

# TÉCNICA DE MARCA DE AGUA EN AUDIO A PARTIR DE LA STFT Y SVD

Bach Franco<sup>1</sup> y Mezzabarba Tomás<sup>2</sup>

<sup>1</sup>Universidad Nacional de Tres de Febrero  
francobach47@gmail.com

<sup>2</sup>Universidad Nacional de Tres de Febrero  
tmezzabarba@gmail.com

**Resumen** — *Diferentes técnicas de marca de agua son utilizadas día a día para protección de derechos de autor, autenticación de datos, protección de contenido, entre otros. En el presente trabajo se presentan la implementación de un algoritmo mediante la transformada de Fourier de tiempo corto (STFT) y la descomposición en valores singulares (SVD), para el incrustado y detección de una marca de agua en señales de audio. El algoritmo es evaluado y se analiza la imperceptibilidad y la robustez de la marca de agua.*

## 1. INTRODUCCIÓN

Los rápidos avances en el campo de la tecnología de audio digital han aumentado la facilidad para almacenar, distribuir y reproducir los archivos de audio. Esto, inevitablemente conlleva un riesgo de seguridad inherente de uso ilegal de datos y violación de derechos de autor.

La técnica de marca de agua de audio digital proporciona una solución prometedora para combatir dichas inseguridades.

La marca de agua en el audio encuentra aplicaciones en diversas áreas como la protección de derechos de autor, autenticación de datos, comunicaciones encubiertas, adición de metadatos, identificación de contenido y subtitulación o etiquetado de datos, entre otros. La misma consta de un proceso de incrustación en el cual se utiliza la firma del propietario o alguna información de derechos de autor en la señal de audio. Ésta generalmente es aplicada mediante texto o imagen.

El rendimiento de esta técnica de audio se evalúa básicamente en tres categorías principales: imperceptibilidad, robustez y carga útil.

La primera refiere a la calidad de la señal de audio a restaurar luego del agregado de la marca de agua. La imperceptibilidad es cuantificada mediante el signal-to-noise ratio (*SNR*) y el peak-signal-to-noise ratio (*PSNR*). La robustez refleja la capacidad de recuperar correctamente los bits de marca de agua con y sin ata-

ques. La misma es evaluada calculando el factor de correlación ( $\rho$ ). Por otro lado, la carga útil se define a través de la capacidad de incrustación del algoritmo de marca de agua.

## 2. MARCO TEÓRICO

El método empleado se basa en la STFT (Short Time Fourier Transform) y en el algoritmo SVD (Singular Value Decomposition). Por este motivo, a continuación se definen términos específicos que serán de utilidad para una mayor comprensión del análisis e implementación posterior.

### 2.1. Short Time Fourier Transform (STFT)

La transformada de Fourier de tiempo corto (STFT) es una secuencia de transformadas de Fourier de una señal ventaneada. Nos proporciona un análisis temporal-frecuencial que extrae el espectro frecuencial de una señal, a través de intervalos ventaneados de la señal. La ecuación de la misma está dada por:

$$STFT_x(t, f) = \int x(\tau)g(\tau - t)e^{-j2\pi f\tau} d\tau \quad (1)$$

$$x(\tau) = \iint STFT_x(t, f)g(\tau - t)e^{j2\pi f\tau} dt df \quad (2)$$

Donde  $g(t)$  es la ventana utilizada y  $x(t)$  la señal. Al deslizar la función  $g$  sobre la señal, multiplicándolos y calculando la transformada de Fourier del producto, obtenemos una representación bidimensional en tiempo y frecuencia de la señal. [1]

## 2.2. Singular Value Decomposition (SVD)

La descomposición en valores singulares (SVD), es la descomposición de una matriz en tres matrices de la siguiente manera:

$$A = USV^T \quad (3)$$

En donde  $A$  es la matriz a descomponer de tamaño  $m \times n$ ,  $U$  es una matriz ortogonal de tamaño  $m \times m$ ,  $S$  es la matriz diagonal de valores singulares de tamaño  $m \times n$ , y  $V^T$  es la transpuesta de una matriz ortogonal de tamaño  $n \times n$ . [1]

## 2.3. Test de prueba

Se realizarán distintas pruebas a las señales con marcas de agua. Principalmente, se evaluará la audibilidad y la robustez de la marca de agua. Por lo tanto se definen las siguientes relaciones que serán utilizadas para medir los parámetros mencionados.

### 2.3.1. Test de audibilidad

$$SNR(dB) = 10 \log_{10} \left( \frac{\sum_{x,y} p_{x,y}^2}{\sum_{x,y} (p_{x,y} - p'_{x,y})^2} \right) \quad (4)$$

$$PSNR(dB) = 10 \log_{10} \left( \frac{XY \max(p_{x,y}^2)}{\sum_{x,y} (p_{x,y} - p'_{x,y})^2} \right) \quad (5)$$

Estos parámetros se miden sobre la matriz STFT de la señal, la cual se representa con  $p$ , y la matriz de la señal con marca de agua, la cual se representa con  $p'$ . Por lo tanto  $p_{x,y}$  y  $p'_{x,y}$  representa un pixel en las coordenadas  $(x,y)$  de cada matriz.  $X$  e  $Y$  representan el número de filas y columnas, respectivamente. [2][3][5]

### 2.3.2. Test de robustez

La robustez se estudiará sobre la matriz de marca de agua, y una matriz de marca de agua extraída luego de aplicar un ataque a la señal con marca de agua. Esto será explicado más adelante.

Luego, para el test se hará uso del factor de correlación  $\rho$ , el cual se define a continuación

$$\rho(w, w') = \frac{\sum_{i=1}^N w_i w'_i}{\sqrt{\sum_{i=1}^N w_i^2} \sqrt{\sum_{i=1}^N w'^2_i}} \quad (6)$$

donde  $w_i$  es la marca de agua original,  $w'_i$  es la marca de agua extraída, y  $N$  representa el número de píxeles. [4]

## 3. PROCEDIMIENTO

En esta sección se describe y desarrolla el procedimiento realizado en este trabajo. Entre ellos, el procesamiento inicial de la señal, el método empleado para incrustar la marca de agua, su detección y extracción, y por último los test realizados, en los cuales se mide la audibilidad y la robustez de las señales con marca de agua.

Para todos los pasos se ha hecho uso de Python 3.9.7, y las siguientes librerías: Numpy, soundfile, librosa, scipy, matplotlib, skimimage, cv2, wave y audioop.

Todo el proceso que se redacta en este trabajo se realiza sobre 3 señales de audio. Un discurso con una  $f_{sampleo} = 88200Hz$ , un canto a capella con  $f_{sampleo} = 48000Hz$  y una canción con  $f_{sampleo} = 44100Hz$ ; las cuales tendrán una duración aproximada de 5 segundos. Sus formas de onda y espectrogramas son exhibidos en las Figuras 1 y 2, respectivamente.

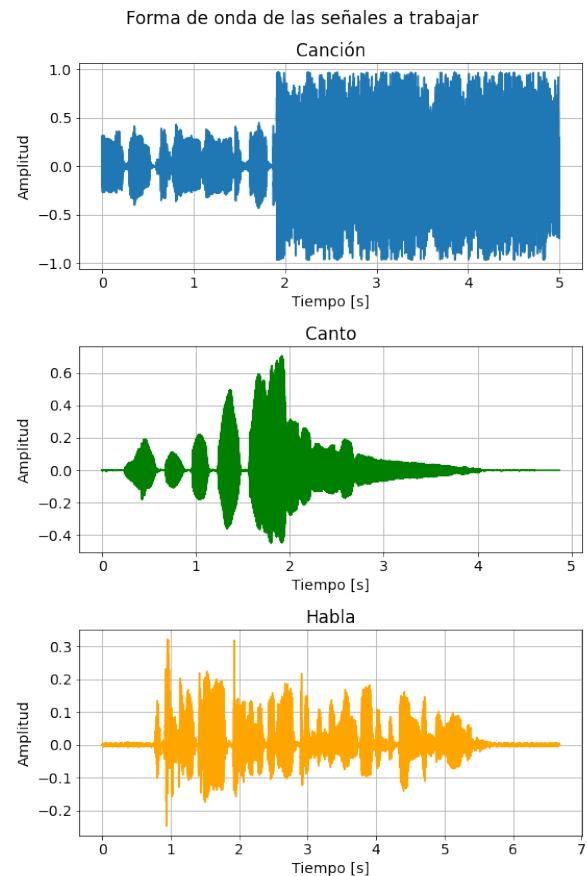


Figura 1: Formas de onda de las señales.

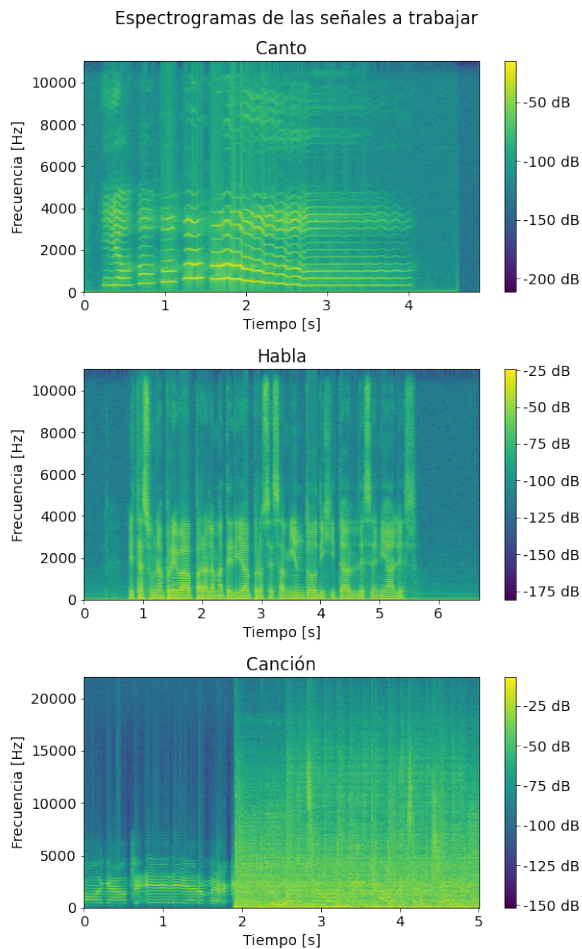


Figura 2: Espectrogramas de las señales.

### 3.1. Procesamiento de la señal

Como se mencionó, se trabaja con la matriz STFT y el algoritmo SVD. Por lo tanto, una vez importada la señal, previamente a realizar el incrustado de la marca de agua, se obtiene la matriz STFT de dicha señal y posteriormente, se la descompone mediante (3) en las matrices características del SVD.

Una vez obtenidas estas tres matrices,  $U$ ,  $D$  y  $V^T$ , se procede al incrustado de la marca de agua, el cual se realiza mediante dos métodos que se desarrollan a continuación.

### 3.2. Incrustado de la marca de agua

Para ambos métodos, se utiliza y trabaja sobre dos matrices principales. Una matriz que corresponderá a la STFT del audio en análisis, y por otro lado, la matriz binaria de la imagen que se usará como implementación de marca de agua.

La Figura 3 representa la imagen utilizada para la implementación de la marca de agua.



Figura 3: Imagen que se utiliza como marca de agua.

De todos modos, se necesita su representación binaria de la misma, por lo tanto, la misma es sometida a un pre procesamiento previo a estar disponible para su utilización como marca de agua.

Este pre procesamiento consiste en pasar la imagen a escala de grises, esto dado que es el formato típico utilizado en la incrustación de marcas de agua. En este formato, cada pixel tiene un valor de intensidad el cual puede ir de 0 a 255. Posteriormente, con la imagen en escala de grises, se la pasa a una representación binaria, de manera tal que queda solo blanco y negro (1 y 0). La elección a partir de que valor se lo pasa blanco o negro se da a partir de un umbral que es elegido mediante prueba, de manera tal que la representación binaria no se vea distorsionada y sea lo más parecida a la imagen original en escala de grises.

Una vez con la imagen en su representación binaria y la matriz STFT del audio, se pasa a la incrustación de la marca de agua. En el presente trabajo, se utilizan los siguientes valores del factor de fuerza  $\alpha$ : 0.0025, 0.0050, 0.0100, 0.0500.

#### 3.2.1. Método 1 - SVMW

El primer método se lo denomina SVMW, y su procedimiento se detalla en la Figura 4 y se lo explica a continuación.

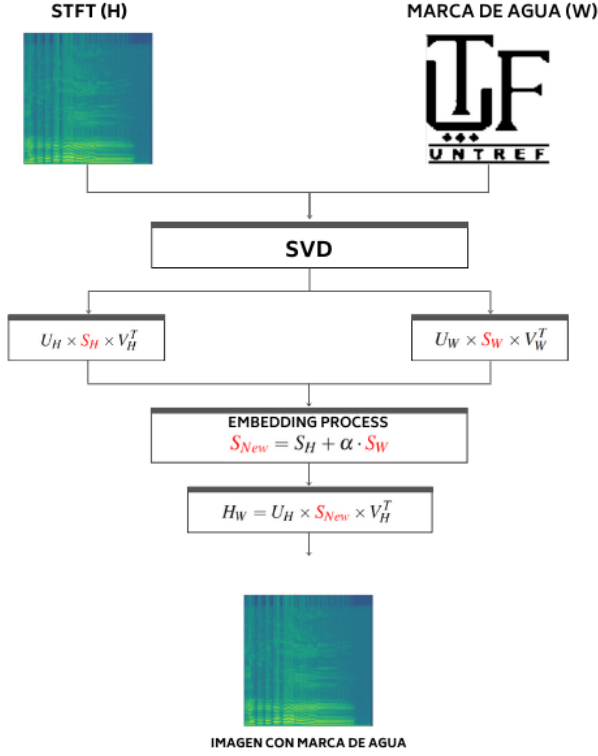


Figura 4: Método SVMW para la incrustación de marca de agua.

Mediante (3) se aplica SVD a la matriz STFT y a la imagen de marca de agua, obteniendo la descomposición en matrices  $H = U_H * S_H * V_H^T$  y  $W = U_W * S_W * V_W^T$ , para la matriz STFT y la matriz de la marca de agua, respectivamente.

Luego, se procede a realizar el incrustado de la marca de agua, el cual se efectúa mediante la siguiente relación,

$$S_1 = S_H + \alpha * S_W \quad (7)$$

en donde  $S_H$  representa la matriz de valores singulares de la matriz STFT del audio y  $\alpha$  es un factor de escala que será analizado más adelante. Esto devuelve una matriz  $S_1$  la cual es utilizada para obtener la matriz final del audio con la marca de agua  $H_W$ , aplicando la inversa de la SVD mediante:

$$H_W = U_H * S_1 * V_H^T \quad (8)$$

Por último, a esta matriz se le aplica la ISTFT para obtener la señal con marca de agua en el tiempo, y así poder exportarla y escucharla. [2]

### 3.2.2. Método 2 - DW

Al segundo método se lo denomina DW, y su procedimiento se detalla en la Figura 5 y se lo explica a continuación.

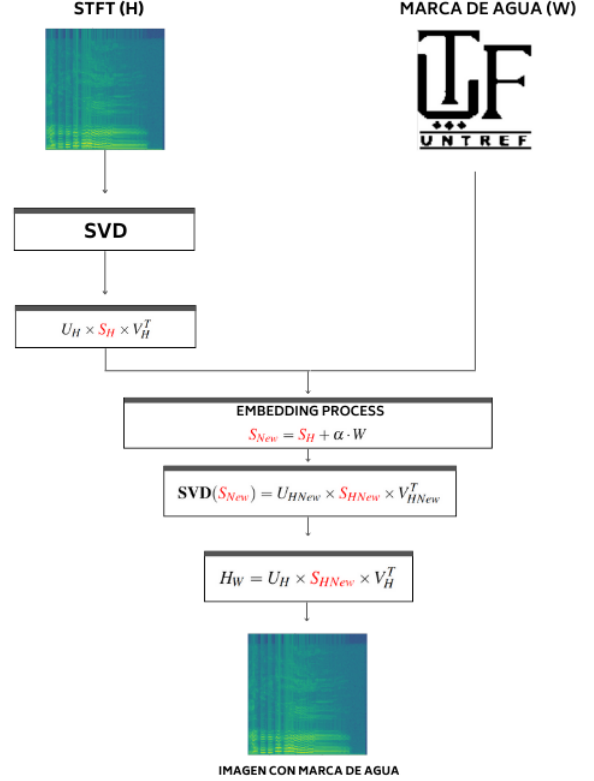


Figura 5: Método DW para la incrustación de la marca de agua.

En este caso, a diferencia del primer método, se aplica el SVD solo a la matriz STFT y con su descomposición de matrices  $U_H * S_H * V_H^T$  se incrusta la marca de agua, a partir de

$$S_2 = S_H + \alpha * W \quad (9)$$

en donde  $W$  representa la matriz en forma binaria de la imagen de marca de agua. Devuelve  $S_2$  y se aplica el mismo proceso que en el método 1, o sea, se busca la matriz de la señal con la marca de audio ya incrustada, y se aplica la ISTFT para poder tener la señal en el tiempo y poder exportarla y escucharla. [2]

Para trabajar computacionalmente, y no tener problemas algebraicos con los tamaños de las matrices, previo a todo el proceso en ambos métodos, se iguala el tamaño de la matriz STFT y el tamaño de la matriz de la marca de agua. Para esto, se realiza un zero padding a la matriz de la marca de agua, la cual originalmente va a tener dimensiones menores a la matriz STFT. Esto dado a las dimensiones de la imagen original, y a las dimensiones de la matriz STFT, la cual estará dada por el tamaño de las ventanas y el número de muestras aplicado en el overlapping. Para este trabajo, se emplea una ventana de hamming de 1024 muestras con un overlapping de 512

muestras; y además, se evalúa el proceso de incrustado y extracción con otros tipos de ventanas.

Se puede observar que ambos métodos comparten en su ecuación de incrustado de marca de agua al factor de escala o de fuerza  $\alpha$ . Este factor será uno de los más importantes a la hora de su elección, dado que será clave en la audibilidad y robustez de la marca de agua en la señal. Se busca un valor que compense ambos factores, o sea, que genere una marca de agua en la señal, pero a su vez que sea lo menos perceptible posible.

Por otro lado, la diferencia principal entre ambos métodos reside en que uno trabaja directamente sobre la matriz de marca de agua (método 2), y el otro trabaja con su matriz de valores singulares (método 1).

De todos modos, en ambos casos se cumplen los siguientes factores:

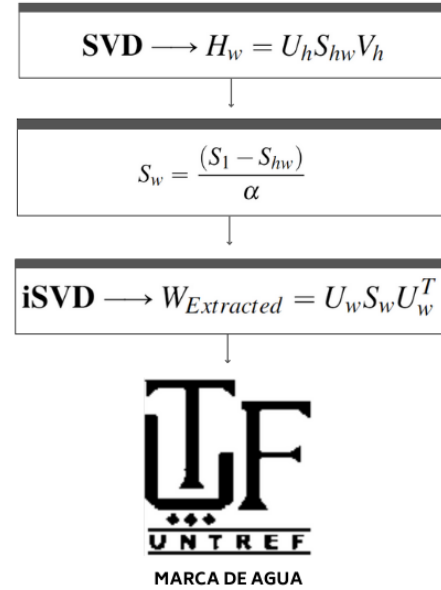
- La matriz de valores singulares  $S$  no contiene información significativa acerca de la estructura de la imagen. Expresa simplemente la luminosidad de dicha imagen.
- Las matrices  $U$  y  $V$  contienen información acerca de la estructura de la imagen y tienen efecto en el contenido de la imagen.
- Por otro lado, si los valores singulares de  $S$  son modificados enteramente, el contenido de la imagen no cambiará.

### 3.3. Detección de la marca de agua

Para la extracción de la marca de agua se emplean dos métodos diferentes para cada uno de incrustación de la misma.

#### 3.3.1. Método 1 - SVMW

El primer método de extracción corresponde básicamente a realizar los pasos inversos a la incrustación con el método SVMW a fin de obtener la imagen de marca de agua utilizada. En la Figura 6 se pueden visualizar los pasos a seguir para la obtención de la misma. Posteriormente se realiza una explicación paso a paso.



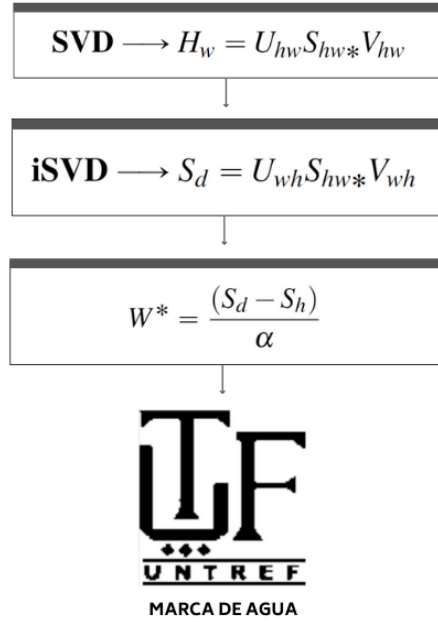


Figura 7: Método DW para la extracción de marca de agua.

Al igual que en el método 1, en este caso se comienza aplicando SVD a la matriz  $H_w$ . Posteriormente se aplica la inversa de SVD utilizando la matriz de valores singulares obtenida anteriormente ( $S_{hw*}$ ) y las matrices  $U_{HNew} - V_{HNew}^T$  para obtener  $S_d$ . Finalmente se obtiene la marca de agua  $W^*$  utilizando la llave ( $\alpha$  y  $S_h$ ).

### 3.4. Análisis y tests de prueba

Como se mencionó, se mide la audibilidad y la robustez de la marca de agua.

#### 3.4.1. Test de Audibilidad

Para medir la audibilidad de la marca de agua se hace uso del signal-to-noise ratio (SNR) y del peak-signal-to-noise ratio (PSNR), mediante (4) y (5), respectivamente.

En este test se mide la imperceptibilidad de la marca de agua. Se busca que sea mínima la percepción de ésta y que la distorsión de la señal sea mínima.

Ambos métodos de medición miden la similitud entre la señal original y la señal distorsionada con la marca de agua. Cuanto mayor sea el valor en ambos métodos, implica mayor relación entre las señales en análisis. Por lo tanto, cuanto más distorsionada sea la señal con marca de agua, menor será el valor de los parámetros que se miden.

#### 3.4.2. Test de Robustez

Para medir la robustez de la señal, primero se tiene que someter la señal con marca de agua a un ataque. Luego, se extrae la marca de agua de la señal atacada, y se la compara con la marca de agua original.

Se proponen los siguientes ataques a realizar a la se-

ñal con marca de agua:

- Agregado de Eco
- Filtro pasa bajos
- Filtro pasa altos
- Filtro pasa banda
- Agregado de ruido Gaussiano
- Rotación de la señal con marca de agua

Cabe aclarar que para definir cada ataque se desarrolla un algoritmo que toma como parámetro entrada a la señal que contiene marca de agua y entrega a la salida la misma señal de entrada pero modificada según el ataque.

En el caso del Eco se utiliza un retardo de  $500ms$ , un factor de  $0,5$  que básicamente va a disminuir la amplitud de la señal cada vez que se repita, y un número de 3 repeticiones en todas las pruebas.

Para los filtros, se implementaron filtros pasa bajos, pasa altos y pasa banda, con  $f_{corte} = 1500$  Hz,  $3000$  Hz y banda de paso de  $500$  a  $5000$  Hz, respectivamente.

Para el agregado de ruido Gaussiano, se sumó a la señal un ruido con  $\mu = 0$  y  $\sigma = 0.1$ , y en cuanto a la rotación de la matriz de la señal con marca de agua, fue de  $45^\circ$ .

Una vez atacada la señal, se procede a medir el factor de correlación (6).

La robustez refleja la capacidad de recuperar correctamente los bits de marca de agua con y sin ataques. El factor de correlación estará dentro de un rango  $(0,1)$ , donde  $0$  y  $1$  representa una relación al azar y perfectamente lineal, respectivamente. De otro modo, cuando la marca de agua extraída y la original sean muy cercanas, el  $\rho \approx 1$ . Si  $\rho = 1$ , la marca de agua original y la extraída son iguales.

## 4. RESULTADOS Y ANÁLISIS

Una vez con las marcas de agua incrustadas con los valores de  $\alpha$  mencionados anteriormente, se procede primero que nada a escuchar dichas señales generadas. Se aprecia que a menor factor de fuerza, menos audibilidad. A medida que dicho factor aumenta, aumenta la distorsión en la señal. De todos modos, no todas las señales responden de la misma manera. La señal canto es la que más fácil distorsiona, o sea, requiere un menor factor de fuerza para que la marca de agua sea audible; por otro lado, en las señales de habla y la canción, se permite más margen para dicho valor. Incluso en la canción, para un valor de  $\alpha$  en el cual el canto se distorsiona en una gran medida, no sucede lo mismo con esta señal, generando un efecto incluso de ecualización, con un realce para las frecuencias graves. Se analiza más adelante cuando se ataca la señal con distintos filtros, como se comporta la marca de agua extraída, pero esto ya da un indicio de que

la marca de agua incrustada se sintoniza en frecuencias medias.

No obstante, la señal canto requiere un factor de fuerza menor, esto puede deberse a que el mismo se asemeja a un canto lírico, el cual tiene saltos pronunciados en amplitud y en frecuencia.

Posteriormente se procede a la extracción de las marcas de agua mediante ambos métodos. Una vez extraídas, se realiza la correlación cruzada entre esta matriz y la matriz de marca de agua original. Para poder ver el porcentaje de relación entre dichas matrices, se realiza la autocorrelación de la matriz marca de agua, y luego se hace un cociente entre el máximo valor de ambas operaciones. No obstante, no se presentan dichos valores dado que la correlación cruzada del método 2 da 1, y el método 1 da 0. Esto es dado a que el método 2 es más eficiente para la extracción de la marca de agua. El método 1 extrae la forma y se puede ver la marca de agua, pero en escala los valores son distintos. Lo mencionado, se aprecia en las Figura 8, 9 y 10, en las cuales se presentan extracciones de marca de agua para las distintas señales, mediante ambos métodos, con un factor de fuerza  $\alpha = 0,005$ .

Se puede ver, que el método 2 extrae la misma o prácticamente la misma imagen que es insertada, y la 1 extrae la forma, pero los valores no son iguales, tendiendo a cero.

La parte que no tiene imagen, sino que solo tiene color negro/gris, es dado a que la imagen de marca de agua se escala a los valores de la STFT de cada señal, para trabajar con dimensiones iguales, y esto se realiza mediante zero padding.

Extracciones de marca de agua con  $\alpha = 0.005$  - Señal: Canto

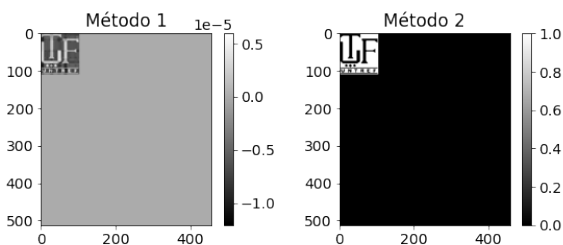


Figura 8: Marcas de agua extraídas con ambos métodos para señal canto ( $\alpha = 0,005$ )

Extracciones de marca de agua con  $\alpha = 0.005$  - Señal: Habla

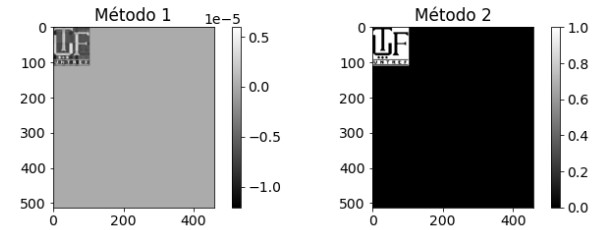


Figura 9: Marcas de agua extraídas con ambos métodos para señal habla ( $\alpha = 0,005$ )

Extracciones de marca de agua con  $\alpha = 0.005$  - Señal: Canción

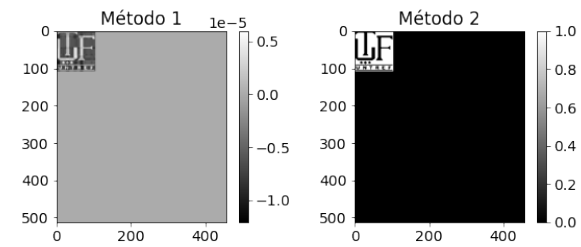


Figura 10: Marcas de agua extraídas con ambos métodos para señal canción ( $\alpha = 0,005$ )

Esta diferencia en la extracción se da principalmente a como es el incrustado de la marca de agua, dado que el método 1 el incrustado se da en base a la matriz de valores singulares de la matriz marca de agua, y el método 2 trabaja sobre toda la matriz en conjunto.

Por otra parte, se analiza el incrustado y la extracción probando la STFT con distintos tipos de ventana. Para esto se utiliza aparte de la de Hamming que es con la que se trabaja durante todo el estudio, se prueba con la de Hann, Blackman y Boxcar. En las Figuras 11 y 12 se muestra la extracción mediante distintas ventanas con un mismo factor de fuerza.

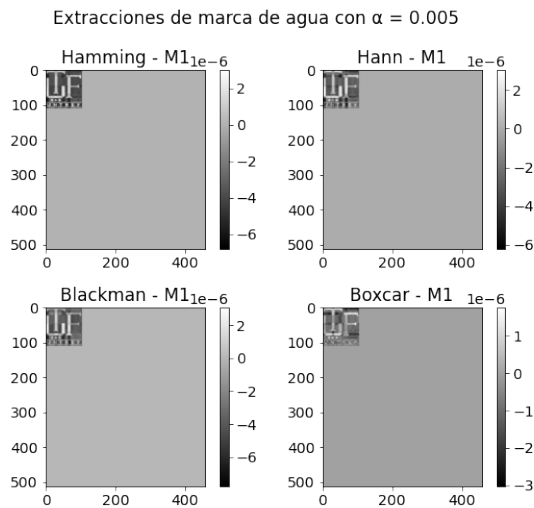


Figura 11: Extracción de marcas de agua con distintas ventanas  
- Método 1

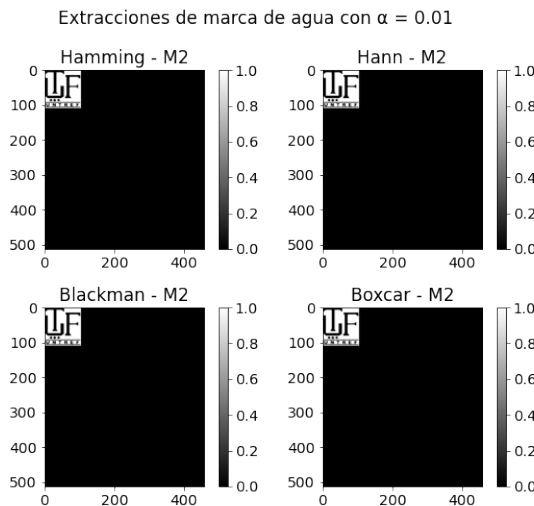


Figura 12: Extracción de marcas de agua con distintas ventanas  
- Método 2

Como se puede observar, en el primer método existe una diferencia entre el tipo de ventana que se aplica, obteniendo una extracción con mayor resolución con las de Hamming, Hann y Blackman, por sobre la de Boxcar. Esto puede estar relacionado a los anchos de los lóbulos principales de cada ventana. Por otro lado, en el método 2 las extracciones no tienen tanta dependencia en el estilo de ventana utilizado, debido a que el incrustado es realizado solo sobre la matriz de marca de agua, y no sobre sus valores singulares como en el primero.

#### 4.1. Test de audibilidad

En las siguientes tablas 1 y 2, se presentan los valores del signal-to-noise ratio y del peak-signal-to-noise ratio,

respectivamente.

Tabla 1: Valores del SNR

Signal-to-noise ratio		
$\alpha$	SVMW [dB]	DW [dB]
Canto		
0.0025	17.57	29.63
0.0050	11.55	22.63
0.0100	5.53	17.37
0.0500	-8.45	-5.88
Habla		
0.0025	13.43	22.50
0.0050	7.41	16.79
0.0100	1.39	9.13
0.0500	-12.59	-10.91
Canción		
0.0025	27.30	27.30
0.0050	21.28	38.60
0.0100	15.26	29.60
0.0500	1.28	7.54

Se puede ver que a medida que se aumenta el factor de fuerza, las relaciones disminuyen, esto dado a que como se mencionó, a mayor valor de factor de fuerza, la audibilidad de la marca de agua es mayor, y por lo tanto, la distorsión en la señal. Por otro lado, los valores obtenidos con los valores más chicos de  $\alpha$  cumplen con lo esperado, una alta relación de dichos parámetros, lo que indica una baja perceptibilidad.

Tabla 2: Valores del PSNR

Peak-signal-to-noise ratio		
$\alpha$	SVMW [dB]	DW [dB]
Canto		
0.0025	51.55	63.61
0.0050	45.53	56.61
0.0100	39.51	48.35
0.0500	25.53	28.10
Habla		
0.0025	53.74	62.80
0.0050	47.72	57.09
0.0100	41.69	49.43
0.0500	27.72	29.39
Canción		
0.0025	59.52	78.58
0.0050	53.50	70.87
0.0100	47.48	61.84
0.0500	33.50	39.76

#### 4.2. Test de robustez

Como se mencionó, previo a calcular el factor de correlación, se ataca la señal y luego se comparan las ma-



trices de marca de agua original y extraída post ataque. En la tabla 3 se presentan los valores obtenidos del factor de correlación, para todas las señales en ambos métodos, con distintos valores del factor de fuerza. M1 y M2 representan el método 1 y 2, respectivamente.

Como se observa, hay ataques en los cuales el método 2 funciona mejor como detección y extracción, y otros en los cuales el método 1.

Esto está relacionado a lo que se mencionó respecto a la escala de valores en las matrices extraídas. El método 1 extrae matrices en los cuales sus valores son próximos a cero. Por lo tanto, si se comparan estas matrices con la original, o con la extraída por el método 2 la cual es prácticamente igual, su porcentaje de relación dará muy bajo.

De todos modos, que un método extraiga la misma matriz y el otro extraiga una con valores próximos a cero, no implica que sea algo sin utilidad o malo. Hay ataques en los cuales un método es más efectivo que el otro y vice versa, o puede ser más útil. Aunque sea, para dar una idea de como es la forma de la marca de agua, como se ve en las Figuras 8, 9 y 10. El método 1 no entrega valores tan certeros sobre los bits y píxeles, pero si da una idea de la forma de la marca de agua, la cual en distintos momentos o aplicaciones puede llegar a ser de utilidad.

Tabla 3: Factor de correlación

$\alpha$	Canto		Habla		Canción	
	M1	M2	M1	M2	M1	M2
Rotación 45°						
0.0025	0.88	0.02	0.92	0.00	0.81	0.04
0.0050	0.92	0.02	0.95	0.01	0.83	0.03
0.0100	0.95	0.01	0.98	0.04	0.86	0.02
0.0500	0.99	0.03	0.99	0.36	0.96	0.04
Eco						
0.0025	0.63	0.83	0.84	0.95	0.95	0.95
0.0050	0.56	0.95	0.83	0.98	0.97	0.97
0.0100	0.26	0.98	0.83	0.99	0.98	0.98
0.0500	0.88	0.93	0.58	0.99	0.99	0.99
Pasa bajos						
0.0025	0.78	0.80	0.18	0.97	0.11	0.06
0.0050	0.84	0.93	0.27	0.99	0.11	0.21
0.0100	0.91	0.98	0.33	0.99	0.13	0.47
0.0500	0.95	0.99	0.30	0.99	0.21	0.94
Pasa alto						
0.0025	0.90	0.01	0.92	0.01	0.11	0.06
0.0050	0.93	0.00	0.96	0.03	0.11	0.21
0.0100	0.96	0.03	0.98	0.06	0.13	0.47
0.0500	0.99	0.15	0.99	0.22	0.21	0.94
Pasa banda						
0.0025	0.64	0.88	0.96	0.28	0.94	0.04
0.0050	0.71	0.96	0.99	0.40	0.95	0.13
0.0100	0.81	0.99	0.99	0.53	0.97	0.25
0.0500	0.93	0.99	0.99	0.78	0.99	0.53
Ruido gaussiano						
0.0025	0.07	0.31	0.11	0.28	0.02	0.27
0.0050	0.07	0.48	0.09	0.42	0.01	0.47
0.0100	0.05	0.72	0.07	0.64	0.00	0.72
0.0500	0.03	0.98	0.03	0.97	0.11	0.98

Por otro lado, continuando con lo mencionado sobre la importancia o uso que puede llegar a tener cada método, dependerá y estará vinculado al tipo de ataque que reciba la señal con marca de agua. Esto, dado que la marca de agua es sintonizada en un rango de frecuencias, y si es atacada por un filtro, puede que la misma esté siendo eliminada y la extracción y/o detección sea más difícil.

Lo mencionado, se observa con los filtros, dado que se dependiendo de la frecuencia de corte de los mismos, la extracción de la marca de agua es más o menos similar a la original. A continuación, se observa la marca de agua extraída, luego de un ataque mediante un filtro pasa banda, variando las frecuencias de corte superior e inferior.

Las figuras corresponden a la señal canto con un  $\alpha = 0,005$ , sometidas a un filtro pasa banda. Primer pasa banda (izquierda),  $f_{inferior} = 100Hz$  y  $f_{superior} = 4000Hz$ . Segundo pasa banda (medio),  $f_{inferior} = 4000Hz$  y  $f_{superior} = 8000Hz$ . Tercer pasa banda (derecha),  $f_{inferior} = 8000Hz$  y  $f_{superior} = 12000Hz$ .

Figura 13 y Figura 14 corresponden a método 1 y 2, respectivamente.

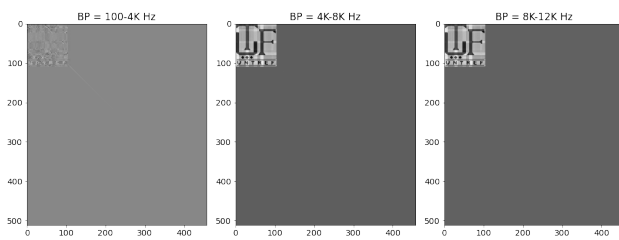


Figura 13: Extracciones de marca de agua, ataque filtro pasa banda. Señal canto, método 1. Izquierda: BP = 100-4K Hz. Centro: BP = 4K-8K Hz. Derecha: BP = 8K-12K Hz.

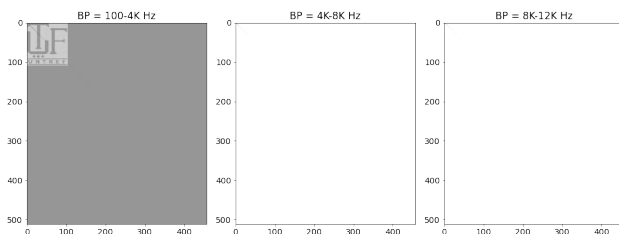


Figura 14: Extracciones de marca de agua, ataque filtro pasa banda. Señal canto, método 2. Izquierda: BP = 100-4K Hz. Centro: BP = 4K-8K Hz. Derecha: BP = 8K-12K Hz.

Todo lo mencionado se corrobora en las figuras presentadas. La diferencia principal radica en la escala de cada método de extracción, dado que si se hace una correlación cruzada y se compara en porcentajes las matrices extraídas con las originales, en un método son prácticamente iguales dado a la similitud de su escala, contrario a lo que pasa en este caso, con el método 2, que se tratan de matrices de valores diferentes, y por lo tanto, la extracción es casi nula.

## 5. CONCLUSIONES

Las marcas de agua son una de las estrategias mas usadas a la hora de hablar de protección de derechos de autor, de autenticación de datos e incluso hasta de comunicaciones encubiertas. En este trabajo se presentan dos métodos diferentes tanto de incrustado como de extracción de las mismas, basadas en métodos que habitan en el dominio frecuencial. Ambos utilizan la STFT para descomponer las señales de audio y también la descomposición de valores singulares para deconstruir y recons-

truir las diferentes matrices, escogiendo un factor de fuerza dependiendo del tipo de marca de agua que se quiera obtener.

Más allá de la diferencia en cada método, los resultados obtenidos de las simulaciones denotan una relación dependiente no solo del factor de fuerza utilizado en cada caso, sino también del objeto de prueba, de las características del tipo de ataque que se esté empleando y de las características del método frecuencial, como la ventana. Como se menciona, si bien un método extrae la marca de agua más fielmente, ambos son de utilidad y dependerá la aplicación en la cual se trabaje.

En futuros estudios se plantea partir de otros métodos que también involucren el dominio frecuencial, como ser la transformada de Wavelet (DWT), o la transformada de coseno (CT), probar una mayor variedad de ataques que involucren por ejemplo ataques de tipo geométricos o de desincronización (jitter, modificación de la escala temporal, entre otros), y probar distintos tipos de señales, y no solo de corta duración, sino evaluar el rendimiento de los algoritmos propuestos en canciones de larga duración, pensando en lo que sería una distribución en la vida real.

## REFERENCIAS

- [1] *An SVD-Based Audio Watermarking Technique* - HAMZA ÖZER, BÜLENT SANKUR, NASIR MEMON - MMSec '05: Proceedings of the 7th workshop on Multimedia and security August 2005, pp. 51-56.
- [2] *Hybrid SVD-Based Image Watermarking Schemes: A Review* - WAFAT HAMDAN ALSHOURA, ZURINAHNI ZAINOL, JE SEN TEH, MOATSUM ALAWIDA, ABDULATIF ALABDULATIF - IEEE Acces, 22 February 2021, pp. 32931 - 32968.
- [3] *An imperceptible and robust audio watermarking algorithm* - ALI AL-HAJ - EURASIP Journal on Audio, Speech, and Music Processing 2014.
- [4] *Imperceptible and Robust DWT-SVD-Based Digital Audio Watermarking Algorithm* - KHALID A. DARABKH - Journal of Software Engineering and Applications, 2014.
- [5] STEFAN KATZENBEISSER, FABIEN A. PETITCOLAS (1999). *Information Hiding Techniques for Steganography and Digital Watermarking*. Artech House. Pp. 109-113.