

MARCA DE AGUA EN SEÑALES DE AUDIO

CAMILO BAUTISTA ¹ , LUCAS RUIZ ², SOL DIMARCO ³

¹ Universidad Nacional de Tres de Febrero, Procesamiento Digital de Señales, Buenos Aires,
Argentina.

camilobautista00@gmail.com

² Universidad Nacional de Tres de Febrero, Procesamiento Digital de Señales, Buenos Aires,
Argentina.

ruizlucas2211@gmail.com

³ Universidad Nacional de Tres de Febrero, Procesamiento Digital de Señales, Buenos Aires,
Argentina.

sollinti@gmail.com

RESUMEN – Se realiza el estudio e implementación de un algoritmo que permite la incorporación de una marca de agua a una señal de audio y su posterior recuperación. Se implementa la Transformada de Tiempo Corto de Fourier (STFT) y el algoritmo de Descomposición de Valor Singular (SVD) para realizar el procesamiento de señal. Posteriormente, se evalúan los resultados obtenidos a partir de parámetros cuantitativos y perceptivos y se realizan conclusiones.

ABSTRACT– The study and implementation of an algorithm that allows the incorporation of a watermark to an audio signal and its subsequent recovery is carried out. The Short Time Fourier Transform (STFT) and the Singular Value Decomposition (SVD) algorithm are implemented to process the signal. Then, the results obtained from quantitative and perceptive parameters are evaluated and conclusions are drawn.

1. INTRODUCCIÓN

La marca de agua es una técnica importante tanto para proteger información multimedia electrónica como para la propiedad intelectual. Suele tener distintas aplicaciones entre las cuales se encuentra la protección de propiedad y la autenticación. El campo de investigación al que aplica este documento es al procesamiento digital de señales para aplicación de marcas de agua en audio y su

correspondiente extracción. En audio, la marca de agua debe cumplir determinadas características: la imperceptibilidad para el oído humano, la resistencia ante posibles manipulaciones o ataques y la imposibilidad de remover la marca de agua sin conocer el algoritmo que la incrusta.

Este documento tiene como principal objetivo plasmar brevemente la metodología de procesamiento y aplicación de marcas de agua en señales de audio a la vez que detallar los

resultados obtenidos a partir del uso de parámetros cuantitativos y perceptivos, tales como la relación entre pico-señal y ruido (PSNR), la correlación normalizada (NC) y la evaluación de distintos valores para el factor de fuerza (α) de la marca de agua [2].

Este documento sigue la siguiente estructura. En primer lugar se mencionan los trabajos de investigación relacionados en los que este documento se apoya, luego, se detallan los conceptos básicos teóricos necesarios para llevar a cabo el procesamiento de señales. A continuación, se detalla el desarrollo experimental en subsecciones y los resultados obtenidos de tal desarrollo. Para finalizar, se presentan las conclusiones obtenidas de la evaluación de resultados.

2. CONCEPTOS BÁSICOS

La marca de agua de audio tiene aplicaciones en diversas áreas, como la protección de derechos de autor, autenticación de datos, comunicaciones encubiertas, adición de metadatos, identificación de contenido y subtítulos o etiquetado de datos. Estas diversas aplicaciones tienen diferentes requisitos de robustez, capacidad de datos e imperceptibilidad. [1]

Se han propuesto varios esquemas de marca de agua digital para los diferentes propósitos mencionados anteriormente. Se los clasifica en Componentes de extracción, Dominio de trabajo, Percepción humana y el Tipo de documento relacionado con el objeto anfitrión como se muestra en la Figura 1.

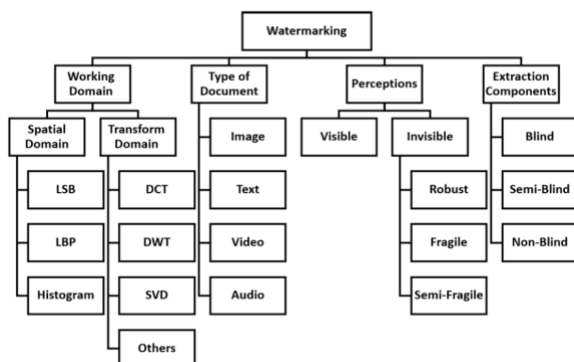


Figura 1. Clasificación de las estrategias de marca de agua digital[2].

- **Componentes de extracción.**

La estructura de marca de agua contiene dos procesos, el embebido y la extracción. Éste último requiere de una información adicional tal como una llave, la marca de audio original y la imagen del anfitrión original sin la marca de agua. Se lo puede clasificar en tres tipos basados en el tipo de información que se requiera respectivamente para su extracción[2]:

- **Blind:** este esquema sólo requiere la imagen y la llave de la marca de agua para la extracción.
- **Non-Blind:** este esquema requiere al menos un componente adicional (la imagen original del anfitrión o de la marca de agua).
- **Semi-Blind:** En este esquema se requiere la llave y también la marca de agua original.[2]

- **Dominio de trabajo**

- **Espacial:**

En este dominio de trabajo la marca de agua es embebida directamente en los píxeles de la imagen anfitrión, esto tiene ventajas destacables como ser un esquema rápido, simple y soporta una gran carga útil.[2]

- **Frecuencial:**

En este dominio de trabajo se utilizan transformadas, como por ejemplo la Transformada discreta de Fourier, la cual se le aplica en la imagen anfitrión para generar los coeficientes de frecuencia. Se los modifica embebiendo la información de la marca de agua, para así luego, obtener la imagen de la marca de agua mediante la transformada inversa. Los pequeños cambios en los coeficientes de frecuencia no son perceptibles para el

sistema auditivo humano. Otra ventaja de trabajar en este dominio es la robustez ante ataques geométricos y no geométricos.[2]

- **SVD (Descomposición en valores singulares)**

La descomposición en valores singulares se utiliza como una herramienta numérica que permite descomponer una matriz en dos matrices ortogonales y sus valores singulares. De esta forma, una matriz A se descompone en $A = U D V^T$.

Siendo A la matriz de F filas y M columnas que se desea resumir. D será una matriz diagonal con las mismas dimensiones que A , cuyos elementos diagonales serán los valores singulares de la matriz A . U será una matriz ortogonal de F filas y F columnas y V será una matriz ortogonal de M filas y M columnas. [1]

Esta herramienta tiene aplicaciones en estadística y otras áreas. En particular, se la suele utilizar para la compresión de imágenes digitales de modo que puedan ser transmitidas a través de distintos medios electrónicos.[2]

- **Percepción humana**

La marca de agua puede ser perceptible o no ante el sistema visual y auditivo humano. Se busca lograr que la marca de agua sea imperceptible, donde se clasifican 4 métodos:

- **Robusto:** Se utiliza para la protección de los derechos de autor y la prueba de propiedad. Con el objetivo de que el esquema sea robusto se utilizan diferentes métodos para embeber la marca de agua. Para prevenir la degradación, manipulación y eliminación de la marca de agua, la marca de agua embebida debe soportar numerosos ataques geométricos y no geométricos.[2]
- **Frágil:** Embebe una marca de agua que es fácilmente

destruible o alterable por mínimas modificaciones.[2]

- **Semi-frágil:** En este método, la marca de audio embebida es modificada hasta un límite, no obstante, la marca de agua sigue siendo vulnerable de ataques maliciosos, produciendo la destrucción de la misma.[2]
- **Híbrido:** Es una combinación de los métodos Frágil y Robusto con las aplicaciones de ambos.[2]

3. DESARROLLO EXPERIMENTAL

3.1 Agregado de la marca de agua

En primer lugar, se aplica la STFT a la señal a la cual se quiere agregar la marca de agua, la cual es procesada en segmentos solapados (Overlap). Cada cuadro de la STFT es ventaneado para reducir los artefactos de los límites. La DTFT de cada cuadro está representada en una matriz de tamaño $F \times M$. (F es la cantidad de cuadros, depende del largo de la señal; M el tamaño del cuadro).

A esta matriz STFT se la somete al proceso de SVD (Singular Value Decomposition). De esta manera se obtiene la matriz A como el producto de dos matrices ortogonales, constituidas por autovectores, y una matriz diagonal que contiene los valores singulares.

$$A = U D V^T$$

Luego se procede al incrustado de la marca de agua en la matriz D . La marca de agua, figura 2, es una imagen que se debe convertir a blanco y negro, es decir, una matriz binaria.



Figura 2. Marca de agua elegida.

El incrustado de la marca de agua puede describirse en la siguiente fórmula,

$$W_D = D + \alpha * W,$$

donde W es la matriz marca de agua, y α es el peso de la misma, cuyo valor se define según cuestiones perceptuales [1]. Cabe recalcar que esta es la manera que se eligió arbitrariamente para el incrustado de la marca de agua, ya que existen otros métodos que implican el agregado de valores sólo en ciertos elementos de la matriz D . Para que este cálculo matricial sea posible se debe contar con cierta coherencia en las dimensiones de las matrices. Para ello se convierten D y W en vectores, y se trunca al vector W para que tenga la misma longitud que el vector D (notar que es necesario que la matriz marca de agua tenga mayor dimensión que la matriz STFT del audio). Luego, se reconstruyen las matrices con las dimensiones originales de D , y se obtiene W_D .

A la matriz con marca de agua W_D resultante se la somete a SVD, correspondiente a la siguiente ecuación.

$$W_D = U_W D_W V_W^T.$$

Con esta nueva matriz D_W se procede a obtener la matriz original con la marca de agua incrustada, mediante el proceso de SVD inverso, es decir, utilizando las matrices de autovectores originales U y V .

$$A_W = U D_W V^T$$

Las matrices U_W y V_W son conservadas para la detección posterior de la marca de agua.

El último paso es aplicar la STFT inversa a la matriz A_W para obtener una señal de audio.

3.2 Detección de la marca de agua

Para la detección de la marca de agua, el receptor debe contar con una señal recibida que contenga la marca de agua, y las matrices U_W y V_W . Primero se debe calcular la STFT a la señal de audio, para luego aplicar SVD y obtener una matriz diagonal D' y dos matrices de autovalores.

Segundo, se realiza la multiplicación matricial de D' con las matrices U_W y V_W , es decir, SVD inverso, y se obtiene una matriz S_W . De esta matriz resultante se debe extraer la marca de agua, mediante algún criterio. Se ha detectado de manera empírica que los elementos de la matriz en los cuales habría un 1 de la marca de agua presentan un valor notablemente mayor que los elementos en los cuales habría un 0 de la marca de agua [3]. Entonces, se procede a calcular el promedio de todos los elementos de S_W , y se construye una matriz marca de agua W_n como lo indica la fórmula siguiente,

$$W_n = \{0 \text{ si } S_{W_{i,j}} \leq \text{avg}(S_W); 1 \text{ en otro caso}\}.$$

3.3 Parámetros evaluativos

Este informe está enfocado a señales de audio, por lo cual la percepción de diferencias entre el audio original y el que contiene la marca de agua, mediante la audición, es la prueba principal para analizar resultados. No obstante, es necesario contar con parámetros objetivos para desvincular la percepción humana del análisis. Para ello, se introduce el parámetro Peak signal-to-noise ratio (PSNR),

$$PSNR = 10 \log \left[\frac{\max(I(i,j))^2}{MSE} \right]$$

con $\max(I(i,j))$ mayor elemento de la STFT del audio original, y MSE el error cuadrático medio entre la STFT del audio original ($I(i,j)$) y la STFT del audio con marca de agua incrustada ($I^W(i,j)$), calculado como

$$MSE = \frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N [I(i,j) - I^W(i,j)]^2,$$

con M y N los números de filas y columnas de tales matrices. Si el valor de PSNR es elevado, indica que hay una diferencia mínima entre las matrices, es decir, que la marca de agua no es distinguible notablemente.

Otro parámetro utilizado es Normalized Correlation (NC),

$NC(W, Wn)$

$$= \frac{\sum_{i=1}^M \sum_{j=1}^N [W(i, j) - \mu_1][Wn(i, j) - \mu_2]}{\sqrt{\sum_{i=1}^M \sum_{j=1}^N [W(i, j) - \mu_1]^2} \sqrt{\sum_{i=1}^M \sum_{j=1}^N [Wn(i, j) - \mu_2]^2}}$$

con W la matriz marca de agua original, Wn la matriz marca de agua extraída por el receptor, μ_1 la media de la matriz W y μ_2 la media de la matriz Wn . NC mide la diferencia entre la marca de agua extraída y la original. Si $NC = 1$, entonces ambas marcas de agua son idénticas.

3.4 Pruebas y ataques

Se realizan distintas pruebas modificando un valor de proporcionalidad α , el cual se lo multiplica a la marca de agua, produciendo así que la marca de agua sea o no perceptible dependiendo del valor de α .

Otra método para analizar la robustez de la marca de agua se generan los siguientes ataques geométricos y no geométricos a la matriz STFT del audio recibido, con la marca de agua:

- **Recorte:** Se reemplazan algunos valores de la matriz por ceros.
- **Corte:** Se reemplazan los valores de algunas filas o columnas de la matriz por ceros.
- **Intercambio:** Se intercambian los valores de algunas filas o columnas por otras.
- **Rotación:** Se rota la matriz STFT del audio con un determinado grado.

Una vez modificada la STFT del audio, se realiza la extracción de la marca de agua para así luego calcular el parámetro NC y verificar su robustez ante estos ataques.

4. RESULTADOS

Tras realizar la correlación normalizada a las señales de audio recibidas con marca de agua, se obtienen los siguientes resultados (Tabla 1).

	Señal de voz	Señal de Música
NC	0,879	0,846

Tabla 1: Correlación normalizada para las señales de audio.

Por otro lado, se presentan los resultados tras evaluar la relación entre pico-señal y ruido (PSNR) a las señales de audio (Tabla 2).

	Señal de voz	Señal de Música
PSNR	52,2 dB	64,3 dB

Tabla 2: Relación entre pico-señal y ruido para las señales de audio.

A continuación, como se observa en la tabla 3, se presentan los resultados de aplicar variaciones al factor de fuerza α , obtenidos a partir del parámetro PSNR.

α	Señal de voz	Señal de música
0,1	31 dB	54,7 dB
1	11,3 dB	20,3 dB
10	11,1 dB	4,9 dB

Tabla 3: Relación entre pico-señal y ruido para las señales de audio modificando el valor de factor de fuerza.

Por último, se presentan en la tabla 4 los resultados evaluados a partir del parámetro de correlación normalizada (NC) para los cuatro tipos de ataques mencionados.

Ataques	Recorte	Corte	Intercambio	Rotación
Señal de voz	0,868	0,866	0,932	0,94
Señal de música	0,815	0,8	0,945	0,944

Tabla 4: Correlación normalizada en las señales de audio para distintos ataques.

5. DISCUSIÓN

Durante el desarrollo de este documento y de un algoritmo para el incrustado y detección de una marca de agua en audio, se observaron una gran cantidad de variaciones con respecto a los métodos utilizados y propuestos. Se optó por una combinación de desarrollos hallados en la bibliografía.

Asimismo, la implementación de ataques se realiza según lo propuesto en [2]. Es importante mencionar que el código desarrollado es un método experimental basado en la recopilación de información de distintas fuentes, lo cual puede traer resultados inesperados e incertezas a la hora del análisis mediante parámetros presentes en estas investigaciones.

Respecto a los resultados de la Tabla 1, se obtuvo un valor de NC relativamente cercano a 1 en ambas pruebas, siendo menor el valor de NC en el caso de la señal musical. Esto podría deberse a las variaciones de contenido energético entre una señal de voz y una señal musical, además del método de extracción de marca de agua propuesto.

Luego, en la Tabla 2 se presentan niveles elevados de PSNR para ambas señales, lo cual indica imperceptibilidad en la marca de agua incrustada. Esto puede deberse al valor de α implementado para esta prueba. En este caso la marca de agua en la señal de música es más imperceptible que en la señal de voz, lo cual es esperable ya que los niveles en el fragmento de la canción se mantienen constantes y elevados, a diferencia de la señal de voz.

En la Tabla 3 se nota una relación indirecta entre el valor de α y el valor de PSNR calculado. Este fenómeno es esperable ya que a mayor factor de fuerza α más perceptible es la marca de agua. Sin embargo, en el código adjuntado a este informe puede escucharse cada prueba con distintos valores de α , y la única diferencia que se puede apreciar es entre los audios originales y cualquier audio con marca de agua. Es decir, entre variaciones de α no se logra escuchar un cambio en la señal. Se deduce que esta diferencia, notable en los parámetros objetivos pero imperceptible a la escucha, se debe a la marca de agua utilizada, que cuenta con una mayor cantidad de ceros que de unos al convertirla en binario. En otras palabras, aumentar el valor de la marca de agua no es un cambio lo suficientemente drástico en la señal de audio resultante como para notar un cambio al reproducirla.

Por último, en la Tabla 4, se observa una reducción en el valor de NC para las señales de audio modificadas con los ataques descritos anteriormente. Esto indicaría una

menor correlación entre la marca de agua original y las marcas de agua extraídas tras los ataques. Sin embargo, este no es el caso para los ataques por intercambio y por rotación, donde los resultados obtenidos demuestran una mayor correlación que los obtenidos en la Tabla 1. Una posible causa de esto puede ser que accidentalmente se hayan realizado los ataques de forma que la marca extraída se asemeje aún más a la marca de agua original. Como se mencionó anteriormente, esto puede ser consecuencia de la elección de la marca de agua y la relación entre cantidades de ceros y unos que ésta presenta. Otra posible causa es el hecho de que el método propuesto para incrustado y detección de la marca de agua es producto de distintas metodologías implementadas en procesos similares pero con transformadas distintas a la STFT.

Un factor importante a tener en cuenta en todo el desarrollo de este informe es la limitación del código propuesto respecto a las señales que puede manipular. Concretamente, las señales de audio a las que se les puede incrustar una marca de agua deben ser de menor dimensión que la marca de agua misma, debido al mecanismo de truncado explicado en la sección 3.1. Por esto, se debe tener en cuenta que si se desea incrustar una marca de agua que no sea truncada, ésta deberá tener exactamente las mismas dimensiones que la STFT del audio que la contendrá.

6. CONCLUSIÓN

Se concluye que el método desarrollado arroja resultados satisfactorios para la incrustación y extracción de marca de agua y su correspondiente robustez ante determinados ataques. Se considera necesario profundizar en el desarrollo y análisis del algoritmo propuesto para mejorar los valores obtenidos a través de parámetros cuantitativos y perceptivos, así como para poder realizar la incrustación de una marca de agua a distintos tipos de señales con distintas dimensiones.

7. REFERENCIAS

[1] Özer H.Sankur B.Memon N. "An SVD-based audio watermarking technique". Proceedings of the 7th Multimedia and Security Workshop 2005, MM and Sec'05 (2006).

[2] Alshoura W. H., Zainol Z., Teh J. S., Alawida M., Alabdulatif A. "Hybrid SVD-Based Image Watermarking Schemes: A Review". IEEE Access (2021).

[3] Imperceptible and Robust DWT-SVD-Based Digital Audio Watermark- ing Algorithm-Khalid A. Darabkh-Journal of Software Engineering and Applications, 2014, 7, 859-871.