



UBA
1821 Universidad
de Buenos Aires

.UBA **fiuba** 
FACULTAD DE INGENIERÍA

TB065
SEÑALES Y SISTEMAS

Análisis de señales

Trabajo Práctico Especial

Integrante
Martin Klöckner
Mateo Lema Roveta
Ernesto Dei Castelli

Legajo
123456
123456
123456

Correo electrónico
mklockner@fi.uba.ar
mlema@fi.uba.ar
edei@fi.uba.ar

Índice

1 Caso práctico: Identificador de canción	2
1.1 Reducción de frecuencia	2
1.2 Generación de huellas acústicas	8
1.3 Resultados	9

Índice de figuras

1 Diagrama de bloques sistema de reconocimiento	2
2 Diagrama de bloques reducción de frecuencia	3
3 Respuesta al impulso filtro pasa-bajos FIR de grado 700	5
4 Respuesta en frecuencia filtro pasa-bajos FIR de grado 700	5
5 Polos y ceros filtro pasa-bajos FIR de grado 700	6
6 Comparación potencia de ventana rectangular y Hamming	7
7 Espectrograma muestra '000002.mp3' filtrado	7
8 Huella acústica muestra '000002.mp3' completa	8
9 Huella acústica muestra '000002.mp3' segmento	9
10 Huella acústica muestra '000141.mp3'	9

1. Caso práctico: Identificador de canción

En esta parte del trabajo se verá un caso práctico de todas las herramientas analizadas en secciones previas. El objetivo es implementar una herramienta que analice un sonido musical, particularmente una canción en formato digital, e identifique la canción basándose en una lista de canciones analizadas previamente.

El procedimiento básico consiste en analizar el segmento de audio, extraer características particulares en el espacio tiempo-frecuencia (es decir, en su espectrograma) que sirvan para identificarlo luego. Al conjunto de estas características se las denomina huella digital acústica, ya que es análogo a como una huella dactilar se identifica por las convergencias, desviaciones, interrupciones y demás particularidades de las crestas papilares. Este procedimiento de extracción de huellas se utiliza tanto para confeccionar la base de datos de canciones conocidas como para posteriormente reconocer segmentos de audio sin identificar, consultando la base de datos.

El sistema de reconocimiento consta de dos bloques principales:

- El algoritmo para extraer las huellas digitales acústicas.
- La base de datos que almacena las huellas acústicas de las canciones conocidas.

El algoritmo para extraer las huellas acústicas toma como entrada el archivo con el audio y, luego de un pre-procesamiento de la señal, extrae las características y entrega como resultado la huella acústica, tal como se muestra en la figura 1.

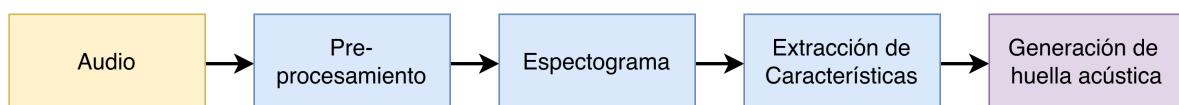


Figura 1: Diagrama de bloques sistema de reconocimiento

1.1. Reducción de frecuencia

El pre-procesamiento de la señal de audio en formato digital consiste en pasarla a canal mono (en caso de que esté en estéreo) y luego sub-muestrearlo, debido a que la información útil para la extracción de características se encuentra en bajas frecuencias. En general, el audio está muestreado a 44100 Hz, pero para el algoritmo a utilizar basta con tenerlo muestreado a 1/8 de su frecuencia original, es decir 5512,5Hz. Esto permite trabajar con menos muestras, aliviando la carga computacional.

Para reducir la frecuencia de muestreo de la señal discreta de 44100Hz a 5512,5Hz se utiliza un decimador que tome muestras de la señal a intervalos regulares de 8 puntos, de esta forma la señal sub-muestreada x_d resulta como en la ecuación 1.

$$x_d(n) = x(n \cdot N); \quad N \in \mathbb{Z} \quad (1)$$

El diagrama de bloques del sistema a utilizar se compone de un filtro pasa-bajos y luego un muestreador o compresor de la señal en dominio temporal como se muestra en la figura 2, en la cual $x(n)$ representa la señal de entrada y $x_d(n)$ la señal decimada. El filtro pasa-bajos se utiliza para evitar aliasing si la señal de entrada x_n no permite decimación en 8 veces, es decir, tiene ancho de banda mayor a $2\pi/8$ o en frecuencia siendo 5512,5Hz

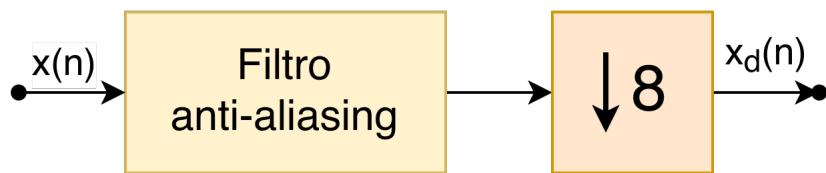


Figura 2: Diagrama de bloques reducción de frecuencia

Para el filtro pasa-bajos se utiliza un filtro con respuesta al impulso finita (FIR) y frecuencia de corte teórica de 2756,25Hz la cual corresponde con la frecuencia de Nyquist, la elección de un filtro de tipo FIR, a diferencia de un filtro con respuesta al impulso infinita (IIR), es que es mas fácil la implementación y, si bien requiere un gran poder de computo a diferencia de los filtros IIR, en este caso el poder de computo no es un problema, a diferencia de sistemas embebidas, por dar un ejemplo.

La forma general de un filtro FIR se muestra a continuación en la ecuación 2, en la cual M representa el grado del filtro, y por consiguiente la cantidad de coeficientes b_k , $y(n)$ la salida del filtro y $x(n)$ la entrada.

$$y(n) = \sum_{k=0}^M b_k \cdot x(n - k) \quad (2)$$

Aplicando la transformada Z a la expresión de la ecuación 2 y por propiedades de la transformada resulta como se ve en la ecuación 3, $H(z)$ es lo que se denomina la transferencia del sistema en el dominio z.

$$\frac{Y(z)}{X(z)} = H(z) = \sum_{k=0}^M b_k \cdot z^{-k} \quad (3)$$

Aplicando la anti-transformada z a la expresión de la ecuación 3 resulta en la respuesta al impulso del filtro FIR que se ve en la ecuación 4. Otra forma de hallar la respuesta al impulso era tomando $x(n) = \delta(n)$ en la ecuación 2.

$$h(n) = \begin{cases} b_n, & \text{si } 0 \leq n \leq M \\ 0, & \text{en otro caso} \end{cases} \quad (4)$$

Para hallar los coeficientes b_n se utiliza el método de ventana, en el cual se iguala la respuesta al impulso del filtro FIR con la multiplicación de la respuesta al impulso de un filtro ideal con una ventana con forma arbitraria, como se muestra en 5.

$$h(n) = h_{ideal}(n) \cdot w(n) \quad (5)$$

Siendo en este caso $h_{ideal}(n)$ la respuesta al impulso del pasa bajos ideal, el cual sabiendo que es un rectángulo en frecuencia, resulta una sinc en el dominio temporal, esto se muestra en la ecuación 6. La multiplicación por una ventana de la ecuación 5 justamente se hace porque los filtros ideales tienen respuesta al impulso infinita, como es el caso de la función *sinc*.

$$h_{ideal}(n) = \frac{\sin(\omega_c \cdot n)}{\pi \cdot n} = \omega_c \cdot \text{sinc}\left(\frac{\omega_c}{\pi} \cdot n\right) \quad (6)$$

Para la ventana se utiliza la ventana de Hamming, ya que es muy simple la implementación, la formula general es común con otras ventanas (como Hanning) y se muestra a continuación en la ecuación 7, en particular para la ventana de Hamming los coeficientes a_0 y a_1 toman los valores aproximados 0,54 y 0,46 respectivamente¹.

$$w(n) = a_0 - a_1 \cdot \cos\left(\frac{2\pi n}{N-1}\right); \quad w_{Hamming}(n) = 0,54 - 0,46 \cdot \cos\left(\frac{2\pi n}{N-1}\right) \quad (7)$$

Teniendo la expresión de la ventana $w(n)$ y la respuesta al impulso del filtro ideal $h_{ideal}(n)$ se obtiene la respuesta al impulso del filtro FIR $h(n)$ de la ecuación 5, en este caso tomando 700 coeficientes, es decir, se tiene un filtro de grado 700. La respuesta al impulso se muestra graficada en la figura 3 en la cual se ve claramente la forma de la función sinc.

La respuesta en frecuencia del filtro se obtiene tomando $z = j\omega$ en la función de transferencia del sistema $H(z)$ mencionada previamente en la ecuación 4. En la figura 4 se grafica la respuesta en frecuencia del filtro pasa-bajos elegido, la magnitud se muestra en color azul y la fase en color rojo. Se nota que a partir de la frecuencia de corte (2627,1 Hz, correspondiente al punto de -3dB) se atenúan drásticamente las señales y en particular para las frecuencias mayores a la de Nyquist (2756,2 Hz)

¹Ventana (función) ([Wikipedia, 2025](#))

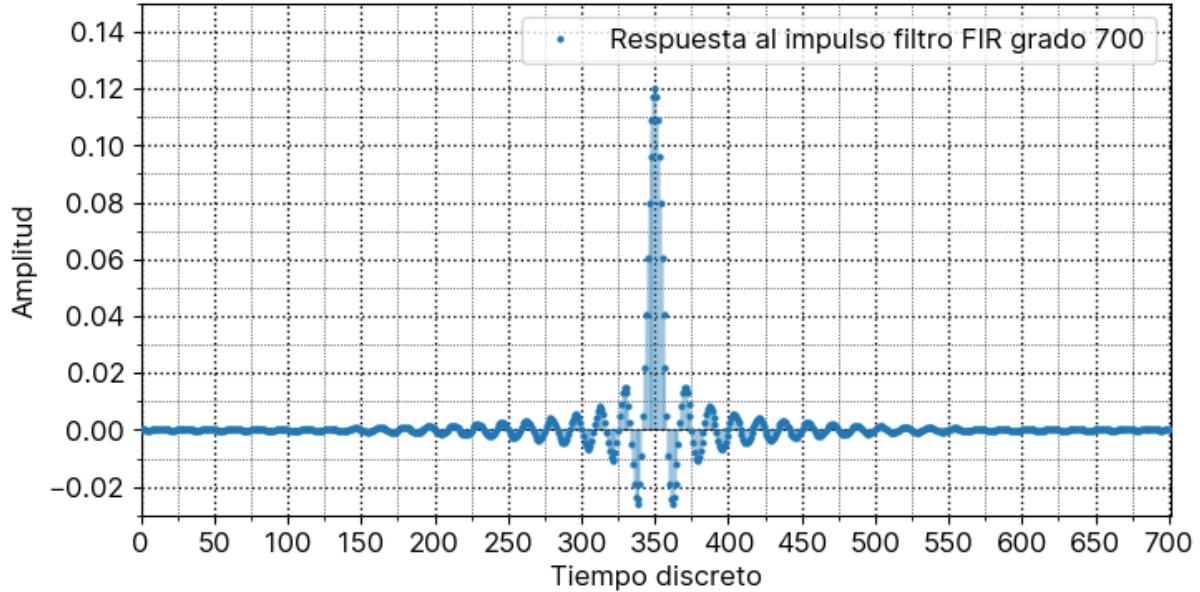


Figura 3: Respuesta al impulso filtro pasa-bajos FIR de grado 700

ya se puede decir prácticamente que se atenúan por completo ya que se tiene una ganancia de $-60dB$, esto es 0,001 veces, lo cual era el objetivo para evitar aliasing al decimar. Además se observa la fase lineal hasta la frecuencia de corte, esta es una de las principales características de los filtros FIR por sobre los IIR, que es que no distorsionan la fase.

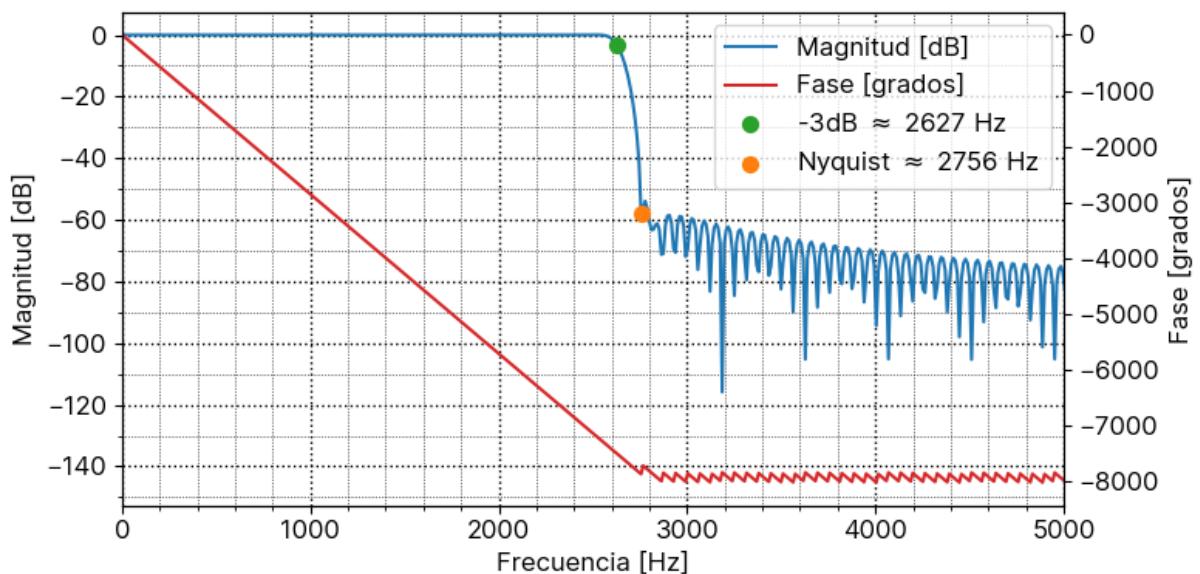


Figura 4: Respuesta en frecuencia filtro pasa-bajos FIR de grado 700

Los ceros z_i y polos p_i se definen como aquellos puntos del plano complejo en donde la función de transferencia del sistema $H(j\omega)$ se anula, en el caso de los ceros y diverge a infinito, en el caso de los polos. La particularidad de los filtros FIR es que no tienen polos distintos de cero, esto sale de analizar la ecuación 3, se tiene una sumatoria de coeficientes no nulos constantes multiplicados por la variable z^{-1} , la única forma de que la transferencia tienda a infinito es que la variable z tienda a 0 de manera de que z^{-1} diverja.

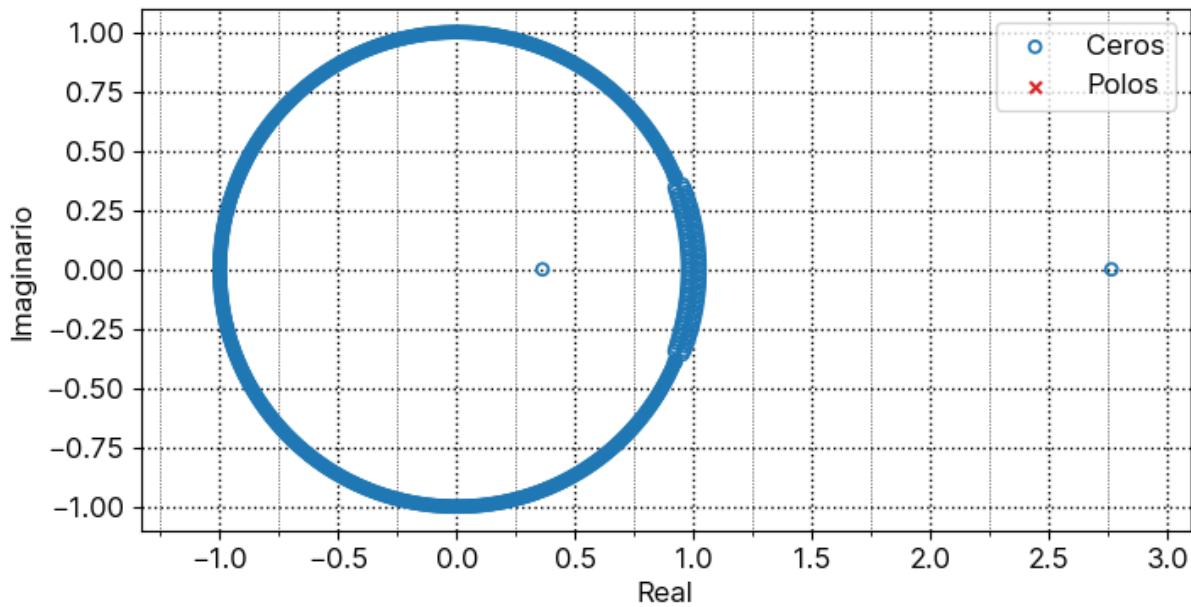


Figura 5: Polos y ceros filtro pasa-bajos FIR de grado 700

La principal ventaja de utilizar la ventana de Hamming en lugar de la rectangular al realizar un espectrograma es la reducción drástica de la fuga espectral, lo que resulta en una representación de frecuencia más limpia y precisa de las señales en el tiempo.

Cuando se calcula un espectrograma, el proceso implica dividir la señal continua en pequeños segmentos (ventanas) para calcular la Transformada Rápida de Fourier (FFT) de cada segmento.

La ventana rectangular es un corte brusco de la señal. Cuando esta ventana se multiplica con la señal en el dominio del tiempo, crea una discontinuidad en los puntos de inicio y fin de la ventana.

En el dominio de la frecuencia, esta discontinuidad se traduce en un espectro con lóbulos laterales muy altos (el primer lóbulo está a solo 13 dB por debajo del pico principal).

La ventana de Hamming tiene un lóbulo principal ligeramente más ancho que la rectangular, lo que resulta en una peor resolución de frecuencia (es más difícil distinguir

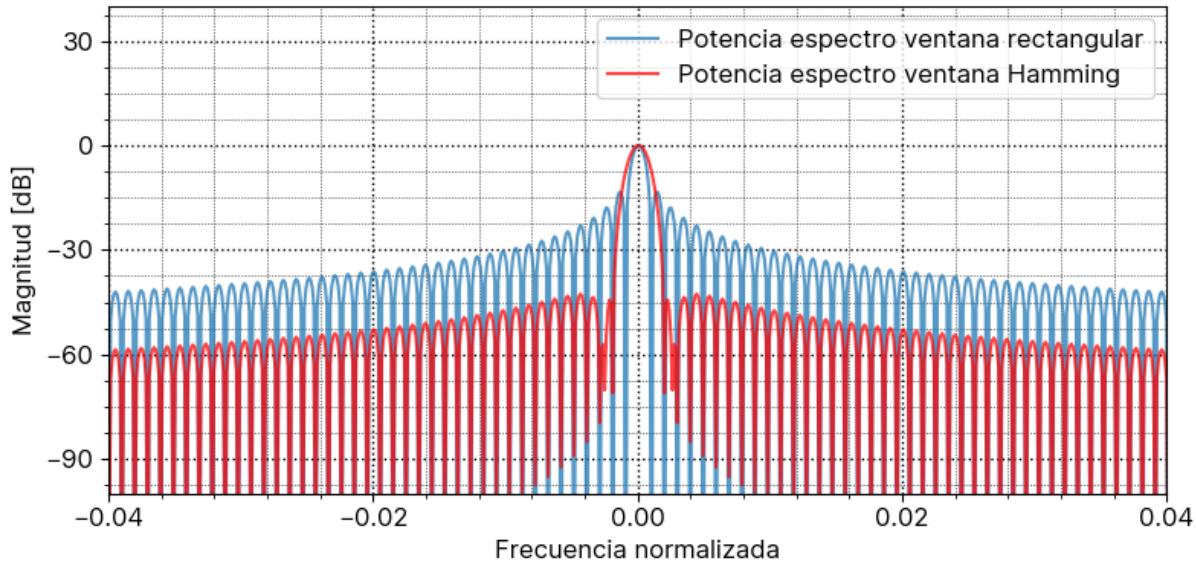


Figura 6: Comparación potencia de ventana rectangular y Hamming

dos frecuencias muy, muy cercanas)

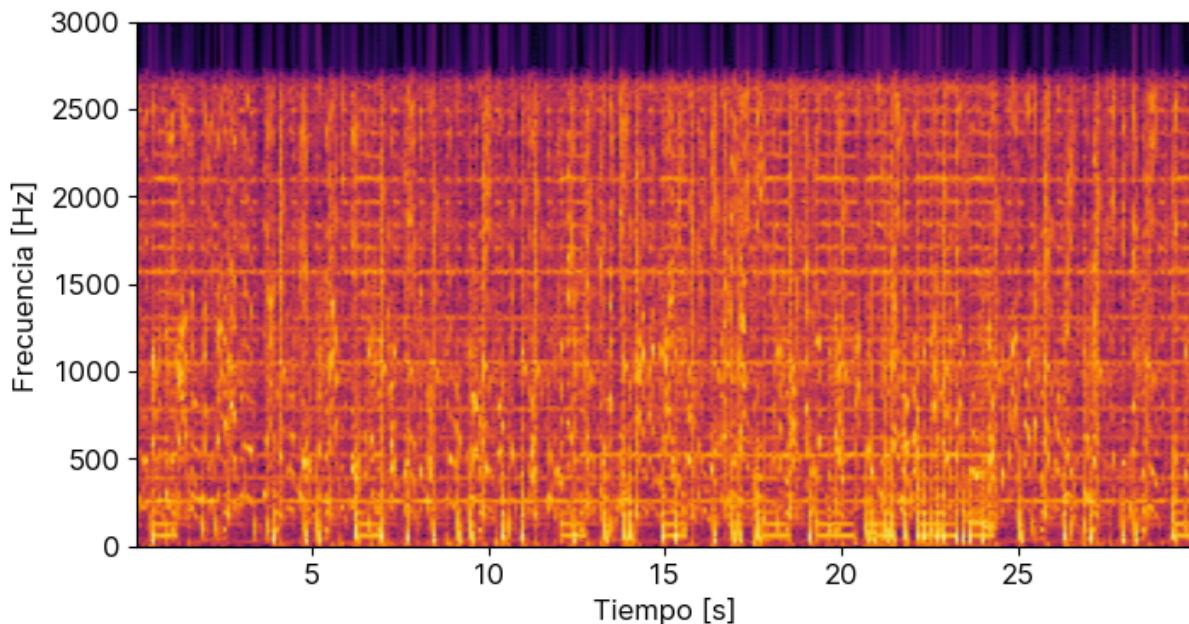


Figura 7: Espectrograma muestra '000002.mp3' filtrado

1.2. Generación de huellas acústicas

El algoritmo de extracción de huellas digitales acústicas extrae características a partir del espectrograma del audio. El conjunto de estas características conforman la huella.

Las características se obtienen mediante una función signo $F(m, n)$ que opera sobre la energía de las bandas $E(m, n)$ según la ecuación 8, donde n es el tiempo y m la banda de frecuencia.

$$F(m, n) = \begin{cases} 1 & E(m, n) - E(n, m + 1) > E(n - 1, m) - E(n - 1, m + 1) \\ 0 & \text{en otro caso} \end{cases} \quad (8)$$

En la figura 8 se muestra un gráfico de la huella generada para una muestra ‘000002.mp3’ utilizando el algoritmo mencionado previamente. Luego en la figura 9 se toma el segmento de 0 a 100 de manera que se aprecie aun mejor el patrón de unos y ceros. En la figura 10 se grafica la huella de otra muestra para el mismo segmento anterior, de 0 a 100, nótese la diferencia entre el patrón generado.

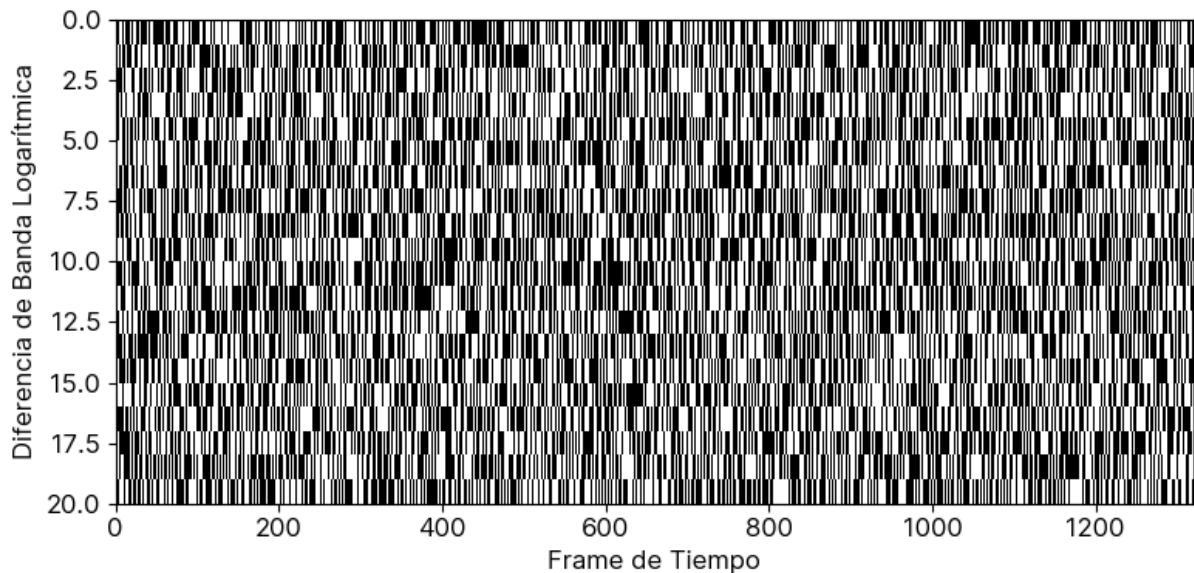


Figura 8: Huella acústica muestra ‘000002.mp3’ completa

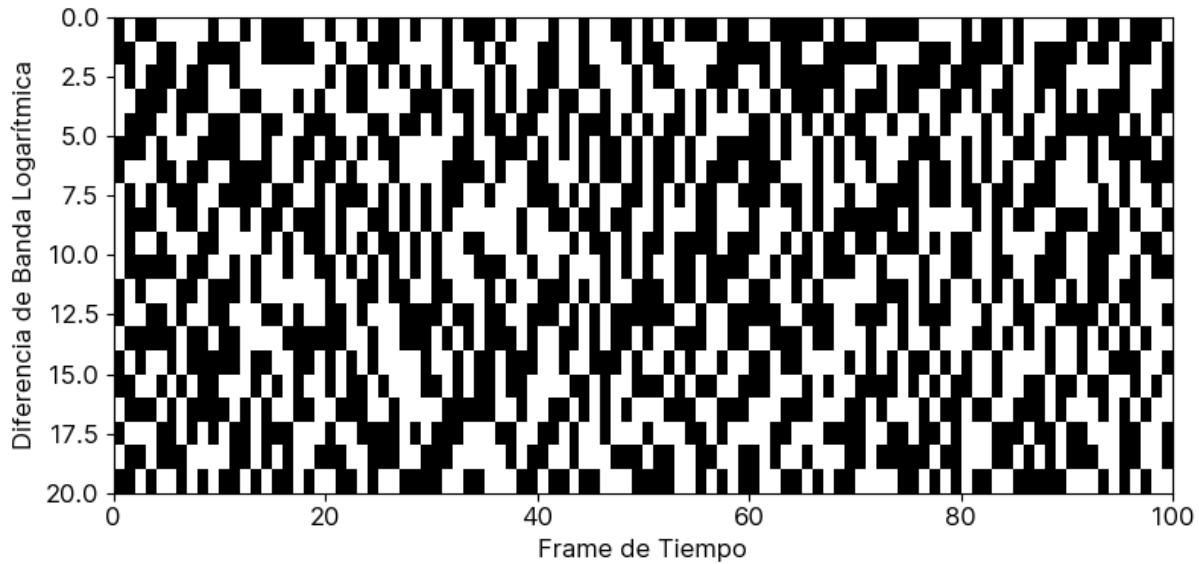


Figura 9: Huella acústica muestra '000002.mp3' segmento

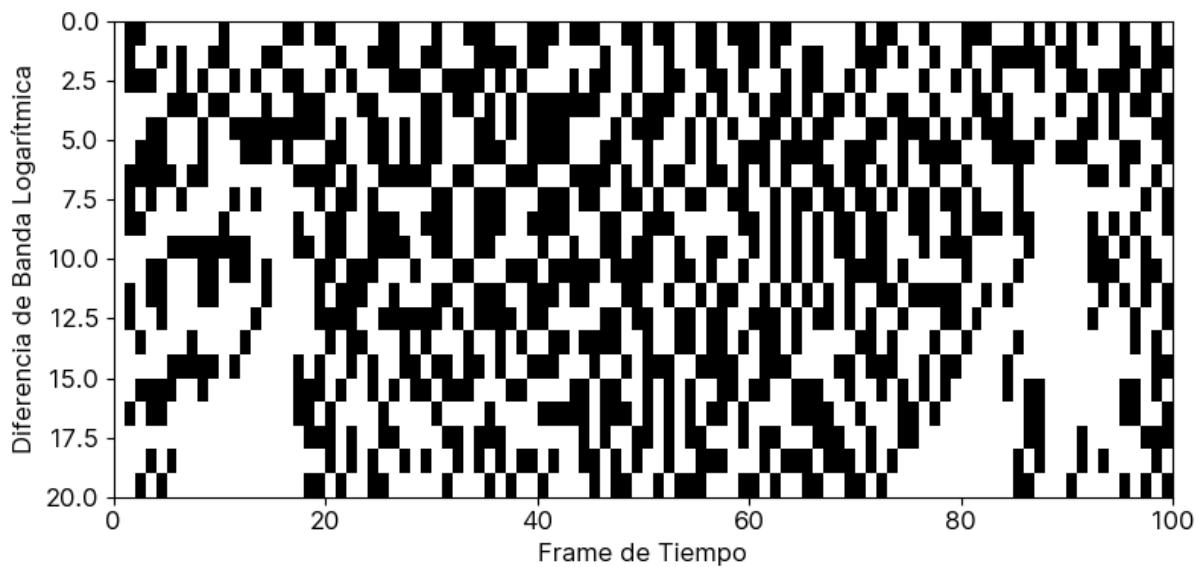


Figura 10: Huella acústica muestra '000141.mp3'

1.3. Resultados

Utilizando 16 canciones de muestra, se generaron las huellas como se menciono previamente y se cargaron estas huellas a la base de datos. Para verificar cual es el error al identificar canciones se generaron pruebas procedurales, en particular se genero un total de 2550 identificaciones de las cuales 29 fueron erróneas, esto resulta en un error porcentual de 1,14 %. En una primera instancia el error fue mayor, pero se

corrigió mediante el ajuste del ancho de ventana y del ancho de solapamiento, a la hora de calcular el espectrograma para generar la huella de las canciones de muestra (las que se guardan en la base de datos).

Referencias

Wikipedia (2025). Ventana (función) — wikipedia, la enciclopedia libre. [Internet; descargado 26-noviembre-2025].