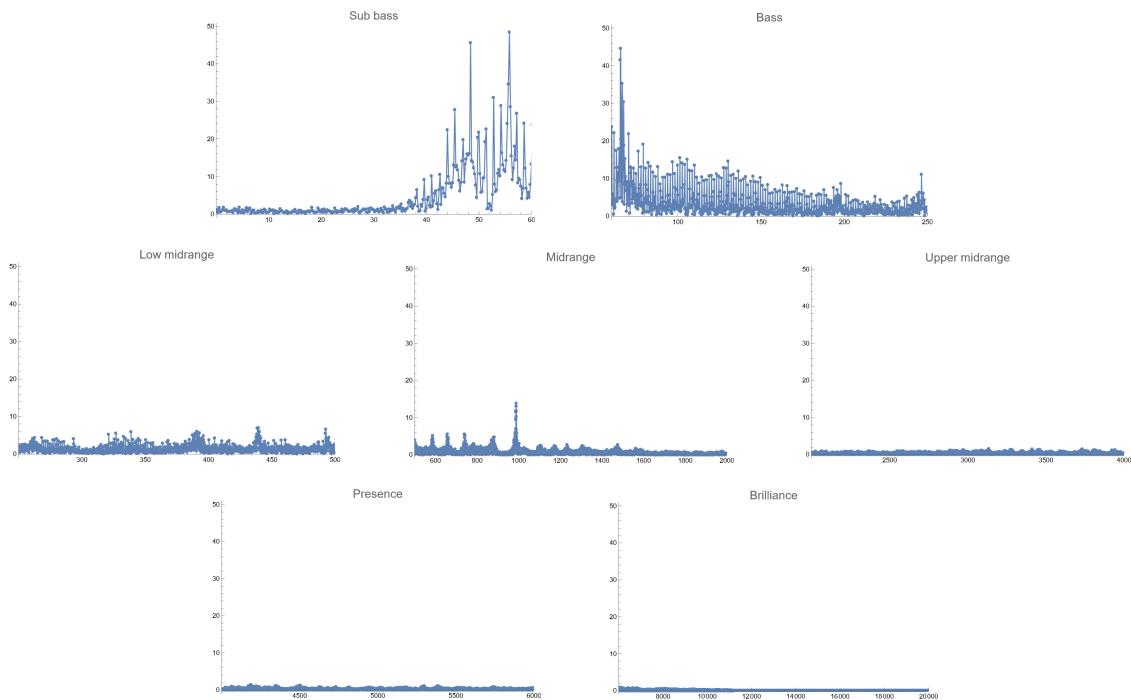


Figura 24: Espectro de frecuencias de la canción 5

Clasificación: Reggaetón

Justificación: En el espectro de frecuencias, se observa que la canción presenta mucho subbass y bass, indicando que es reggaetón, y el espectrograma nos lo confirma, al no verse un patrón armónico tan marcado como en la música instrumental.

Canción 6

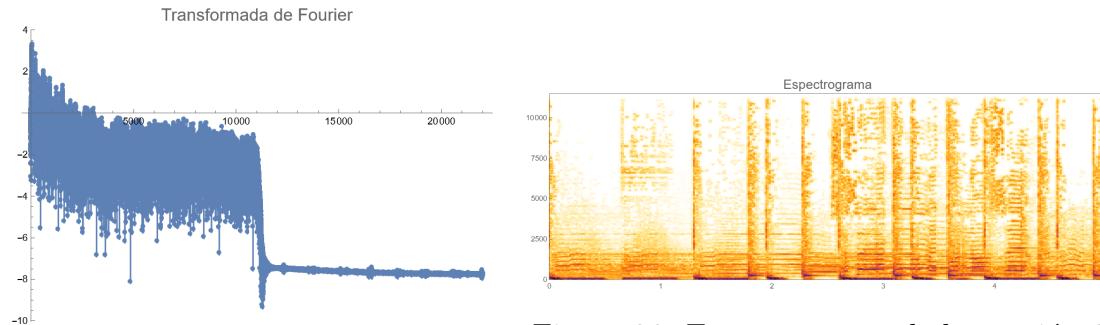


Figura 25: Logaritmo de la transformada de Fourier de la canción 6

Figura 26: Espectrograma de la canción 6

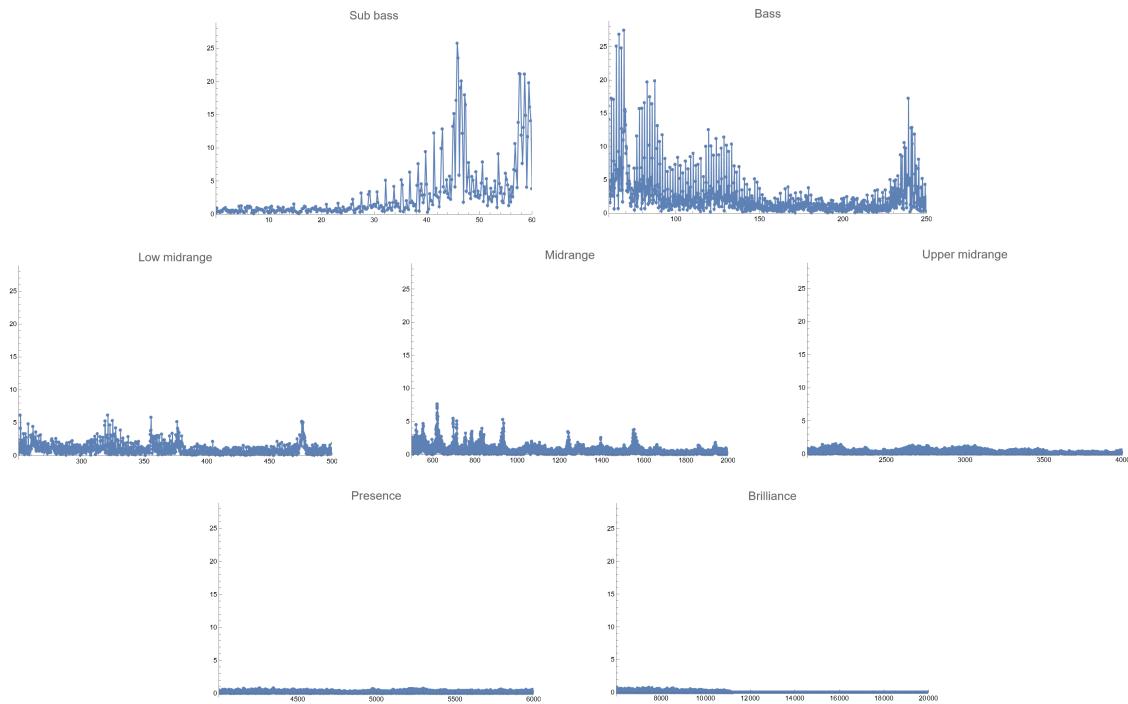


Figura 27: Espectro de frecuencias de la canción 6

Clasificación: Reggaetón

Justificación: Nuevamente hay una gran predominancia en las frecuencias que se encuentran en el subbass y el bass de la canción, significando así que la canción es de reggaetón. Asimismo, el espectrograma se asemeja a las demás canciones clasificadas como reggaetón, debido a su falta de sonidos armónicos preponderantes.

Canción 7

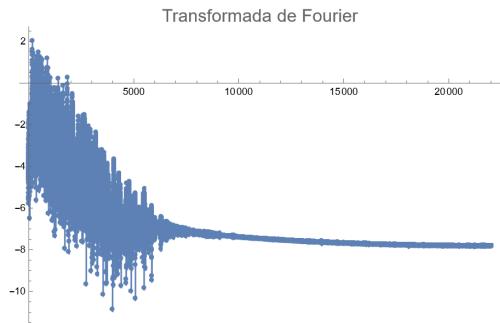


Figura 28: Logaritmo de la transformada de Fourier de la canción 7

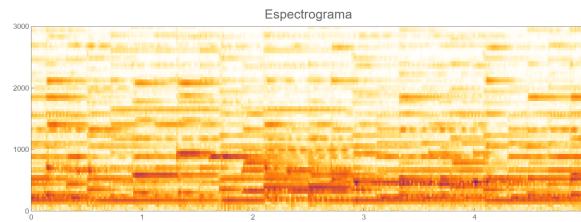


Figura 29: Espectrograma de la canción 7

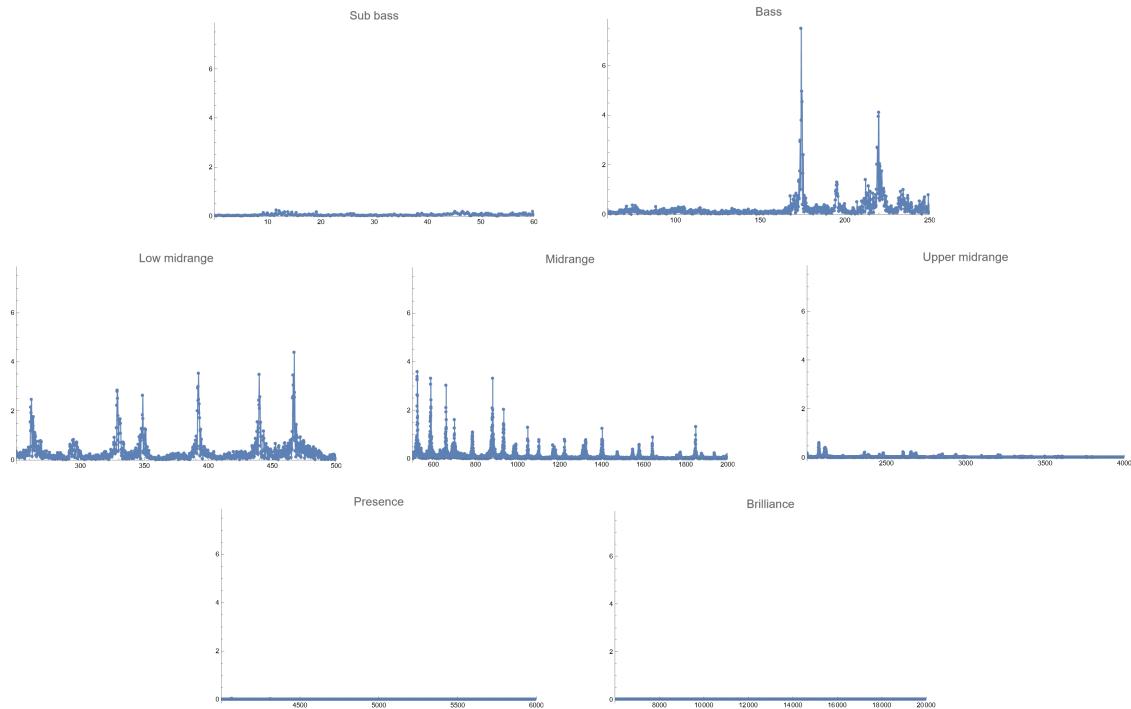


Figura 30: Espectro de frecuencias de la canción 7

Clasificación: Instrumental

Justificación: La canción no presenta una gran predominancia de frecuencias dentro del subbass, y su espectrograma muestra muchos sonidos armónicos, por lo que la clasificamos como instrumental.

Canción 8

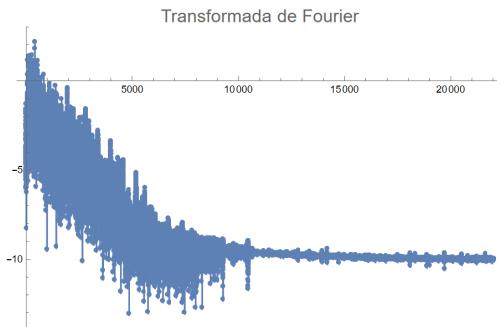


Figura 31: Logaritmo de la transformada de Fourier de la canción 8

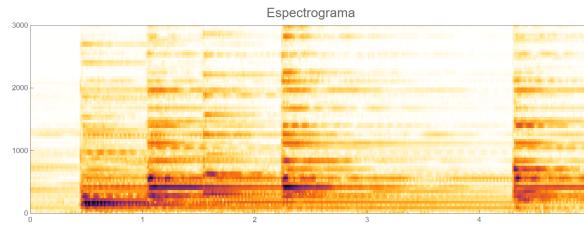


Figura 32: Espectrograma de la canción 8

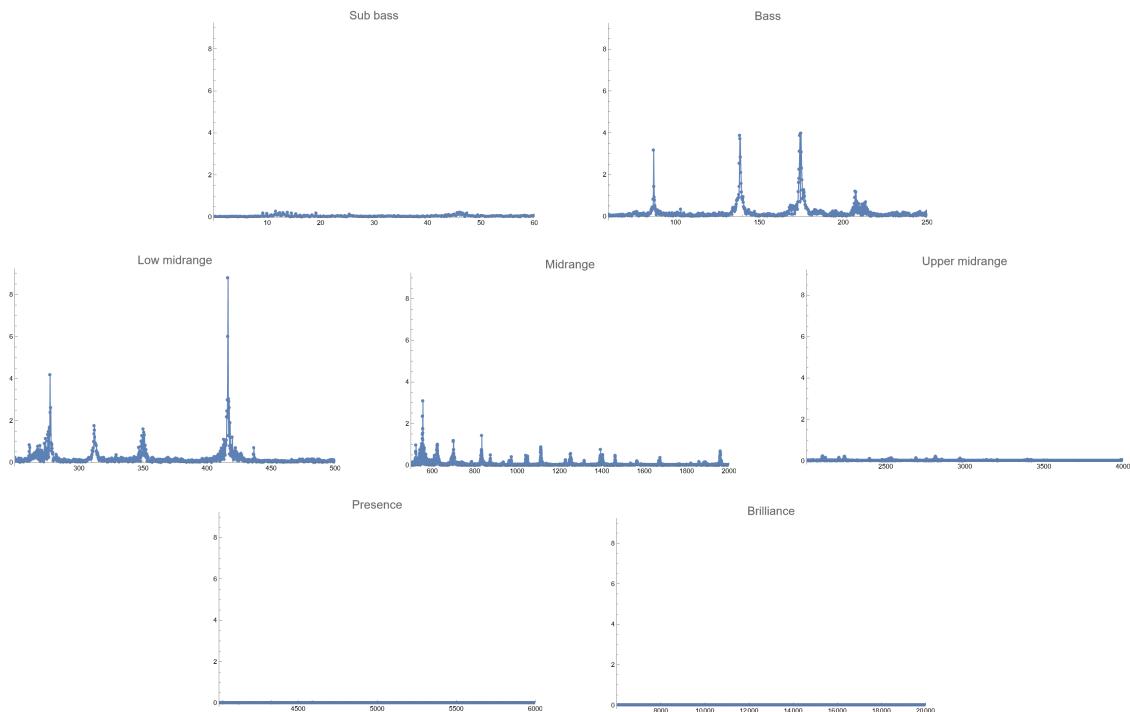


Figura 33: Espectro de frecuencias de la canción 8

Clasificación: Instrumental

Justificación: Nuevamente, observamos poco subbass en el espectro de frecuencias, y sonidos armónicos en el espectrograma, que significan una canción instrumental.

Canción 9

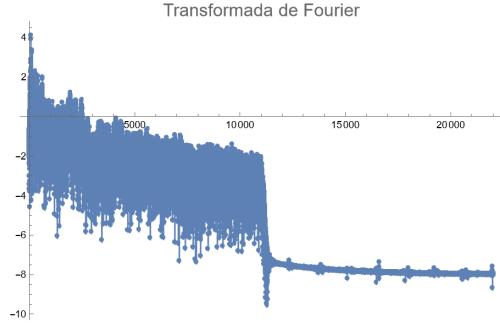


Figura 34: Logaritmo de la transformada de Fourier de la canción 9

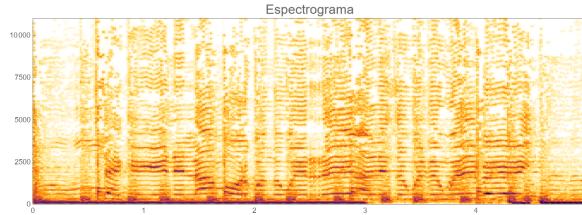


Figura 35: Espectrograma de la canción 9

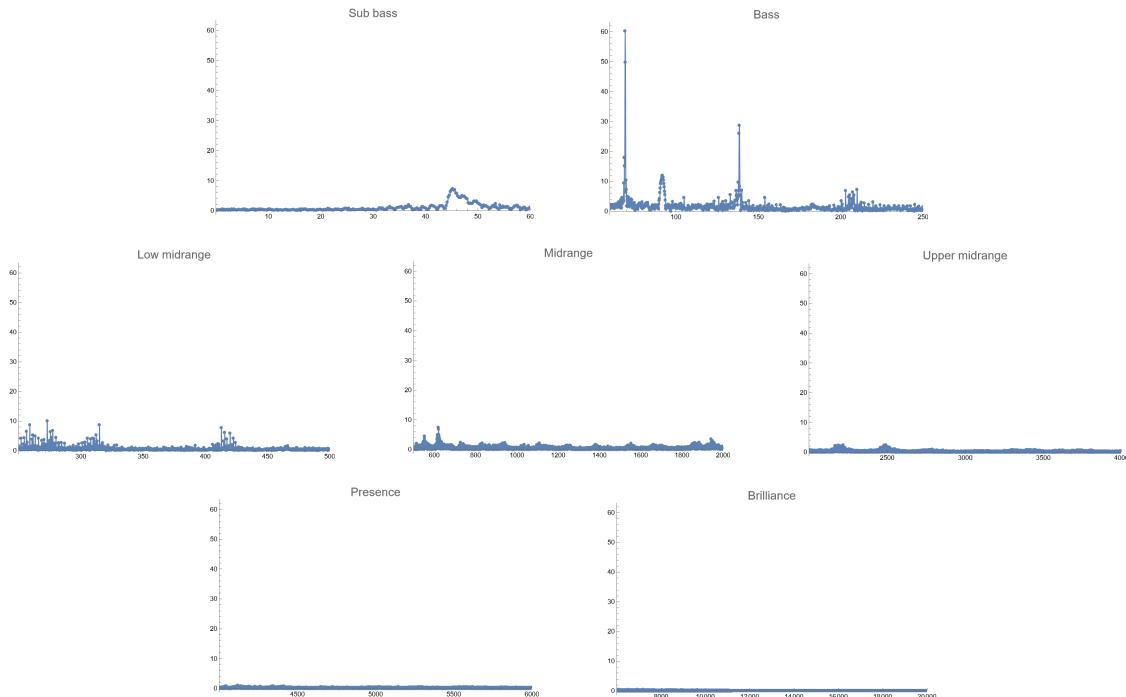


Figura 36: Espectro de frecuencias de la canción 9

Clasificación: Reggaetón

Justificación: Esta canción no parece presentar un subbass tan dominante, aunque no se observa tan insignificante como en otras canciones instrumentales. Observando más a detalle, las magnitudes de las frecuencias en el subbass de otras canciones instrumentales han sido menores a 1, y estas llegan a más de 7. Asimismo, observando el espectrograma de esta canción, no se presentan sonidos armónicos. Con todo ello, concluimos que esta canción es de reggaetón.

Canción 10

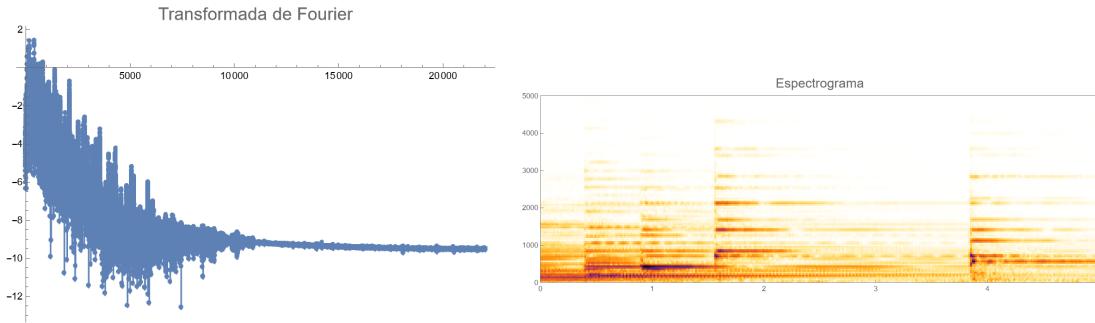


Figura 37: Logaritmo de la transformada de Fourier de la canción 10

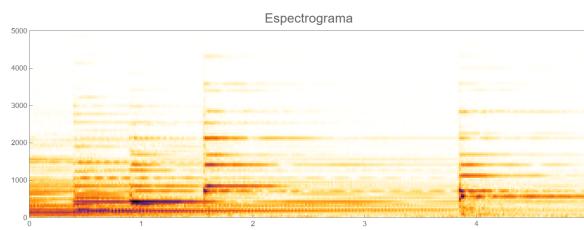


Figura 38: Espectrograma de la canción 10

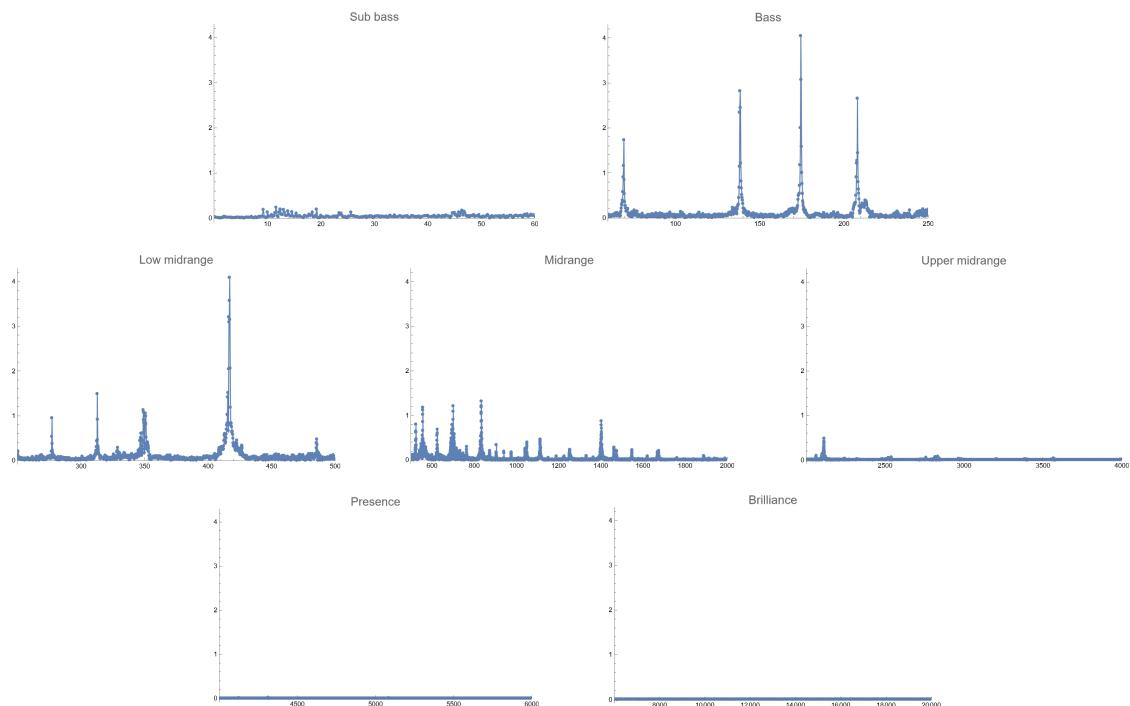


Figura 39: Espectro de frecuencias de la canción 10

Clasificación: Instrumental

Justificación: El espectrograma muestra sonidos armónicos, y el espectro de frecuencias no presenta magnitudes altas de frecuencias, por lo que clasificamos esta canción como instrumental.

4 Conclusiones

References

- [1] H. D. Young and Roger A. Freedman, *University Physics with Modern Physics*, Addison-Wesley, San Francisco, 2012.
- [2] Al Hwaitat et Al, Journal of Experimental & Theoretical Artificial Intelligence, 2022, 34, 749-780.
- [3] N. Lenssen, *PhD Thesis: Applications of Fourier Analysis to Audio Signal Processing: An Investigation of Chord Detection Algorithms*, Claremont University, 2013.
- [4] Britannica, <https://www.britannica.com/art/octave-music>, (accesado 28 Octubre 2024).
- [5] A. Montenegro, *CORE*, 2009, <https://core.ac.uk/download/pdf/6448967.pdf>, (accesado Octubre 28, 2024).
- [6] J. Bernal, P. Gómez y J. Bobadilla, *Estudios de fonética experimental*, 1999, **10**, 75-105.
- [7] L. O’Gorman, *DIBS Methods Meetings*, 2023, <https://dibsmethodsmeetings.github.io/fourier-transforms/>.
- [8] Y. M. G. Costa, L. S. Oliveira, A. L. Koerich y F. Gouyon, *IEEE*, 2011, **18**, 1-4.
- [9] Luis Colomer Blasco, Capítulo 10. Figura 10. Espectrograma serie armónica, <https://www.youtube.com/watch?v=RzitKHMUoeg>, (accesado Octubre 28, 2024).
- [10] Luis Colomer Blasco, Capítulo 10. Figura 11. Espectrograma de envolventes de amplitud, <https://www.youtube.com/watch?v=idEvt5HuQTc>, (accesado Octubre 28, 2024).
- [11] Luis Colomer Blasco, Capítulo 10. Figura 13. Espectrograma de ruido blanco y sonido simple, <https://www.youtube.com/watch?v=ZoffBnGMctI>, (accesado Octubre 28, 2024).
- [12] Luis Colomer Blasco, Capítulo 10. Figura 14. Espectrograma de tráfico con lluvia y locutora de radio, <https://www.youtube.com/watch?v=XUvLZQctg7I>, (accesado Octubre 28, 2024).
- [13] Headphonesty, <https://www.headphonesty.com/2020/02/audio-frequency-spectrum-explained/>, (accesado Noviembre 04, 2024).
- [14] R. Fink, M. Latour y Z. Wallmark, *The Relentless Pursuit of Tone: Timbre in Popular Music*, Oxford University Press, 2018.
- [15] L. M. Garcia, *Sound Studies: An Interdisciplinary Journal*, 2016, **1**, 59-76.