Esquema Robusto ante Ataques de Desincronización para Marcas de Agua en Audio

Por

María Alejandra Menéndez Ortiz

Tesis sometida como requisito parcial para obtener
el grado de

MAESTRA EN CIENCIAS EN LA ESPECIALIDAD DE CIENCIAS COMPUTACIONALES

en el

Instituto Nacional de Astrofísica, Óptica y Electrónica

Tonantzintla, Puebla

Supervisada por:

Dr. René Armando Cumplido Parra
Investigador del INAOE
Dra. Claudia Feregrino Uribe
Investigadora del INAOE

Agradecimientos

Agradezo profundamente a mi familia, por su apoyo incondicional y sus constantes enseñanzas. A mis padres, Angélica Ortiz Cabrera y Miguel Ángel Menéndez López, mis abuelas, mis hermanos y mis sobrinos.

En estas simples líneas, quiero reconocer el invaluable apoyo y dirección de los doctores Claudia Feregrino Uribe y René Cumplido Parra, quienes con sus sabios consejos me ayudaron a concluir este trabajo. Asimismo, agradezco a mis sinodales: Dr. Miguel Arias, Dr. Carlos García y Dr. Manuel Montes, por sus amables comentarios.

A mis amigos, por su tiempo, comprensión y compañía; ha sido un honor compartir esta amistad con ustedes.

Al INAOE, por haberme otorgado la oportunidad de realizar mis estudios de posgrado en esta reconocida institución. De igual manera, reconozco la excelente labor de todos los trabajadores de esta organización.

Agradezo al pueblo de México, que a través del CONACYT me fue otorgada la beca No. 243961, con la cual tuve la oportunidad de realizar este posgrado.

Agradezco al CONCYTEP por el apoyo otorgado a través del Programa de Becas-Tesis CONCYTEP 2012, el cual me ayudó en la culminación de este proyecto de tesis.

Dedicatoria

Así como mi vida, trabajo y esfuerzo, dedico este trabajo a mi Creador.

Con gran amor, respeto y admiración, para mis padres,
abuelas, hermanos y sobrinos.

Resumen

Las marcas de agua digitales son una técnica que ayuda a la protección de derechos de autor, agregando información del propietario de una obra o el comprador de la misma. El objetivo fundamental de las marcas de agua es insertar datos adicionales en un medio electrónico (imágenes, audio, video) de tal manera que no puedan ser percibidos por un observador humano.

En este trabajo de investigación se desarrolla un esquema de marcas de agua para audio que presenta gran robustez ante ataques de desincronización. Este tipo de ataques son de gran interés en el área, porque causan que la mayoría de esquemas actuales de marcas de agua para audio fallen. La estrategia del esquema propuesto para combatir estos ataques es tomar las bajas frecuencias del dominio DWT para construir un histograma, del cual se toman grupos de bins adyacentes y realizar la inserción de la marca de agua. La detección de la misma se lleva a cabo reconstruyendo el histograma e interpretando los mismos grupos de bins adyacentes para construir la marca.

Abstract

Digital audio watermarking is a recent technique that assists with copyright protection. This technique inserts aditional data of the owner or the purchaser of a work. The fundamental aim of digital watermarking is to insert aditional data in an electronic file (such as images, audio or video), in a way that a human observer cannot detect its presence.

In this research project we develop an audio watermarking scheme that is robust against desyncronization attacks. These kind of attacks are of great interest in the area, because they make most of the recent watermarking schemes to fail. The strategy used in the proposed scheme agains these attacks is to take low-frequency DWT coefficients in order to construct a histogram; adjacent groups of bins from this histogram are used to insert the watermark. The detection is performed in a similar manner, the histogram is reconstructed and the same groups of adjecent bins are interpreted as the inserted watermark.

Índice general

Índice de figuras

Índice de tablas

Capítulo 1

Introducción

1.1. Descripción del problema

En las últimas décadas, el Internet ha sido un medio de comunicación sumamente importante, ya que a través de esta red se comparte gran cantidad de información. Sin embargo, la creación de sitios de almacenamiento de archivos o las redes punto a punto ha dado paso a la copia, modificación y redistribución ilegal de contenido.

Esta situación ha traído consigo grandes pérdidas económicas para las compañías productoras de contenidos multimedia; particularmente en la industria de la música y el entretenimiento, donde se ha luchado durante años contra la distribución ilegal de sus productos. La Alianza Internacional de Propiedad Intelectual (IIPA, por sus siglas en inglés) ha estimado una pérdida global por concepto de piratería en grabaciones y música de más de $1,000 millones de dólares tan sólo en el año 2009, excluyendo a Europa y Estados Unidos [10].

Para tratar de combatir la piratería se han desarrollado tecnologías, como la criptografía, que ayudan al cumplimiento de las leyes de derechos de autor. No obstante, la criptografía deja desprotegido el material una vez que los usuarios descifran las claves, por lo que ha surgido un nuevo enfoque de seguridad, llamado marcas de agua digitales.

El objetivo de los esquemas de marcas de agua es insertar información en material digital, de tal forma que sea imperceptible a un observador humano pero que sea detectado por un algoritmo de computadora. Una marca de agua es un patrón de información transparente e invisible que se inserta en los datos originales utilizando un algoritmo específico. Dichas marcas son señales que se agregan a medios digitales (como audio, video o imágenes) y éstas pueden detectarse o extraerse posteriormente para hacer alguna afirmación acerca del medio [22]. Por ejemplo, se puede incluir una marca con los datos del propietario de un video o canción.

La imperceptibilidad de la información insertada se logra aprovechando las imperfecciones en los sentidos de los seres humanos, razón por la que éstos se estudian cuidadadosamente. Por lo tanto, las marcas de agua en imágenes y video dependen de las características del Sistema Visual Humano (HVS, por sus siglas en inglés), mientras que las marcas de agua en audio explotan las características del Sistema Auditivo Humano (HAS, por sus siglas en inglés).

Actualmente se han reportado menos trabajos de investigación en marcas de agua en audio, comparados con los trabajos de marcas de agua en imágenes. Esto ocurre porque las marcas de agua en audio representan un reto mayor que las

marcas de agua en imágenes, dado que el Sistema Auditivo Humano es significativamente más sensible que el Sistema Visual Humano. Por lo tanto, es difícil diseñar esquemas de marcas de agua en audio que sean eficaces [2].

Además de la imperceptibilidad de los datos insertados, otro de los objetivos de las marcas de agua es la robustez de la marca. La robustez se refiere a la resistencia que tiene la información insertada ante ataques. Un ataque puede describirse como cualquier procesamiento que evite el propósito de la marca [1].

Pérez-Meana [19] propone una clasificación de ataques para esquemas de marcas de agua en audio (Figura 1.1). De esta clasificación, los ataques geométricos son de interés especial, pues no eliminan la marca sino que destruyen la sincronización entre el detector y la marca. La sincronización es importante para la detección de marcas, especialmente cuando el audio es atacado, debido a que la mayoría de los esquemas de marcas de agua para audio están basados en posiciones; es decir, las marcas se insertan en lugares específicos del audio y la detección se logra a partir de los mismos. Por lo tanto, los ataques que cambian las posiciones de las marcas hacen que la mayoría de esquemas actuales de marcas de agua para audio fallen [11].

En la actualidad, la mayoría de los esquemas de marcas de agua en audio son robustos ante operaciones comunes de procesamiento de señales, tales como compresión MP3, ruido aditivo, filtros pasa-bajas, entre otros. Sin embargo, estos esquemas tienen graves problemas con los ataques de desincronización. Aunque existen diferentes técnicas para combatir la desincronización, los trabajos actuales aún puede ser mejorados, ya que se enfocan en uno o dos ataques de desincroni-

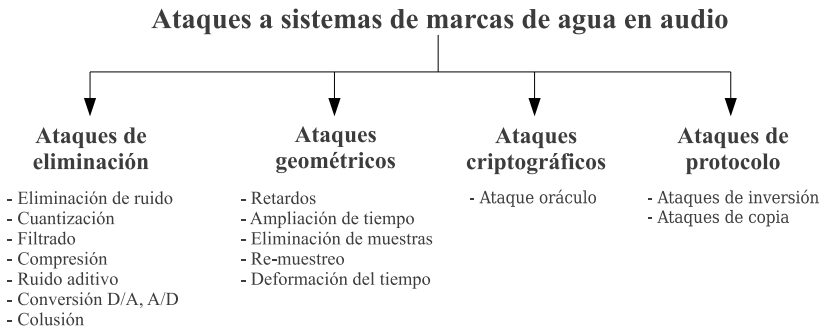


Figura 1.1: Clasificación de ataques en sistemas de marcas de agua en audio.

zación en particular, pero son débiles contra el resto.

A continuación se detallan los objetivos de este trabajo de investigación, con el que se ofrece una solución al problema de desincronización en las marcas de agua para audio.

1.1.1. Objetivos

Objetivo general

Diseñar e implementar un esquema de marcas de agua en audio, robusto ante ataques de desincronización, que mejore el desempeño de los métodos propuestos en la literatura actual.

Objetivos específicos

- Definir una mejora respecto a la robustez de los esquemas de marcas de agua existentes en la literatura actual.
- Desarrollar un esquema de marcas de agua en audio, robusto ante ataques de desincronización que incorpore la mejora propuesta.

1.2. Metodología

Para dar cumplimiento a los objetivos de esta investigación, se plantea una serie de pasos que conforman la metodología propuesta (Figura 1.2). A continuación se describen brevemente cada uno de estos pasos.

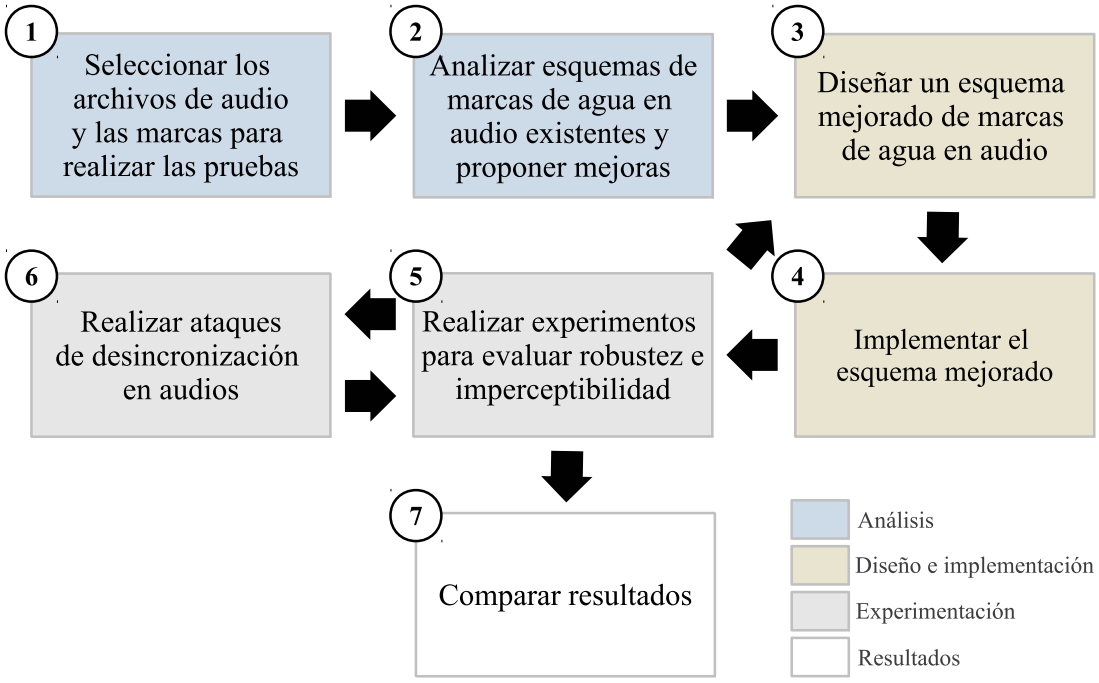


Figura 1.2: Diagrama de la metodología propuesta.

Seleccionar los archivos de audio y las marcas para realizar las pruebas

Como su nombre lo indica, en esta etapa se buscan archivos de audio utilizados por otros autores para insertar las marcas de agua. Se seleccionan audios de distintos géneros musicales para evaluar el desempeño del esquema. Asimismo, se definen las marcas de agua para insertar en los archivos de audio originales.

Analizar esquemas de marcas de agua en audio existentes y proponer mejoras

Durante la revisión de la literatura, se consideran las técnicas utilizadas por los esquemas de marcas de agua en audio que ofrecen una solución al problema de la desincronización. Se estudian los resultados reportados por los diferentes trabajos y se elige el que presente los mejores. Se analizan sus ventajas y desventajas para determinar las posibles mejoras.

Diseñar un esquema mejorado de marcas de agua en audio

En el paso anterior se identifica el esquema de marcas de agua más prometedor de la literatura y se analizan sus características. Con base en ello, se identifican mejoras con respecto a la robustez del esquema, que permitan resistir mejor los ataques de desincronización.

Implementar el esquema mejorado

En este punto se realiza una implementación en Matlab; con ésta, en las siguientes etapas se analiza el desempeño del esquema de marcas de agua propuesto. El esquema consiste en dos algoritmos, uno para la inserción de las marcas de agua y otro para la detección y recuperación de las mismas.

Realizar experimentos para evaluar robustez e imperceptibilidad

En este punto, se realizan dos tipos de experimentos: experimentos con audios no atacados y experimentos con audios atacados. Debido a que se pueden tener diferentes versiones del esquema propuesto, se llevan a cabo experimentos con audios sin atacar cada vez que existen modificaciones en el esquema. De ahí el

ciclo entre esta etapa y el diseño de un esquema mejorado de marcas de agua en audio. El objetivo de esta interacción es refinar la estrategia y obtener los mejores resultados posibles. Una vez que se alcanzan resultados de robustez e imperceptibilidad satisfactorios, se avanza a la siguiente etapa para obtener los audios atacados. Con estos audios, nuevamente se realizan experimentos para evaluar la robustez de las marcas.

Realizar ataques de desincronización en audio

Antes de realizar ataques en los audios, es necesario insertarles una marca a cada uno de ellos. Hecho esto, se prosigue a aplicar los ataques previamente elegidos. Para determinar los ataques que se llevan a cabo, se estudian los ataques presentados en el *benchmark* Stirmark y los ataques presentados en otros esquemas relevantes. Se eligen los ataques de desincronización más significativos, así como algunos ataques de procesamiento de señales.

Comparar resultados

Esta etapa consiste en realizar una comparativa entre los resultados obtenidos con el esquema de marcas de agua propuesto y los resultados reportados en otros trabajos de la literatura. Cabe mencionar que es necesario implementar el esquema base, para poder reproducir los resultados presentados.

En el siguiente capítulo se describen los conceptos más relevantes de las marcas de agua digitales, lo que permitirá comprender mejor este trabajo de investigación.

1.3. Estructura de la tesis

Este trabajo está organizado de la siguiente manera. En el capítulo 2 se presentan las bases teóricas que dan sustento a esta investigación; se presenta una introducción a las marcas de agua digitales, conceptos básicos de marcas de agua para audio y la problemática de sincronización en marcas de agua para audio.

En el capítulo 3 se detallan los esquemas de marcas de agua para audio que ofrecen soluciones al problema de sincronización; asimismo, se examina cómo estos trabajos emplean las estrategias teóricas, además de sus características y funcionamiento.

En el capítulo 4 se describen los esquemas base y propuesto. Se detallan las transformadas *wavelet* que se utilizan en cada esquema, así como los algoritmos de inserción y detección. Además, se presentan las mejoras incluidas en el esquema propuesto, que permiten alcanzar la robustez contra ataques de desincronización.

En el capítulo 5 se presentan los experimentos realizados para probar el esquema propuesto, así como los resultados obtenidos. Además, se explican los conjuntos de datos utilizados para construir el banco de pruebas y las métricas de evaluación usadas. También se presenta una comparativa con otros esquemas de marcas de agua para audio que tratan de combatir el problema de sincronización.

Finalmente, en el capítulo 6 se dan las conclusiones generales del trabajo de investigación; se analizan sus ventajas, desventajas y el trabajo futuro.

Capítulo 2

Fundamento teórico

En este capítulo se describen los aspectos teóricos que sientan la base para este trabajo de investigación. En la primera sección se describen características generales de las marcas de agua digitales, las aplicaciones donde éstas son empleadas y las propiedades que deben cumplir. La siguiente sección describe los fundamentos de las marcas de agua para audio, las técnicas más populares que se utilizan, los ataques para estos medios, así como la forma de evaluación. Finalmente, se describen la sincronización en marcas de agua para audio y las estrategias que pretenden resolver esta problemática.

2.1. Introducción a las marcas de agua digitales

Actualmente es muy común almacenar documentos, imágenes, videos y audio en formatos digitales, y posteriormente compartir estos archivos a través de Internet. Gracias a estas tecnologías, las actividades cotidianas se han vuelto más cómodas de lo que eran antes. Sin embargo, junto con las ventajas que esto ofrece, también prevalecen algunos inconvenientes.

Debido a la naturaleza de la información digital, es fácil crear una cantidad ilimitada de copias fidedignas a partir de los medios digitales orginales, además que estas copias pueden ser modificadas y distribuidas rápidamente a través de Internet [25]. El Internet es un excelente sistema de distribución debido a que las transacciones hechas por este medio no tienen costo, se elimina la necesidad de almacenamiento y control de existencias, y las entregas se realizan casi instantáneamente [3].

Con lo anterior, se vuelve evidente que se necesita proveer protección a los archivos digitales, bien sea demostrando la propiedad de la información o verificando la manipulación de la misma. Con este objetivo, se crearon los sistemas de *Administración de Derechos Digitales* (DRM por sus siglas en inglés), los cuales permiten controlar y restringir el acceso a información multimedia. Los sistemas DRM incluyen cifrado, control de acceso, administración de llaves y control de copias. Una tecnología clave de estos sistemas, que sirve para la identificación de información y el control de copias, son las marcas de agua digitales [7].

Como se mencionaba anteriormente, el cifrado es uno de los métodos utilizados en los sistemas DRM. La información digital es cifrada antes de entregarla y únicamente se proporciona una llave a los clientes que han adquirido legalmente las copias del contenido. El archivo cifrado puede entregarse via Internet, pero será inútil para una persona que no cuente con la llave apropiada. Sin embargo, una vez que el contenido es descifrado, no se puede garantizar su uso posterior; es decir, una persona puede comprar el producto, utilizar la llave de descifrado para obtener una copia desprotegida y distribuir copias ilegales del mismo. Por

lo tanto, se aprecia que el cifrado permite proteger el contenido que está siendo transmitido, pero una vez descifrado, éste queda desprotegido [3]. Dadas estas razones, se necesita una solución alternativa o complementaria al cifrado, que permita proteger el contenido aún después de haber sido descifrado. Las marcas de aguas digitales tienen el potencial de satisfacer esta necesidad, porque insertan información dentro del contenido, donde nunca es removida bajo usos normales [3].

2.1.1. Generalidades de las marcas de agua

El objetivo fundamental de las marcas de agua es insertar información adicional a un archivo digital, también conocido como *medio original*; al medio que ha sido modificado para llevar consigo esta información se le conoce como *medio marcado*. De manera ideal, se espera que no haya una diferencia perceptible entre estos medios [21]. Aunque existen marcas de agua perceptibles (*marcas de agua visibles*), en este documento sólo se discutirán las marcas de agua que no pueden percibirse (*marcas de agua invisibles*).

En general, un esquema de marcas de agua consiste en dos etapas (Fig. 2.1), el *proceso de inserción* y el *proceso de extracción*. En el proceso de inserción se tiene un medio original X, al cual se le insertará un mensaje M, también conocido como *marca de agua* o simplemente *marca*; opcionalmente, en la inserción puede influir una llave K que permitirá darle seguridad al esquema. Luego del proceso de inserción, se tiene un medio marcado Y que como se mencionaba anteriormente, será muy parecido al medio original [9]. El algoritmo de inserción se define como [19]:

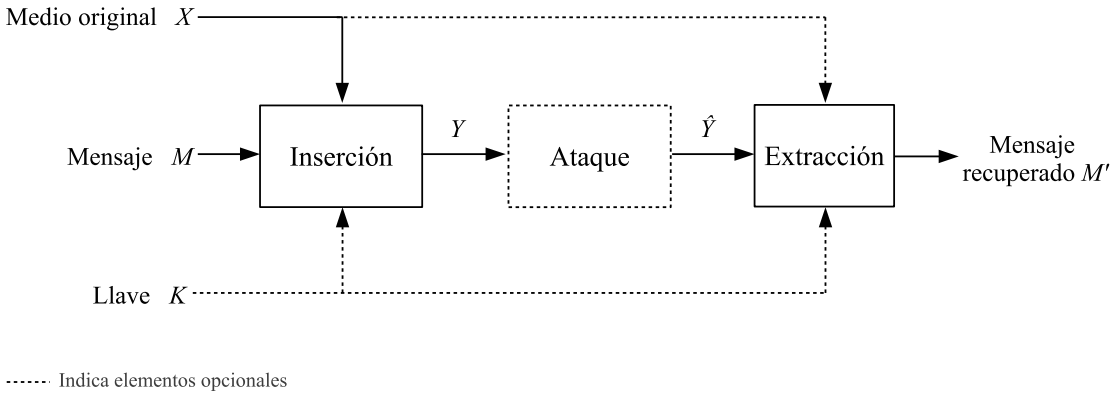


Figura 2.1: Diagrama general de un esquema de marcas de agua.

$$Y = I(X, M, K) \tag{2.1}$$

Posteriormente, el medio marcado será enviado, bien sea a través de Internet o de algún otro sistema de transmisión. Durante esta transferencia, el medio marcado puede sufrir algunas modificaciones o ataques. El *proceso de extracción* permite obtener la marca, a partir del medio $\hat{Y}$. Dependiendo de la seguridad del esquema, se podrá o no tener una llave K. Cabe mencionar que la mayoría de los esquemas de marcas de agua son similares a los sistemas criptográficos simétricos, donde se utiliza una misma llave para la codificación y decodificación [9]. El proceso de extracción puede definirse como [19]:

$$M' = E(\hat{Y}, X, K) \tag{2.2}$$

La extracción a su vez, puede constar de dos procesos, uno de *detección* y otro de *decodificación*. En el proceso de detección se identifica la existencia de una marca en el medio que se está evaluando, pero no se conoce la marca. El proceso de decodificación es el encargado de obtener la marca M' a partir del medio $\hat{Y}$.

Cabe señalar que en la ecuación 2.2 el medio original X podría no ser requerido en ciertos esquemas. Si se utiliza el medio original se trata de un *detector informado*, en caso contrario es un *detector ciego* [8].

2.1.2. Aplicaciones

Las marcas de agua digitales pueden ser utilizadas en múltiples aplicaciones. A continuación se describen las más comunes [3] [9].

Monitoreo de las transmisiones de radiodifusión

Diferentes personas u organizaciones pueden estar interesadas en monitorear lo que se transmite a través de las estaciones de radio y televisión, bien sea para asegurarse que se cumplan con tiempos contratados por un anunciante, que se les paguen las regalías adecuadas por alguna canción o para cerciorarse que no existan re-transmisiones ilegales de sus obras. Las marcas de agua digitales son una alternativa, que permite insertar información a los medios originales previo a su transmisión. Posteriormente, se puede tener un sistema de cómputo que lleve a cabo el monitoreo y detecte las marcas en el contenido transmitido. Esta técnica tiene como ventaja que la información insertada es persistente y compatible con el equipo de radiodifusión, pero tiene la desventaja de que el proceso para incrustar la marca es complejo y se puede degradar la calidad visual o auditiva del medio.

Protección de derechos de autor

Las formas tradicionales para proteger un medio pueden ser costosas o no ser lo suficientemente robustas para un medio transmitido por Internet. Las marcas de agua permiten insertar información de la propiedad intelectual de un medio,

de manera imperceptible y con la robustez necesaria para soportar distintos tipos de procesamiento. Sin embargo, las marcas de agua aún no se han utilizado en corte para probar legalmente la propiedad de un medio.

Control de distribución de copias

Cuando se venden medios digitales a través de Internet, es importante tener el control de las personas que han adquirido un mismo medio. Una estrategia que se utiliza para facilitar esta tarea, se conoce como *fingerprinting* o seguimiento de transacciones. El término *fingerprinting* se utiliza haciendo la analogía con una huella digital humana, que permite identificar de manera única a cada individuo. Algo similar puede realizarse con los medios digitales, insertando en ellos información del comprador y que permite identificar su copia de las demás. Si alguno de los clientes comienza a realizar copias ilegales del contenido que adquirió, gracias a esta estrategia será posible identificar al responsable.

Autenticación de contenido

En algunos casos, puede ocurrir que la persona que distribuye un medio desee que quien lo reciba obtenga un trabajo original, sin que haya sufrido ninguna alteración durante la transmisión. Las marcas de agua frágiles pueden ser utilizadas para cumplir con este objetivo. La característica de las marcas de agua frágiles es que desaparecen con cualquier tipo de ataque o procesamiento, por leves que éstos sean. Si el receptor puede detectar y extraer las marcas de agua frágiles que se insertaron previamente, significa que tiene el medio original y que éste no ha sufrido modificaciones. Si no se pueden extraer esas marcas, significa que el medio ha sido modificado.

Protección contra copias ilegales

En este tipo de aplicación, a diferencia de otras donde el objetivo es verificar si se está usando indebidamente algún medio protegido, se desea prevenir que los usuarios puedan caer en malos usos de los medios protegidos. Por ejemplo, si una persona desea copiar un CD o DVD; se les puede insertar una marca de agua que indique que no pueden ser copiados, con esta marca, se puede controlar el dispositivo de grabación digital para que no lleve a cabo la copia; en el dispositivo de grabación se necesitaría tener un detector, que reconozca la marca que le indique cuándo no debe realizar la copia.

Metadatos adicionales

Existen enfoques para marcas de agua en donde el principal objetivo no es la protección del contenido, sino enriquecer el medio con información adicional, útil para el usuario final. En audio, se puede insertar información de bases de datos con contenidos enriquecidos que permitan obtener datos adicionales para una canción, como autor, álbum, compañía disquera, etc. Otra posible aplicación para los metadatos adicionales, es agregar información a un medio, que permita separar una mezcla de audio transmitido por un medio analógico convencional (transmisiones FM) [20].

2.1.3. Propiedades

Los esquemas de marcas de agua pueden caracterizarse por ciertas propiedades que los definen. La importancia de cada una de estas propiedades es relativa y depende de los requerimientos de la aplicación que utilice el esquema. En esta subsección, primero se describirán las propiedades asociadas con el proceso de

inserción: efectividad, fidelidad y carga útil de la marca; posteriormente, aquellas típicamente asociadas con la detección: detección ciega o informada, probabilidad de error y robustez [3] [9].

Efectividad

La efectividad de un esquema de marcas de agua, es la probabilidad de que la salida del proceso de inserción estará marcada. Se dice que un medio está marcado si el detector encuentra una marca en dicho medio. La efectividad puede determinarse analíticamente o estimarse empíricamente. La estimación empírica puede realizarse insertando una marca en un conjunto grande de audios de prueba, por ejemplo. El porcentaje de audios de salida que resulten en una detección positiva será la probabilidad de efectividad aproximada. Esta aproximación será válida si se tiene un conjunto de prueba suficientemente grande y si los audios tienen la misma distribución que los audios que se esperan en la aplicación.

Fidelidad

La fidelidad de un esquema es la similaridad perceptual entre un medio sin marcar y otro marcado, en el momento que se presentan a un consumidor. El canal de transmisión puede determinar la fidelidad necesaria para las marcas, por ejemplo, un audio que sea distribuido en DVD necesitará una fidelidad mucho mayor que un audio transmitido por una radiodifusora AM. En la sección 2.2.3 se detalla cómo medir esta propiedad.

Carga útil de la marca

La carga útil se refiere al número de bits que una marca codifica dentro de una unidad de tiempo o dentro de un medio. Un esquema que codifica N bits puede

utilizarse para insertar 2^N mensajes. Dependiendo de la aplicación donde se vaya a implementar el sistema, se requerirán distintas cargas útiles. Por ejemplo, una aplicación para música y video podría requerir 4-8 bits, mientras que una de televisión podría necesitar 24 bits. Como se mencionó previamente (2.1.1), el proceso de decodificación puede consistir en dos etapas: determinar si está presente una marca y, en tal caso, identificar cuál de los 2^N mensajes se ha insertado. La decodificación tendría entonces $2^n + 1$ salidas posibles; los 2^N mensajes, además del mensaje "sin marca".

Detección ciega o informada

En aplicaciones para protección de derechos de autor y monitoreo de transmisiones, los algoritmos de detección pueden tener acceso al medio sin marcar para extraer el mensaje, aunque también pueden existir los casos donde únicamente se utilice información derivada del medio original; estos detectores son conocidos como *informados*. En aplicaciones para protección contra copias, no se tiene acceso al medio original al detectar el mensaje, lo que puede ser más complicado; a dichos detectores se les llama *ciegos*. Aquellos esquemas que utilizan detección informada, se conocen como *sistemas de marcas de agua privados*; mientras que los que usan detección ciega son conocidos como *sistemas de marcas de agua públicos*.

Probabilidad de error

En cualquier esquema de marcas de agua, es de suma importancia saber si un medio está marcado o no, por tanto, la probabilidad de error debe ser muy pequeña cuando se detecta una marca. Existen dos tipos de errores; el primero es el *falso negativo* y ocurre cuando el decodificador no encuentra la marca en un medio donde ésta sí estaba presente; el segundo error, conocido como *falso*

positivo, sucede cuando el decodificador detecta una marca en un medio que no la tenía. Las aplicaciones donde un esquema de marcas de agua pueda ser utilizado, dependerán de la probabilidad de ocurrencia de estos errores.

Robustez

La robustez es la habilidad de detectar una marca después de haber aplicado una técnica de procesamiento (ataque no intencional) o un ataque intencional a un medio marcado. Algunos ejemplos de ataques no intencionales para audio son compresión MP3, ruido aditivio, re-muestreo, filtrado, conversión analógico-digital (A/D) y digital-analógico (D/A), entre otros. Asimismo, existen los intentos de remoción intencional de la marca o análisis intencionales para estimar la marca incrustada; esto se logra mediante un ataque de confabulación (o colusión) entre los propietarios de algunos medios marcados. Según la aplicación donde se utilice el esquema, pueden existir algunos donde la robustez sea irrelevante o indeseable, como en las aplicaciones para autenticación donde se emplean marcas de agua frágiles. Sin embargo, en la mayoría de las aplicaciones, no debe existir la forma de remover la marca sin que el medio resulte fuertemente modificado, de manera que éste quede inutilizable debido a la degradación perceptual. En la sección 2.2.3 se detalla cómo medir esta propiedad.

En la siguiente sección se detallan aspectos relativos a las marcas de agua para audio. Aunque éstas comparten las mismas propiedades hasta ahora descritas, las marcas de agua para audio son de interés especial en este trabajo de investigación.

2.2. Marcas de agua para audio

Las marcas de agua digitales se originaron relativamente hace poco tiempo y los primeros esquemas se enfocaron en desarrollar técnicas que trabajaran con imágenes y videos, fue poco tiempo después que surgieron los esquemas de marcas de agua para audio. Desde entonces, se han desarrollado algoritmos para insertar y extraer marcas en secuencias de audio que utilizan diferentes técnicas y pretenden alcanzar distintos objetivos. Un factor determinante para diseñar esquemas de marcas de agua para audio es el Sistema Auditivo Humano (HAS, por sus siglas en inglés); éste puede presentar tanto ventajas como desventajas para los esquemas, dependiendo de las metas que se deseen alcanzar [4].

Comparado con el Sistema Visual Humano (HVS, por sus siglas en inglés), el HAS es mucho más sensible. Algunos esquemas aprovechan las propiedades de este último para insertar marcas en una señal original, de forma que sea perceptualmente transparente. Sin embargo, insertar esta información adicional es una tarea más compleja que en imágenes o videos, dado que es más fácil que una persona detecte modificaciones realizadas en un audio [8].

2.2.1. Técnicas de marcas de agua para audio

Como se mencionaba, existen esquemas de marcas de agua para audio con distintos objetivos. Por lo tanto, hay diferentes técnicas que se utilizan para intentar alcanzarlos. A continuación se describen brevemente las técnicas de marcas de agua para audio más comunes [8].

Codificación LSB

Es una de las primeras técnicas para ocultar información en audio. Consiste en codificar o reemplazar el bit menos significativo (LSB, por sus siglas en inglés) de una muestra de audio. En su implementación más simple, el bit menos significativo de la señal original es reemplazado por el bit de la marca que se desea ocultar. En un escenario más seguro, el proceso de inserción utiliza una llave secreta para elegir un conjunto pseudo-aleatorio de muestras. El reemplazo de los bits se realiza únicamente en ese conjunto de muestras elegidas; en el proceso de extracción se utiliza esa misma llave para encontrar los bits ocultos en los audios recibidos.

La ventaja de la codificación LSB es su gran capacidad de inserción. Por ejemplo, si se utilizan audios con calidad de CD (frecuencia de muestreo de 44.1 kHz y 16 bits por muestra), se pueden ocultar 44,100 bits por segundo (bps); algunos esquemas incluso utilizan los últimos 3 ó 4 bits menos significativos, con lo que logran capacidades de 132.3 o 176.4 kbps. Otra ventaja de esta estrategia es su sencillez de implementación, tanto la inserción como la decodificación requieren un costo computacional muy bajo, lo que permite llevar estos esquemas a aplicaciones en tiempo real.

No obstante, esta técnica tiene algunas desventajas. El reemplazo aleatorio de muestras genera ruido blanco Gaussiano aditivo (AWGN, por sus siglas en inglés), este tipo de ruido es muy perceptible para el Sistema Auditivo Humano y por lo tanto, se genera una molesta distorsión audible. Otra desventaja es su poca robustez; ataques simples, como recortes aleatorios o intercambios de muestras destruyen la marca insertada. Asimismo, la profundidad para los bits menos sig-

nificativos a utilizar está limitada; considerando un audio con 16 bits por muestra, únicamente se pueden utilizar 4 para realizar la inserción, si se desea mantener una distorsión audible mínima.

Patch Work

La técnica conocida como *Patch Work* originalmente se desarrolló para marcar imágenes, pero posteriormente también fue utilizada para audio. El algoritmo utiliza una hipótesis estadística en dos conjuntos con una gran cantidad de muestras, que sirven para ocultar la información. Este método es apropiado para marcas de agua en audio, ya que el audio tiene una gran cantidad de muestras. En un escenario simple de codificación, se utiliza una llave secreta para seleccionar de manera aleatoria muestras que se dividirán en dos conjuntos; a estos conjuntos se les conoce como "parches".

La amplitud de cada conjunto se modifica ligeramente en sentido inverso; es decir, la amplitud de un conjunto se incrementa una pequeña cantidad d de veces, mientras que el otro conjunto se decrementa la misma cantidad de veces. El valor de d se elige cuidadosamente, procurando que no sea demasiado pequeño, para que sea robusto ante posible ruido en la transmisión; y que no sea demasiado grande como para introducir una distorsión audible significativa. Esto se puede ilustrar de la siguiente manera:

$$\begin{aligned} {a_i}^w &= a_i + d \\ {b_i}^w &= b_i - d \end{aligned} \tag{2.3}$$

Donde a_i y b_i son la i-ésima muestra de los conjuntos aleatorios A y B, respectivamente. ${a_i}^w$ y ${b_i}^w$ son las muestras después de haber sido modificadas. En el

decodificador, se utiliza la misma llave aleatoria para elegir los mismos conjuntos que en la codificación. La expectativa de la diferencia entre esos dos conjuntos de datos se calcula con la fórmula 2.4, si es igual a *2d*, se considera que la señal está marcada. El proceso puede describirse como:

$$E(A-B) = \frac{1}{N}\sum_{i=1}^{N}(a_i{}^w - b_i{}^w) = 2d + \frac{1}{N}\sum_{i=1}^{N}(a_i - b_i) \tag{2.4}$$

Sin embargo, debido a que los datos de los conjuntos se seleccionaron de forma aleatoria, se espera que la última parte de la ecuación sea cero, entonces:

$$E(A-B) = 2d \tag{2.5}$$

El problema con la técnica de *Patch Work* es que en una aplicación real, la diferencia promedio entre dos conjuntos de datos seleccionados aleatoriamente no siempre es cero. Aunque la distribución de la diferencia promedio de esos "parches" marcados está recorrida *2d* veces hacia la derecha de la versión sin marcar, aún existe un traslape entre estas dos distribuciones. Por lo tanto, existe la probabilidad de una detección errónea. Si se incrementara el valor de la modificación d, esto haría que la detección fuera más precisa; sin embargo, existiría el riesgo de introducir más distorsión auditiva.

Modulación del índice de cuantización

La modulación del índice de cuantización (QIM, por sus siglas en inglés) es otra técnica popular para marcas de agua en audio, que oculta información al cuantizar muestras. Se encuentra la muestra con valor máximo del audio original, con ésta se determina el paso de cuantización d. Posteriormente, cada muestra es cuantizada con el paso de cuantización determinado, modificando la muestra

ligeramente, dependiendo del bit que se desee ocultar.

Una implementación simple de la técnica QIM sería la siguiente. Supóngase que el audio original es x, el paso de cuantización es d, la función de cuantización es $q(x, d)$, el bit de la marca a ser ocultado es w (que puede valer 0 ó 1), entonces, la muestra marcada y se denotaría de la siguiente manera:

$$y = q(x, d) + \frac{d}{4} * (2 * w - 1) \tag{2.6}$$

La función de cuantización se define como:

$$q(x, d) = \left[\frac{x}{d}\right] * d \tag{2.7}$$

donde $[x]$ es una función que redondea hacia el entero más cercano a x.

En la figura 2.2, la muestra x primero se cuantiza hacia $q(x, d)$ o el círculo negro. Si el bit a insertar es 1, entonces se añade $\frac{d}{4}$ a la muestra cuantizada, lo que mueve la muestra hacia el círculo blanco. De lo contrario, se resta $\frac{d}{4}$ a la muestra cuantizada, lo que mueve la muestra hacia la cruz (x).

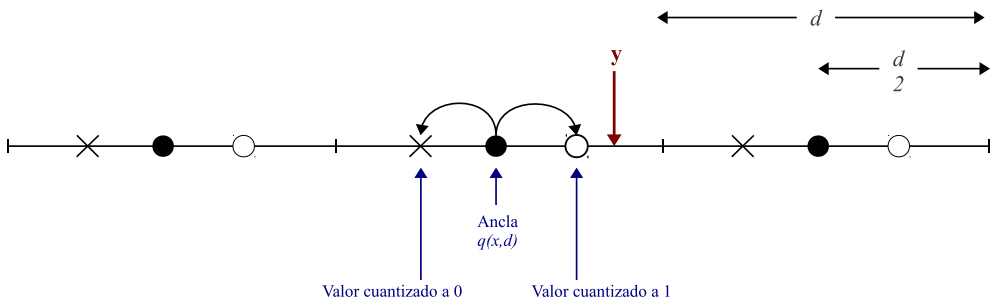


Figura 2.2: Inserción para el esquema QIM.

En el proceso de decodificación, se calcula la diferencia entre la muestra recibida y el valor cuantizado. Si está entre $(0, \frac{d}{4})$, entonces el bit de la marca extraído es "1". Si la diferencia está entre $(-\frac{d}{4}, 0)$, entonces el bit de la marca es "0". Para cualquier otro caso, se dice que la señal recibida no está marcada. Suponiendo que la señal recibida es y, el bit de la marca incrustada es w y el paso de cuantización es d, el proceso de detección puede definirse como:

$$\begin{aligned} w = 1, \quad \text{if} \quad & 0 < y - q(y,d) \quad \leq \tfrac{d}{4} \\ w = 0, \quad \text{if} \quad & -\tfrac{d}{4} \leq y - q(y,d) \quad < 0 \end{aligned} \tag{2.8}$$

La técnica QIM es muy sencilla de implementar y es robusta ante ciertos ataques de adición de ruido. Siempre y cuando el ruido agregado a un canal de transmisión sea inferior a $\frac{d}{4}$, se puede detectar correctamente la marca. No obstante, si el ruido excede este valor, la marca podría no ser detectada con precisión. Por lo tanto, existe un compromiso entre robustez y transparencia; un paso de cuantización d mayor permitirá tener una marca más robusta, con el riesgo de crear una distorsión audible en la señal.

Ocultamiento de eco

Esta técnica inserta información a un audio original introduciendo un eco. La compensación (o retraso) entre el audio original y el eco es tan pequeño que el eco se percibe como una resonancia. Existen cuatro parámetros para este método: la amplitud inicial, la tasa de decaimiento, la compensación "uno" y la compensación "cero". En el proceso de inserción, ilustrado en la figura 2.3, la señal original se mezcla con el kernel "uno" o "cero", dependiendo del valor del bit de la marca que está siendo insertado; con esta mezcla se obtiene la señal marcada. En la figura 2.4 se muestra un ejemplo de cómo un eco es añadido a la señal original.

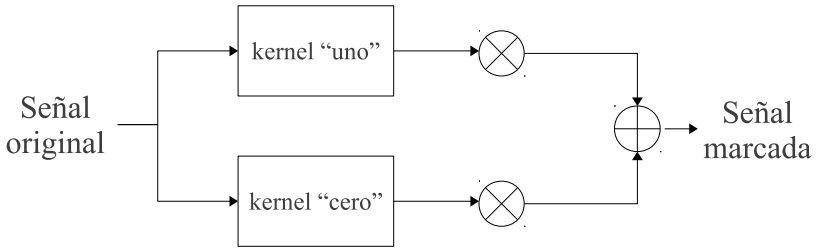


Figura 2.3: Inserción para el esquema de ocultamiento de eco.

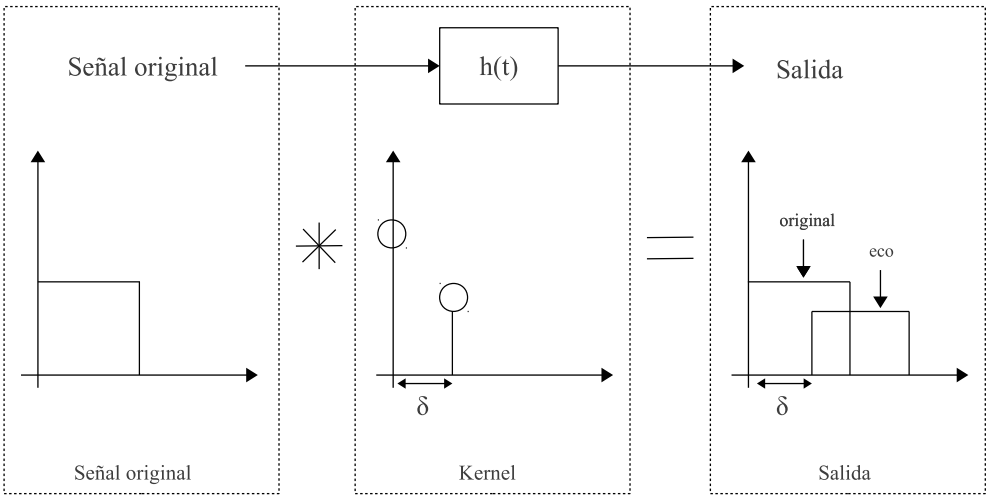


Figura 2.4: Ejemplo de la inserción de eco.

En el proceso de decodificación, se detecta el espacio entre el eco y la señal original y la marca insertada se extrae de acuerdo al espacio. Para hacer esto, se necesita examinar la magnitud de la autocorrelación del *cepstrum* (la transformada de Fourier del logaritmo del espectro de la señal) de la señal insertada en dos posiciones.

$$F^{-1}(ln_{complex}(F(x))^2) \tag{2.9}$$

donde F representa la Transformada de Fourier y F^{-1} la Transformada de Fourier Inversa. En cada segmento, el pico del auto-*cepstrum* detecta la información insertada. Como se mencionaba anteriormente, la naturaleza del eco añadido es la de una resonancia en la señal original, sin llegar a añadir ruido. Es posible insertar la información adicional dentro del audio, conservando sus mismas características estadísticas y perceptuales.

Aunque el ocultamiento de eco provee muchos beneficios, entre ellos robustez, imperceptibilidad y sencillez en los procesos de inserción y decodificación, para lograr una marca de agua robusta es necesario insertar un eco con mucha energía, lo que provoca una distorsión audible. Por lo tanto, con esta técnica también existe un compromiso entre inaudibilidad y robustez.

Codificación de fase

Aunque el Sistema Auditivo Humano es mucho más sensible que el Sistema Visual Humano, es menos sensible a los componentes de fase del audio. [1]. Por lo tanto, es posible insertar marcas en el dominio de la fase [2]. El procedimiento para la codificación de fase es el siguiente:

1. Leer el audio de entrada y dividirlo en N segmentos cortos.

2. Aplicar la Transformada Discreta de Fourier (DFT, por sus siglas en inglés) a cada uno de los N segmentos y almacenar tanto la magnitud como la fase de cada segmento.

[1] En el análisis espectral, el sonido se separa en componentes de amplitud y fase de las frecuencias que lo forman. La fase se refiere a la posición de una onda de sonido con respecto a su inicio.

[2] Se refiere a los componentes de fase obtenidos de un análisis espectral.

3. Calcular y almacenar la diferencia de los segmentos vecinos.

4. Sea el vector de fase del primer segmento de la señal $\frac{\pi}{2}$ si el bit a ocultar es 0 ó $-\frac{\pi}{2}$ si el bit a ocultar es 1.

5. El vector de fase resultante debe ser la suma del vector de fase del segmento anterior con la diferencia de fase que se almacenó en el paso 2.

6. Usar el vector de fase del paso 5 y la magnitud del paso 2 para realizar la DFT inversa para obtener la señal marcada.

Después del proceso de inserción, la fase absoluta de la señal marcada es diferente de aquella en la señal original; sin embargo, la diferencia relativa de la fase se mantiene en la señal marcada. Al no alterar esta fase relativa de la fase, es posible insertar marcas de agua inaudibles.

En la decodificación es necesario hacer algunas sincronizaciones antes de la propia decodificación. Se debe conocer de antemano la longitud de cada segmento, los puntos DFT y los intervalos de los datos. Con esta información, se puede calcular la DFT y detectar el vector de fase para el primer segmento de la señal, este primer segmento es el que contiene el mensaje oculto.

Aunque la codificación de fase es uno de los métodos más efectivos, en términos de la relación señal-ruido percibido e inaudibilidad de la marca, tiene algunas desventajas. La primera, es la distorsión introducida en la señal marcada, causada por el cambio en la relación entre cada componente de frecuencia; esta distorsión es conocida como dispersión de fase. A mayor distorsión de fase, se puede insertar

más información, pero también se puede producir más distorsión audible. Otra desventaja es la distorsión causada por la tasa de cambio del modificador de fase; al cambiar la fase lo suficientemente lento, la distorsión se puede reducir, logrando una marca inaudible.

A pesar de todos estos inconvenientes, la codificación de fase aún tiene un rol muy importante en las marcas de agua para audio, debido a su potencial para insertar gran cantidad de información y su facilidad de implementación. Asimismo, existen muchos trabajos en los últimos años, en los que se han podido insertar marcas de agua inaudibles de manera exitosa.

Codificación de espectro disperso

Existen varios esquemas de marcas de agua basados en la técnica de espectro disperso y se han propuesto desde mediados de los 90's, cuando Cox introdujo por primera vez el espectro disperso para marcas de agua. Un esquema típico de espectro disperso basado en cuadros para marcas de agua en audio se muestra en las figuras 2.5 y 2.6, que representan el proceso de inserción y decodificiación, respectivamente. "T" denota transformadas, como Fourier, Transformada de Coseno Discreta (DCT, por sus siglas en inglés), *wavelet*, entre otras; "TI" es la transformada inversa correspondiente.

El proceso de inserción es el siguiente:

1. El audio original se segmenta en cuadros superpuestos con longitud de N muestras.

2. Si la marca se inserta en el dominio de la transformada, cada cuadro debe

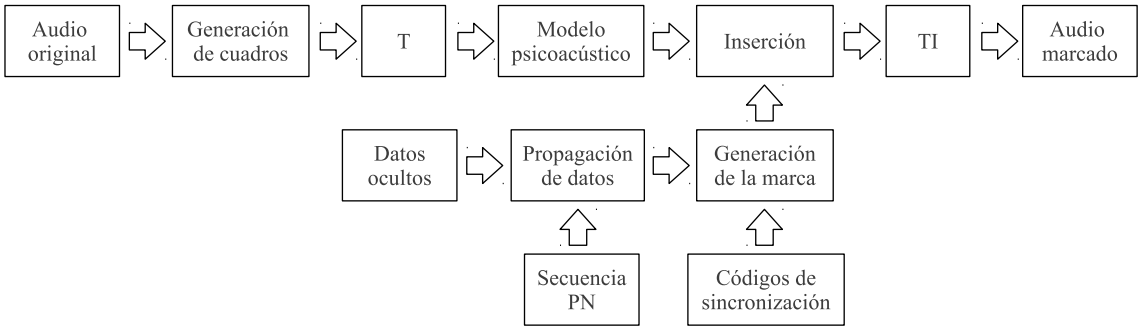


Figura 2.5: Inserción para el esquema de espectro disperso.

convertirse a ese dominio (por ejemplo, frecuencia, sub banda, *wavelet*, etc.).

3. Se aplica un modelo psicoacústico para determinar los límites de enmascaramiento de cada cuadro, de manera que los datos insertados sean inaudibles.

4. Se dispersan los datos, según una secuencia pseudo aleatoria (secuencia PN).

5. Los códigos de sincronización se añaden a los datos dispersos, con lo que se producen las marcas finales a ser insertadas.

6. La inserción de las marcas está condicionada por los umbrales de enmascaramiento.

7. Si la marca se insertó en el dominio de la transformada, es necesario realizar la transformada inversa a los cuadros marcados, para regresar el audio marcado al dominio del tiempo.

El proceso de detección trabaja de manera inversa a la inserción, consiste en los siguientes pasos:

1. Primero, el audio de entrada se segmenta en cuadros superpuestos.

2. El cuadro se convierte al dominio de la transformada, si es necesario.

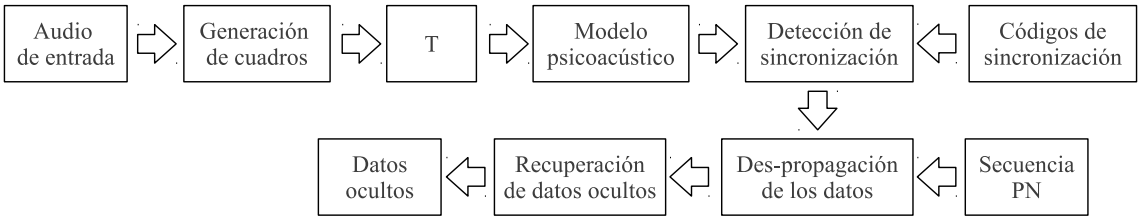


Figura 2.6: Decodificación para el esquema de espectro disperso.

3. Se aplica el mismo modelo psicoacústico que en la inserción para determinar los umbrales de enmascaramiento.

4. Se buscan los códigos apropiados para realizar la sincronización.

5. Los datos se des-propagan para poder detectar y recuperar la información oculta.

Las ventajas de esta técnica son que no se necesita el audio original para realizar la decodificación y se logra una robustez relativamente alta, razón por la que esta técnica sea tan popular. Sin embargo, tiene la desventaja que debe existir una perfecta sincronización entre la señal marcada y la marca o secuencia pseudo aleatoria, durante el proceso de decodificación. También, la longitud de la marca debe ser suficiente para alcanzar una probabilidad de error aceptable, lo que incrementa la complejidad y el tiempo para realizar la detección.

En esta técnica, la señal por sí misma es vista como una fuente de interferencia que puede degradar la calidad del audio. Además, la secuencia de longitud máxima se utiliza como secuencia pseudo aleatoria, lo que tiene dos inconvenientes; el primero es que la longitud de la tasa de fragmentos está limitado por $2^M - 1$, donde M es la longitud de la secuencia; el segundo es que las diferentes secuencias que

pueden generarse también están limitadas, por lo que es inseguro en términos criptográficos. Finalmente, existe un problema al trabajar con audios comprimidos, ya que la mayoría de los esquemas deben introducir una descompresión completa para poder realizar la detección, lo que hace que el proceso tarde más tiempo.

2.2.2. Ataques

Como se menciona en la sección anterior de este capítulo, un medio marcado puede sufrir manipulaciones durante su transmisión, por ello es deseable que los esquemas de marcas de agua consideren la robustez ante diferentes tipos de manipulaciones. En el caso del audio y dependiendo del medio de transmisión, pueden realizarse diferentes operaciones. Por ejemplo, un audio que se transmita en una estación de radio puede ser normalizado, comprimido, ecualizado o filtrado; por otro lado, un audio transmitido por Internet, usualmente sufre de compresión MP3 o AAC; estas operaciones son los ataques. Actualmente existe el *benchmark* Stirmark, que presenta ataques para marcas de agua en audio y su clasificación es la siguiente [23]:

Dinámicos

- **Compresión.** Se utilizan algoritmos de compresión de audio con pérdida, basados en efectos psicoacústicos (MPEG y AAC), éstos reducen el tamaño de un audio en un factor de 10 o más.

- **Eliminación de ruido.** Como su nombre lo indica, los eliminadores de ruido remueven el ruido de las señales de audio. Se utiliza un parámetro que indica el volumen de las señales que se interpretan como ruido.

Filtrado

- **Pasa altas.** Se trata de un filtro que elimina todas las frecuencias que se encuentran debajo de un umbral determinado, para este caso de 50 Hz.
- **Pasa bajas.** Similar al filtro pasa altas, se eliminan las frecuencias arriba de un umbral de 15 kHz.
- **Ecualizador.** Se utiliza un ecualizador para reducir en 48 dB los canales de cierta frecuencia.
- **División de canales.** Se emplea un efecto de ecualizador para incrementar la imagen estereo percibida; es decir, las frecuencias se reducen en un canal, mientra que se incrementan en el otro.

Ambientación

- **Retraso.** Se añade una copia retrasada a la señal original, esto se utiliza para simular espacios abiertos.
- **Reverberación.** Este efecto se utiliza para simular el ambiente en habitaciones o edificios. Es similar al ataque anterior, donde se añade una copia retrasada de la señal original, pero los tiempos para retrasar dicha copia son más cortos y además se agregan reflecciones del audio.

Conversión

- **Remuestreo.** En estos ataques, se modifica la frecuencia de muestreo de la señal. Por ejemplo, una señal original de 44.1 kHz se submuestrea a 29.4 kHz. Con esto, se reduce la frecuencia más alta posible y el resultado es similar a un filtro pasa bajas.

- **Inversión.** Este es un ataque inaudible que cambia el signo de las muestras.

Ruido

- **Ruido aleatorio.** A las muestras originales se les añaden números generados aleatoriamente; para generarlos, se tiene un parámetro que restringe la cantidad relativa del número, comparado con la señal original.

Modulación

- **Coros.** A la señal original se le añade un eco modulado con varios tiempos de retraso, modulaciones y número de voces.

- **Flanger.** Este efecto generalmente se crea al mezclar una señal con una copia ligeramente retrasada, donde la longitud del retraso se cambia constantemente.

- **Potenciador.** Se utiliza un potenciador para incrementar la cantidad de frecuencias altas de una señal. Este efecto también es conocido como "excitador".

Ampliación de tiempo y cambio de tono

- **Cambio de tono.** Este efecto se utiliza para cambiar la frecuencia base sin cambiar la velocidad del audio. Es uno de los algoritmos para edición de audio más sofisticados y existen diferentes versiones, especializadas en brindar diferentes calidades, dependiendo de las características de la señal original.

- **Ampliación de tiempo.** Es similar al anterior, pero se utiliza para incrementar o decrementar la duración del audio, sin cambiar su tono.

Permutación de muestras

- **Inserción de silencios.** En este ataque, se buscan muestras con valor 0 y se agregan 20 ceros más en esa posición, con lo que se crea una pequeña pausa en la señal. La distancia mínima entre pausas es de un segundo.
- **Copiar muestras.** Se seleccionan aleatoriamente muestras que son repetidas en la señal, por lo que se incrementa la duración del audio.
- **Intercambiar muestras.** Se toman muestras en posiciones aleatorias, posteriormente, una muestra se cambia a la posición de una segunda muestra y viceversa.
- **Cortar muestras.** Se elimina de la señal una secuencia aleatoria de muestras.

En esta clasificación no existe una sección donde se hable propiamente de ataques de desincronización. No obstante, los ataques de retraso, reverberación, cambio de tono, ampliación de tiempo, inserción de silencios, intercambio y corte de muestras son ataques que alteran la sincronización de las marcas al momento de la detección, por lo que en la literatura actual éstos son considerados ataques de desincronización.

2.2.3. Evaluación

Hasta este punto se han presentado las técnicas más utilizadas para insertar las marcas de agua y los posibles ataques a los que se pueden enfrentar los audios marcados. Sin embargo, es necesario conocer el rendimiento de los esquemas desde diferentes puntos de vista. A continuación se analizarán algunas medidas para marcas de agua en audio [14].

Fidelidad

Como se ha mencionado anteriormente, el proceso de inserción de la marca dentro de un audio puede producir una degradación en el medio original. Sin embargo, se desea que esta degradación sea lo suficientemente pequeña como para que no pueda ser percibida por una persona. Para ello, es necesario medir el impacto que tiene la inserción; esto se logra midiendo la fidelidad del audio marcado. La fidelidad se refiere a la similitud entre los audios original y marcado, por lo que necesita usarse una medida estadística. Esta medida puede ser de dos tipos: métricas de diferencia o métricas de correlación.

Como su nombre lo indica, las *métricas de diferencia* miden la diferencia entre el audio original X y el audio marcado Y. La métrica de diferencia más común es la *relación señal a ruido* (SNR, por sus siglas en inglés). La fórmula para calcular el SNR en decibeles es la siguiente:

$$SNR(dB) = 10log_{10}\frac{\sum_n X_n^2}{\sum_n (X_n - Y_n)^2} \tag{2.10}$$

donde X_n corresponde a la n-ésima muestra del audio original X y Y_n es la n-ésima muestra del audio marcado Y. Esta medida de calidad refleja la distorsión que una marca impone sobre una señal. Aunque el ruido tolerable depende de la aplicación y las características del audio original, se espera tener distorsiones con SNR de 35 dB. Otra métrica de diferencia es la *relación señal a ruido pico* (PSNR, por sus siglas en inglés), la cual mide la relación señal a ruido máxima en una señal de audio. Su fórmula es la siguiente:

$$PSNR = \frac{N(max_n X_n)^2}{\sum_n (X_n - Y_n)^2} \tag{2.11}$$

Las *métricas de correlación* miden la distorsión basada en la correlación estadística entre los audios original y marcado. En esta categoría están la *correlación cruzada normalizada* (NC, por sus siglas en inglés) y la *calidad de correlación* (QC, por sus siglas en inglés).

Cuando se utiliza una métrica que da resultados en decibeles, es difícil hacer comparaciones ya que su escala es logarítmica. Es más sencillo presentar los resultados utilizando una clasificación de calidad normalizada. La clasificación de calidad ITU-R Rec.500 es ideal para esta tarea, pues proporciona una clasificación con una escala de 1 a 5. La tabla 2.1 presenta la clasificación junto con la calidad que representa. Esta clasificación de calidad se calcula con la fórmula:

$$Calidad = F = \frac{5}{1 + N * \mathrm{SNR}} \tag{2.12}$$

donde N es una constante de normalización y SNR es la relación señal a ruido medida. El resultado corresponde a la fidelidad F de la señal marcada.

Tabla 2.1: Clasificación de calidad ITU-R Rec.500

Clasificación	Descripción	Calidad
5	Imperceptible	Excelente
4	Perceptible, no es molesto	Buena
3	Poco molesto	Razonable
2	Molesto	Pobre
1	Muy molesto	Mala

Robustez

La robustez permite saber si un esquema de marcas de agua es resistente ante diferentes tipos de operaciones, ya sean intencionales o no. Se dice que un esquema es robusto cuando es capaz de tolerar ataques que degraden la marca, al grado

que ésta sea removida, o que el proceso de detección falle. Esto significa que sólo con interferir con el proceso de detección se puede tener un ataque exitoso. Sin embargo, en algunos casos es posible utilizar códigos de corrección de errores o un detector más fuerte para tratar de compensar estos ataques. Si se utiliza un código de corrección de errores, no es necesario recuperar la marca completa para conocer el mensaje oculto. También se puede diseñar un esquema con detectores de diferente fuerza.

Con lo anterior, se puede apreciar que es necesaria una métrica que de como resultado diferentes niveles de robustez y el procedimiento básico para calcularla es el siguiente:

1. Para cada audio en un determinado conjunto de prueba, insertar una marca aleatoria W en el audio X, con la máxima fuerza posible que no lleve la fidelidad por debajo del mínimo especificado.

2. Aplicar un conjunto relevante de operaciones de procesamiento al audio marcado Y.

3. Finalmente, para cada audio marcado, extraer la marca W utilizando el detector correspondiente y medir el éxito del proceso de recuperación.

La *tasa de bits erróneos* (BER, por sus siglas en inglés) es una métrica que permite tener una escala de valores; se define como la razón entre el número de bits extraídos erróneos y el total de bits insertados, se puede expresar con la siguiente fórmula:

$$BER = \frac{B}{L} * 100\,\% \tag{2.13}$$

$$B = \sum_{n=1}^{L} \begin{cases} 1, & W'_n \neq W_n \\ 0, & W'_n = W_n \end{cases} \tag{2.14}$$

donde B es el número de bits detectados erróneamente, L es la longitud de la marca, W_n corresponde al n-ésimo bit de la marca insertada y W'_n al n-ésimo bit de la marca recuperada.

Perceptibilidad

La calidad del audio es un tema muy ambigüo y la opinión varía de un individuo a otro. Existen algunas medidas subjetivas, como el *grado de diferencia subjetiva* (SDG, por sus siglas en inglés); sin embargo, en este trabajo solamente se hablará de las medidas objetivas.

La meta de los algoritmos de medición objetiva es sustituir las pruebas subjetivas de escucha, al modelar el comportamiento de la escucha humana; estos algoritmos permiten obtener una medida de calidad que consiste en un solo número para describir qué tan audible es una distorsión introducida. Estos algoritmos de medición objetiva de calidad de audio tienen una arquitectura general, que se muestra en la figura 2.7.

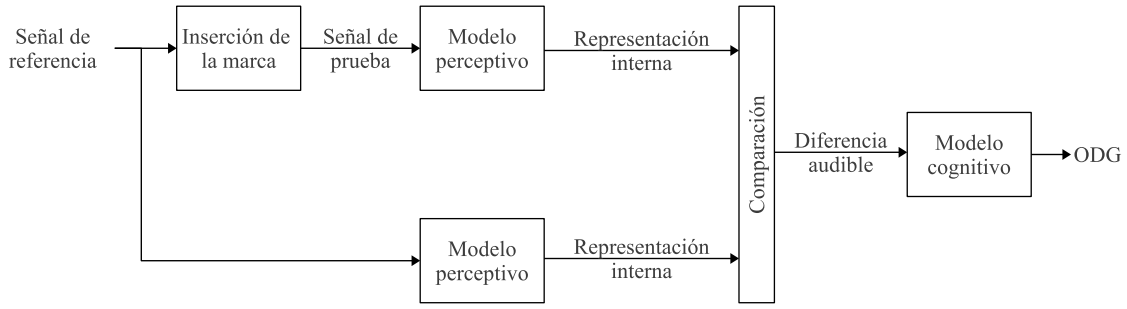


Figura 2.7: Arquitectura general para medición objetiva de calidad de audio.

Se utiliza una técnica para medir la diferencia entre la señal de referencia (original) y la señal de prueba (marcada); ambas son procesadas por un sistema auditivo, que calcula un estimado de los componentes audibles de la señal. Estos componentes pueden ser considerados como la representación de las señales en el sistema auditivo humano. La *representación interna* está relacionada con el umbral de enmascaramiento. A partir de estas dos representaciones internas, se calcula la *diferencia audible*. El *modelo cognitivo* modela el procesamiento de las señales que realiza el cerebro durante las pruebas de escucha. La salida del sistema completo es el *grado de diferencia objetiva* (ODG, por sus siglas en inglés).

Hasta este punto, se han analizado los aspectos más importantes de las marcas de agua para audio; sin embargo, existe el caso de la sincronización entre la marca y el detector, que representa un reto importante para los esquemas de marcas de agua actuales. En la siguiente sección de este capítulo se trata el problema de la sincronización en marcas de agua para audio, así como las técnicas utilizadas para combatirlo.

2.3. Sincronización en marcas de agua para audio

En las marcas de agua para audio, existe una problemática que es de especial interés, porque aún no se ha logrado resolver completamente y se trata de la sinronización entre la marca y el detector. En la figura 2.8 se representa un esquema de marcas de agua, visto desde un enfoque de comunicaciones [8].

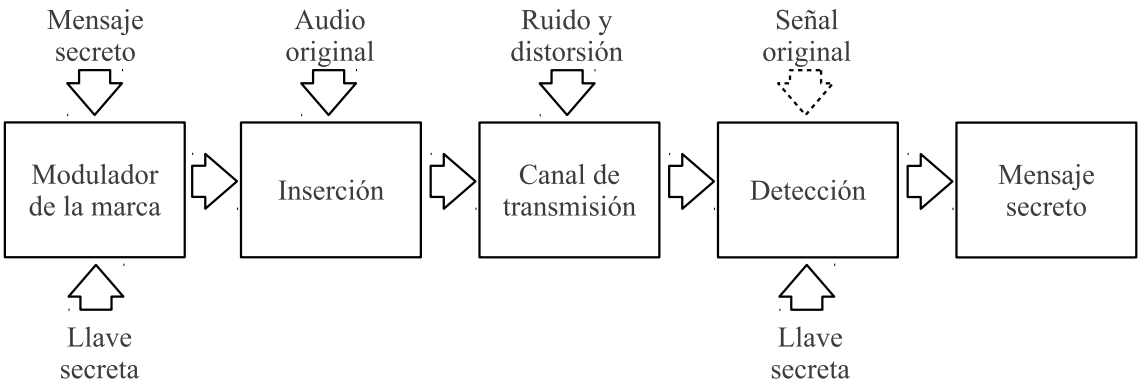


Figura 2.8: Esquema de marcas de agua desde un enfoque de comunicaciones.

Supóngase que el mensaje a ser transmitido es m, que usualmente está modulado a la marca w y el audio original x, entonces el audio marcado es:

$$s = x + aw \tag{2.15}$$

donde a es un factor entre 0 y 1, que permite controlar la fuerza de la marca. Este factor ayuda a prevenir que se introduzca una distorsión perceptible al medio marcado.

En el proceso de detección, se realiza una correlación entre la señal recibida y la marca. Un valor alto de correlación significa la existencia de la marca. Si la longitud de la marca es N y el ruido introducido por el canal de transmisión es n, la señal recibida r es:

$$r = s + n = x + aw + n \tag{2.16}$$

y la correlación entre r y w es:

$$\begin{aligned} c &= \frac{1}{N}\sum_i r_i w_i \\ &= \frac{1}{N}\sum_i x_i w_i + \frac{1}{N}\sum_i a w_i^2 + \frac{1}{N}\sum_i n_i w_i \\ &= d_x + d_w + d_n \end{aligned} \tag{2.17}$$

Asumiendo que N es lo suficientemente grande, la marca w, el audio original x y el ruido n son independientes y tienen una distribución Gaussiana; entonces:

$$d_x \approx 0 \quad \text{y} \quad d_n \approx 0 \tag{2.18}$$

El umbral de detección óptimo es:

$$\tau = \frac{1}{2} d_w \tag{2.19}$$

Durante el proceso de detección se necesita conocer la ubicación inicial de la marca en la señal recibida; de ahí el problema de sincronización. La ecuación 2.17 es válida sólo cuando el audio recibido está perfectamente alineado con la marca; es decir, cuando la señal r está sincronizada con la marca w, el valor de la correlación c alcanzará su máximo. Existen algunas estrategias clásicas, con las que se ha pretendido compensar esta problemática y se mencionan a continuación [3] [8].

2.3.1. Búsqueda exhaustiva

Es la técnica más simple para combatir los ataques de desincronización. La estrategia consiste en definir un rango probable de valores para cada parámetro de distorsión y por cada parámetro, una resolución de búsqueda; posteriormente se examinan cada una de estas combinaciones de parámetros.

El rango de búsqueda puede limitarse al asumir que, si se realizaron distorsiones, éstas tuvieron un impacto mínimo en la calidad perceptual del audio. La resolución de la búsqueda se puede determinar analizando la robustez natural que tiene la marca contra ciertos ataques y enfocarse a buscar sólo aquellos de interés. Cada combinación de valores para los parámetros de distorsión representan una distorsión hipotética que pudo haber ocurrido. En la búsqueda se puede invertir cada distorsión hipotética y realizar la detección para cada marca posible.

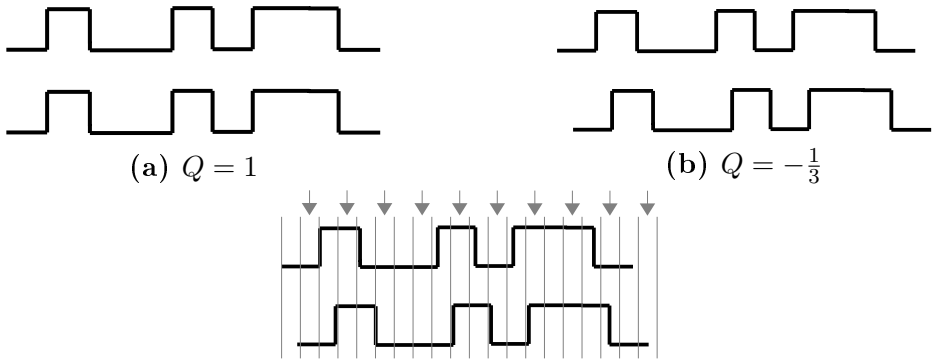
Sin embargo, esta estrategia tiene dos desventajas principales. La primera es el costo computacional, pues los cálculos requeridos aumentan según el tamaño del espacio de búsqueda. La segunda es con respecto a la probabilidad de errores en el detector; es decir, la probabilidad de falsos positivos puede ser muy alta si se prueba con una gran cantidad de audios sin marcar.

2.3.2. Marcas redundantes

En esta técnica, la idea es repetir N veces la marca antes de ser insertada en el audio, con esto, se espera que exista una mayor probabilidad de lograr la sincronización con el detector. En la figura 2.9 se ilustra un ejemplo de detección utilizando este método.

En la figura 2.9a se muestra una sincronización perfecta entre la marca insertada y la detectada, por lo que el valor de la correlación normalizada es $Q = 1$. La figura 2.9b representa la situación donde la marca está recorrida una posición hacia la derecha, debido a ataques de desincronización, por lo que la correlación normalizada es $Q = -\frac{1}{3}$. Finalmente, en la figura 2.9c la marca se repite tres veces antes de la inserción y la correlación normalizada se calcula únicamente con las

muestras del centro de cada región, por lo que se obtiene el valor de $Q = 1$, a pesar de la desincronización.



(a) $Q = 1$ **(b)** $Q = -\frac{1}{3}$

(c) $Q = 1$

Figura 2.9: Ejemplo de detección para marcas redundantes.

Con este enfoque, la correlación es correcta siempre y cuando el desplazamiento lineal sea menor a $\lfloor \frac{N}{2} \rfloor$ muestras. Su principal desventaja es la poca capacidad de inserción, que se reduce N veces (el mismo número de veces que se repite la marca).

2.3.3. Dominio invariante

Con este enfoque, se plantea insertar la marca en un dominio que sea invariante a posibles ataques. El cambio de dominio se logra utilizando transformadas, como la Transformada Discreta de Fourier.

Originalmente, esta idea se implementó para marcas de agua en imágenes. En estos primeros trabajos, los autores propusieron diferentes esquemas de marcas de agua que trabajaban en el dominio Fourier-Melin. Este proceso incluye tres pasos: aplicar una DFT a la señal, mapear las magnitudes DFT a un sistema de coordenadas polares y, finalmente, aplicar una segunda transformada DFT.

En los primeros intentos de introducir la transformada Fourier-Melin a las marcas de agua para audio, se realizó un esquema de *fingerprinting* resistente a cambios de velocidad, que se basaba en esta transformada.

2.3.4. Patrón de sincronización

Los patrones de sincronización son códigos con características especiales, que se conocen tanto en la inserción como en la decodificación. Usualmente no contienen información de la marca y se utilizan únicamente con propósitos de sincronización. El detector puede utilizar las características especiales de los patrones para realizar una búsqueda exhaustiva, logrando con ello la sincronización y manteniendo una tasa de error baja. Las secuencias pseudo-aleatorias (PN, por sus siglas en inglés) son ampliamente utilizadas como patrones de este tipo, ya que tienen propiedades bien conocidas de autocorrelación.

Sin embargo, esta técnica tiene como desventajas su poca capacidad de inserción, puesto que los patrones mismos no contienen información de la marca. También presentan problemas de seguridad, ya que los oponentes pueden adivinar las secuencias PN utilizadas y realizar ataques de plantillas de estimación.

2.3.5. Auto-sincronización

Las marcas auto-sincronizadas se utilizan con el objetivo de resolver los problemas existentes con los patrones de sincronización. Estas marcas están diseñadas especialmente para que la autocorrelación de las marcas tenga uno o varios picos, con los que se logra la sincronización.

Aunque existen métodos que han utilizado esta estrategia para realizar marcas exitosas, también son susceptibles a ataques de estimación. Una vez que conoce la marca, el oponente puede realizar una autocorrelación entre el audio marcado y la marca, removiendo los picos periódicos, logrando que la detección falle.

2.3.6. Puntos característicos

La sincronización mediante puntos característicos o basada en contenido, se considera un método de segunda generación. Los métodos de primera generación trabajan en el dominio del tiempo o la transformada para realizar la inserción de la marca; los de segunda generación, por su parte, utilizan la noción de características del medio o puntos característicos [13]. Los puntos característicos son secciones estables en la señal original y son invariantes o casi invariantes ante procesamiento de señales o ataques. Estos puntos se extraen tanto en la inserción como en la decodificación para lograr la sincronización.

Existen trabajos que utilizan diferentes propiedades para encontrar estos puntos característicos. Por ejemplo, al realizar un análisis de contenido del audio, se pueden detectar áreas de rápido incremento de energía. Dichas áreas son perceptualmente importantes y por lo tanto, estables ante procesamiento de señales.

Otra estrategia basada en puntos característicos, es detectar los ritmos en la música, pues son los eventos más robustos. En la inserción, se estima el periodo promedio de los ritmos y las marcas se insertan al inicio de éstos. En la detección, primero se calcula el periodo promedio de ritmos y posteriormente se localiza el inicio de cada uno de ellos, con lo que se logra la sincronización con la marca.

Las características estadísticas también pueden ser utilizadas como puntos característicos. Un esquema que lleva a cabo esto, transforma el audio original mediante una Transformada *Wavelet* Discreta (DWT, por sus siglas en inglés) y posteriormente localiza características estadísticas en el dominio *wavelet*. La media de los coeficientes en la sub-banda de aproximaciones, proporciona información sobre la distribución de las bajas frecuencias del audio. Estas bajas frecuencias representan los componentes que contienen mayor información de la señal, por lo que son relativemente estables antes procesamientos comunes.

Los métodos basados en puntos característicos no necesitan insertar información adicional para lograr la sincronización, lo que permite una capacidad de inserción mayor, sin introducir demasiada distorsión al audio. No obstante, es difícil que estos métodos logren una sincronización correcta muestra por muestra.

En el siguiente capítulo se analizarán los trabajos actuales que tratan de resolver la problemática de sincronización en marcas de agua para audio.

Capítulo 3

Trabajo relacionado

3.1. Estrategias de sincronización en marcas de agua para audio

Como se menciona en el capítulo anterior, la sincronización es un aspecto fundamental para los esquemas actuales de marcas de agua para audio, dado que ciertos ataques no permiten que el detector encuentre las marcas adecuadas. En la sección 2.3 del capítulo anterior, se mencionan las estrategias utilizadas en los esquemas para tratar de combatir los ataques de desincronización. Ahora, se realizará un análisis de los trabajos existentes en la literatura actual que pretenden combatir esta problemática. Se describirán los esquemas que caen dentro de cada estrategia de la clasificación presentada anteriormente.

3.1.1. Búsqueda exhaustiva

El trabajo de Kirovski, *et al.* [12] emplea una búsqueda exhaustiva para realizar la detección de la marca. Utilizan la técnica de Espectro Disperso con Transfor-

mada *Lapped* Compleja Modulada (MCLT, por sus siglas en inglés) para llevar a cabo tanto la inserción como la detección de la marca. Se seleccionan bloques de audio que se trasforman con la MCLT y los bits se insertan en los componentes de magnitud que resultan de la transformada. Para la detección, se generan versiones de las marcas que corresponden a distintos ataques y se comparan con la marca del bloque que se está evaluando.

Para que este esquema tenga una tasa de falsos positivos aceptable, se necesitan audios grandes que son difíciles de sincronizar. Además, la búsqueda exhaustiva que se realiza en la detección, hace que este esquema sea impráctico para aplicaciones en ambientes muy ruidosos o con muchos ataques, puesto que el espacio de búsqueda sería demasiado grande. No obstante, por las características de este trabajo, sería más eficiente para marcas de agua en video.

3.1.2. Patrón de sincronización

Erçelebi y Batakçi [5] utilizan patrones de sincronización que permiten detectar las marcas insertadas. Transforman los audios de entrada al dominio *wavelet*, separando componentes de alta y baja frecuencia. Se genera una marca aleatoria que se inserta en los componentes de baja frecuencia, dependiendo de los valores de una imagen binaria, la cual determina si se inserta el bit correspondiente de la marca.

Este esquema no es robusto contra ataques de compresión MP3 y sólo se reportan resultados para el ataque de desincronización de remuestreo. Sin embargo, la novedad de este trabajo es el uso de la imagen binaria para controlar la inser-

ción y detección, con lo que agregan seguridad al esquema. Aunque los atacantes descubran el funcionamiento de los métodos de inserción y detección, si no cuentan con la imagen binaria con la que se ocultó la marca, será imposible recuperarla.

Otro trabajo donde utilizan patrones de sincronización, es en el de Niu, Wang y Lu [18]. En éste, el audio original se segmenta y cada segmento se divide en dos partes. En la primera parte, se inserta un código de sincronización en la media estadística de las muestras. Las muestras de la segunda parte se mapean de un plano 1D a 2D y posteriormente se calculan los momentos *wavelet*. A partir de los momentos *wavelet* de orden bajo, se calcula el módulo del valor promedio y en éste se inserta el bit de la marca.

No obstante, los códigos de sincronización se insertan en el dominio del tiempo, lo que puede llevar a desincronización si se utilizan ataques muy severos; asimismo, no presenta robustez contra ataques de *jittering* [3]. Pero tiene la ventaja de presentar robustez contra ataques híbridos; es decir, ataques de procesamiento de señales combinados con ataques de desincronización.

3.1.3. Auto-sincronización

Un trabajo que utiliza la técnica de auto-sincronización es el de Martínez-Noriega, *et al.* [16]. Ellos dividen el audio en grupos de muestras y para insertar los bits de la marca, se modifican las amplitudes de las medias de estos grupos. La marca se codifica con códigos LDPC, con lo que se logra la auto-sincronización.

[3] *Jittering* es un ataque de desincronización que elimina una de cada N muestras de un audio.

Este esquema tiene la desventaja de trabajar en el dominio del tiempo, por lo que es probable que ciertos ataques de desincronización muy severos afecten gravemente su desempeño. Aunque en su documento dicen tener robustez contra ataques de desincronización, no reportan los resultados de experimentos con este tipo de ataques. Sin embargo, al trabajar en el dominio del tiempo y no necesitar patrones de sincronización, tiene una carga útil mayor que la de otros trabajos.

3.1.4. Puntos característicos

Esta estrategia es de interés especial, ya que en los últimos años ha tenido un gran auge. De este tipo de esquemas, ha surgido interés especial en aquellos que utilizan características estadísticas para tratar de combatir la desincronización, en especial utilizan la media estadística y los histogramas del audio. A continuación se describirán algunos trabajos representativos que utilizan algunas técnicas estadísticas para extraer los puntos característicos de los audios.

Xiang, Huang y Yang [27] construyen un histograma, en el que insertan la marca y con esto tratan de combatir la desincronización. Extraen el histograma del audio, con bins de igual tamaño y el histograma se calcula a partir de un rango de amplitudes, relacionado con la media. La inserción de los bits se realiza reasignando las relaciones entre tres bins vecinos. Sin embargo, al trabajar en el dominio del tiempo, puede provocar que no sea robusto ante distintos tipos de ataques. Dentro de los ataques de desincronización, sólo se enfocan a los de modificación de la escala de tiempo [4] (TSM, por sus siglas en inglés) y recortes. También tiene como desventaja que el desempeño del esquema depende del valor de la media

[4]En el *benchmark* Stirmark, este ataque es equivalente al ataque de ampliación de tiempo.

estadística de los audios, lo que no es una característica necesariamente estable para distintos tipos de audio. Pero, precisamente por trabajar en el dominio del tiempo, es un esquema rápido y fácil de implementar.

Zhang, Yin y Yu [30] también construyen un histograma en el dominio del tiempo para tratar de mantener sincronización con el detector. A partir de un audio de entrada, se selecciona un rango de inserción y con éste, se construye un histograma en el dominio del tiempo. Tanto la inserción como la detección se realiza a partir de los bins de este histograma. Al insertar, se modifican las relaciones entre cuatro bins consecutivos y al detectar, se verifica que dichas relaciones cumplan con los criterios de detección. Sin embargo, se enfoca contra ataques de TSM y filtros pasa bajas; además, se asume que los coeficientes de los audios tendrán una distribución normal, lo que probablemente no se cumpla para cualquier tipo de audio. Igual que el esquema anterior, es fácil de implementar por trabajar en el dominio del tiempo.

Wang, Niu y Yang [25] utilizan el promedio estadístico para realizar la inserción y detección de las marcas. Se segmenta un audio de entrada y cada segmento se divide en dos partes. Se inserta un código de sincronización en el promedio estadístico de las muestras de la primera parte. La segunda parte se divide en secciones, a las que se les aplica una transformada DWT y los bits de la marca se insertan en el promedio estadístico de los componentes de frecuencia obtenidos con la transformada. Sin embargo, los códigos de sincronización se insertan en el dominio del tiempo, por lo que se pueden perder con algunos ataques severos; además, este esquema presenta poca robustez contra ataques de TSM y *jittering*.

Fan y Wang [6] también utilizan el promedio estadístico dentro de su esquema. Se segmenta un audio original y se aplica una transformada DFT a cada cuadro. Se calcula el promedio absoluto para cada cuadro y se seleccionan las regiones de inserción, a partir de cuadros con distintos promedios. Cada cuadro de inserción se divide en segmentos y se calculan relaciones entre tres segmentos vecinos, los bits se insertan a partir de estas relaciones. Tiene como desventajas que su objetivo principal es tener robustez contra ataques de modificación de la velocidad de reproducción (PSM, por sus siglas en inglés); además, sus resultados de SNR y robustez contra filtros pasa bajas no son lo suficientemente altos. No obstante, esta estrategia es robusta contra ataques de TSM y variación de amplitud.

Tao, *et al.* [24] segmentan un audio original y cada segmento se divide en dos secciones, a cada una se le aplica una transformada *wavelet* con *lifting* (LWT, por sus siglas en inglés). Un código de sincronización y la marca se insertan en las secciones, respectivamente, al modificar el promedio estadístico de los coeficientes de la sub-banda. Este esquema sólo se enfoca contra recortes aleatorios y no es robusto contra ataques de TSM. Sin embargo, la implementación es eficiente por utilizar la transformada LWT, además que la fuerza de inserción se calcula de manera que se adapte a cada audio.

Yang, Wang y Ma [29] utilizan estadísticas de orden alto para realizar la inserción y detección de las marcas. Antes de comenzar el procedimiento, se realiza una eliminación de ruido con *wavelets* del audio original, posteriormente éste se segmenta y cada segmento se divide en dos partes. En la primera parte, se inserta un código de sincronización en el promedio estadístico y se obtienen estadísticas de orden alto mediante la distancia de Hausdorff. La marca se inserta en el domi-

nio wavelet utilizando las estadísticas de orden alto. Como desventajas, se tiene que el código de sincronización se inserta en el dominio del tiempo, por lo que se puede perder con ataques severos y el esquema presenta poca robustez contra TSM.

Xiang [26] construye un histograma para tratar de combatir la desincronización. Se extrae el histograma a partir de los componentes de baja frecuencia de un filtro Gaussiano. Se utilizan dos bins consecutivos del histograma para insertar las marcas, al reasignar las relaciones entre éstos. Este esquema no es robusto contra recortes y realiza dos búsquedas en los coeficientes de la transformada, por lo que su implementación no es eficiente cuando se tienen audios muy grandes. Sin embargo, presenta robustez contra TSM.

Finalmente, Yang *et al.* [28] también construyen un histograma para realizar la inserción y detección de la marca. Aplican una transformada *wavelet* no-decimada (UDWT, por sus siglas en inglés), con la que se obtienen coeficientes de alta y baja frecuencia. Se selecciona un rango de valores en los coeficientes de baja frecuencia, a partir del cual se construye un histograma invariante. Se seleccionan cuatro bins consecutivos para insertar un bit de la marca, al reasignar las relaciones que existen entre éstos. Tiene como desventajas que presenta poca robustez contra ataques de TSM; además que se asume que los coeficientes UDWT de baja frecuencia presentan una distribución normal, lo que puede no ser cierto para todos los audios. Sin embargo, el histograma construído a partir de estos coeficientes es robusto contra distintos tipos de ataques, pues a pesar de los procesamientos conserva su forma.

3.2. Resumen de las estrategias de sincronización

En la tabla 3.1 se presenta un resumen de los esquemas anteriormente descritos, donde se pueden apreciar sus resultados contra distintos tipos de ataques. En la tabla se presentan los autores del trabajo, el año de publicación, la técnica empleada en el esquema, el dominio en el que realiza la inserción de la marca, algunos ataques reportados con sus respectivos resultados de BER, PSNR y correlación (según sea el caso), el *payload* (carga útil dada en bits por segundo) reportado y finalmente, algunas observaciones sobre el esquema.

El esquema con la estrategia de búsqueda exhaustiva no resiste gran cantidad de ataques, ya que se deben realizar comparaciones entre la marca que se detecta y las posibles marcas resultantes de los ataques; con ello, se vuelve un esquema impráctico y que sirve sólo con algunos ataques de desincronización.

El esquema con estrategia de patrón de sincronización realiza la sincronización de sus marcas en el dominio del tiempo, haciéndolo sensible ante ataques de desincronización severos, pues es probable que al modificar la posición de las muestras se pierda la información necesaria para detectar la marca.

Los esquemas que trabajan con la técnica de puntos característicos parecen ser los que presentan mejores resultados; sin embargo, Yang *et al.* [28], y Yang, Wang y Ma [29] reportan más ataques de desincronización que los otros autores, además de los ataques comunes de procesamiento de señales. De estos últimos esquemas, los resultados de Yang *et al.* son mejores que los de Yang, Wang y Ma para los mismos ataques, lo que sugiere que dicho esquema es más robusto.

El trabajo de Yang *et al.* [28] se utiliza como esquema base por ser el que presenta mayor robustez ante más ataques de desincronización. Al analizar algunos resultados experimentales del esquema base, se pudo observar que el desempeño depende del rango de coeficientes que se utilice para construir el histograma; por lo tanto, la hipótesis de este trabajo es que el esquema de Yang *et al.* puede mejorarse al tomar distintos rangos de coeficientes, considerando que las bajas frecuencias en el dominio *wavelet* no necesariamente tienen una distribución normal.

En el siguiente capítulo se explicará con detalle el esquema base y las modificaciones realizadas para el esquema propuesto.

Tabla 3.1: Resumen de los trabajos relacionados con ataques de desincronización en audio

Autores	Año	Técnica	Dominio	Ataque	BER	SNR (dB)	Correlación	Payload	Observaciones
Kirovski, Malvar [12]	2003	Búsqueda exhaustiva	MCLT	Compresión	—	—	1.000	—	Impráctico en ambientes con muchos ataques
				Remuestreo			1.000		
				Intercambio			0.970		
				Cortar muestras			0.970		
Niu, Wang, Lu [18]	2010	Patrón de sincronización	*Wavelet*	Remuestreo (36.75kHz)	0.0039	—	—	—	Los códigos de sincronización se insertan en el dominio del tiempo por lo que son susceptibles ante ataques severos
				Recorte (6s)	0.0000				
				Cambio de pitch	0.0000				
				Jittering (1/5000)	0.2500				
Xiang, Huang, Yang [27]	2007	Puntos característicos	Tiempo	MP3 (112kbps)	0.0167	43.97	—	2 bits/s	El desempeño del esquema depende de los coeficientes de los audios
				Recorte (3s)	0.0000				
				Jittering (1/5000)	0.0000				
				TSM (-30 %)	0.0500				
Zhang, Yin, Yu [30]	2008	Puntos característicos	Tiempo	MP3 (64kbps)	0.0666	—	—	3 bits/s	Se asume que los coeficientes tienen una distribución normal
				Pasa bajas (5kHz)	0.0333				
				Re-cuantización (32 bits)	0.0000				
				TSM (130 %)	0.0500				
Wang, Niu, Yang [25]	2009	Puntos característicos	DWT	Sin ataque	0.0000	46.30	—	≈420 bits/s	El código de sincronización se inserta en el dominio del tiempo, haciéndolo susceptible ante ataques severos
				Remuestreo (8kHz)	0.0320	27.51			
				MP3 (64kbps)	0.0010	47.33			
				Recortes (10 %)	0.0000	34.62			
				Cambio de pitch	0.0000	26.58			
				TSM (-4 %)	0.2070	26.84			
				Jittering (1/5000)	0.2120	26.79			
Fan, Wang [6]	2011	Puntos característicos	DFT	Sin ataque	0.0000	22.50	—	3 bits/s	Considera solamente un ataque de desincronización (PSM)
				Ruido (40dB)	0.0166	—			
				MP3 (48kbps)	0.0166				
				PSM (70 %)	0.1000				
Tao, Zhao, Wu, Gu, Xu, Gu [24]	2010	Puntos característicos	LWT	Añadir (10 %)	—	—	1.000	≈12 bits/s	No es robusto contra TSM
				Recortes (10 %)			1.000		
				Jittering (1/100)			0.823		
				Recorte aleatorio			0.910		

Continúa en la siguiente página

Tabla 3.1 Resumen de los trabajos relacionados ... (continuación)

Autores	Año	Técnica	Dominio	Ataque	BER	SNR (dB)	Correlación	Payload	Observaciones
Yang, Wang, Ma [29]	2011	Puntos característicos	DWT	Remuestreo (8kHz)	0.1152	—	—	≈102 bits/s	El código de sincronización se inserta en el dominio del tiempo, haciéndolo susceptible ante ataques severos
				Pasa bajas (4kHz)	0.1846				
				MP3 (64kbps)	0.1035				
				Recortes (10 %)	0.0000				
				Cambio de pitch	0.0342				
				TSM (-4 %)	0.2461				
				Jittering (1/5000)	0.1729				
Xiang [26]	2011	Puntos característicos	Frecuencia	Pasa bajas (6kHz)	0.1000	—	—	3 bits/s	No es robusto ante recortes
				Ruido (45dB)	0.0166				
				MP3 (56kbps)	0.0333				
				Recortes (15 %)	0.1833				
				TSM (-4 %)	0.0000				
Yang, Bao, Wang, Niu [28]	2010	Puntos característicos	UDWT	Pasa bajas (1kHz)	0.3000	—	—	3 bits/s	El desempeño del esquema depende de los valores de la transformada, porque se asume que éstos tienen distribución normal
				Ruido aditivo (100)	0.3000				
				MP3 (64 kbps)	0.0833				
				Recortes (15 %)	0.0000				
				Cambio de pitch	0.0000				
				Jittering (1/3000)	0.0000				
				TSM (-4 %)	0.0000				
				TSM (+11 %)	0.2833				
Menéndez, Cumplido, Feregrino	2012	Puntos característicos	DWT	—	—	—	—	3 bits/s	Se espera que al modificar el cálculo del rango de inserción se tenga robustez ante ataques de desincronización, sin importar la distribución en los valores de la transformada

Capítulo 4

Esquema de marcas de agua para audio utilizando wavelets

En este capítulo se describirán con detalle los esquemas base [28] y propuesto. De cada uno de ellos, se explicarán brevemente las transformadas *wavelet* en las cuales se basan, la construcción de los histogramas, así como los algoritmos empleados para la inserción y detección de la marca; finalmente, se presentará una discusión sobre los esquemas, analizando sus ventajas y desventajas.

4.1. Esquema base

4.1.1. Transformada UDWT

La transformada *wavelet* discreta no-decimada (UDWT, por sus siglas en inglés) fue descubierta independientemente en diferentes ocasiones y se conoce con distintos nombres, por ejemplo, la transformada *wavelet* invariante a traslaciones, la transformada *wavelet* estacionaria o la transformada *wavelet* redundante. La clave de esta transformada es que es redundante, invariante a traslaciones, es lineal y provee una mejor aproximación a la transformada *wavelet* continua que la aproximación dada por la transformada *wavelet* discreta (DWT, por sus siglas en inglés) ortonormal.

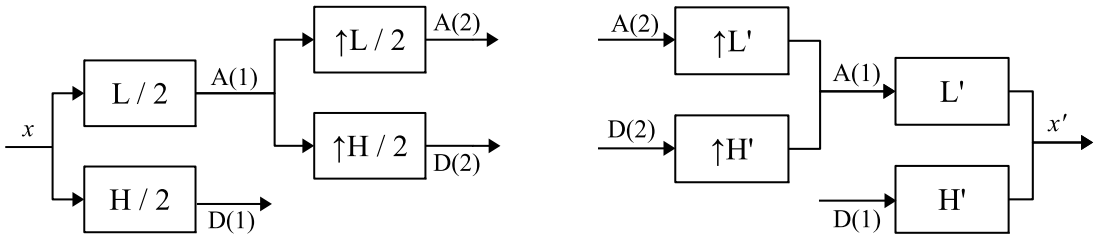
En una transformada *wavelet* ortonormal, existe una función de escalamiento $\phi(t)$ y una *wavelet* madre $\psi(t)$. La función de escalamiento $\phi(t)$ se puede construir a partir de un análisis de multiresolución en $L^2(R)$. El conjunto de funciones $\{2^{\frac{m}{2}}\phi(2^{\frac{m}{2}}l-n)\}$ es una base ortonormal de V_m. El conjunto de funciones $\{2^{\frac{m}{2}}\psi(2^{\frac{m}{2}}l-n)\}$ forma una base ortonormal de W_m, donde $V_{m+1} = V_m \oplus W_m = V_0 \oplus W_0 \oplus W_1 \oplus \cdots \oplus W_m$. Con esto, se puede descomponer la señal $x(t) \in L^2(R)$ en $\{V_0, W_0, W_1, \cdots, W_m\}$. Una descomposición *wavelet* ortonormal de una señal continua, con *wavelet* madre $\psi(t)$ resulta en:

$$\begin{aligned} w_j^k(x) &= \left\langle x(t), \frac{1}{2^{\frac{j}{2}}}\psi(\frac{t}{2^j}-k)\right\rangle \\ &= \frac{1}{2^{\frac{j}{2}}}\int_{-\infty}^{+\infty} x(t)\psi^*(\frac{t}{2^j}-k)dt, (k,j) \in Z^2 \end{aligned} \tag{4.1}$$

Un algoritmo eficiente para implementar la transformada *wavelet* discreta ortonormal con filtros de media banda fue desarrollado por Mallat [15]. A diferencia de la DWT que submuestrea los coeficientes de aproximación y detalle en cada

nivel de descomposición, la UDWT no incorpora estas operaciones; por lo tanto, los coeficientes de aproximación y detalle en cada nivel tienen la misma longitud que la señal original.

La transformada UDWT sobremuestrea los coeficientes de los filtros pasa altas y pasa bajas en cada nivel. La operación de sobremuestreo es equivalente a dilatar las *wavelets*. La resolución de los coeficientes UDWT decrece al incrementar los niveles de descomposición. Desde el punto de vista de banco de filtros, se mantienen las muestras tanto pares e impares y se continúan dividiendo las bandas bajas. La estructura de la transformada *wavelet* discreta no-decimada se muestra en la figura 4.1, donde H y L son idénticos a los filtros utilizados en la transformada *wavelet* discreta ortogonal, $\uparrow H$ y $\uparrow L$ representan un sobremuestreo diádico; es decir, que se realiza en potencias de 2.



(a) UDWT **(b)** IUDWT

Figura 4.1: Diagrama de la transformada UDWT.

La figura 4.1a muestra el diagrama para la transformada directa, en este ejemplo se utiliza una descomposición de dos niveles; se tiene una señal de entrada x que se divide en una aproximación ($A(1)$) y detalle ($D(1)$), al realizar otra descomposición, la aproximación pasa por dos filtros (las versiones sobremuestreadas de H y L), al final de la transformada se tienen una señal de aproximación y dos

señales para el detalle. En la figura 4.1b se muestra el diagrama para la transformada inversa, en el ejemplo nuevamente se utiliza una descomposición de dos niveles; a partir de una señal de aproximación y dos para el detalle, se reconstruye la señal x'.

4.1.2. Construcción del histograma

Sean $A = \{a(i), i = 1, 2, \ldots, L_A\}$ el audio original con L_A muestras, $F = \{f(i), i = 1, 2, \ldots, L_A\}$ los coeficientes de baja frecuencia en el dominio UDWT y $W = \{w(i), i = 1, 2, \ldots, L_W\}$ una marca binaria con L_W elementos. El histograma en dominio UDWT se representa como:

$$H = \{h(i) | i = 1, 2, \ldots, L\} \tag{4.2}$$

donde H es un vector que denota el histograma construido a partir de los coeficientes en el vector F y L es el número de bins del histograma, tal que $L \geq 4L_W$. $h(i) \geq 0$ denota el número de coeficientes en el i-ésimo bin y satisface que $\sum_{i=1}^{L} h(i) = L_A$.

Para calcular el histograma, se selecciona un rango de inserción $B = [(1 - \lambda)\mu, (1 + \lambda)\mu]$, donde μ representa la media estadística de F. La media de una señal dada se calcula al sumar los valores absolutos de las muestras y dividir por el total de éstos. Es decir:

$$\mu = \frac{1}{L_A} \sum_{i=1}^{L_A} |f(i)| \tag{4.3}$$

donde $f(i)$ es el i-ésimo coeficiente UDWT de baja frecuencia del vector F. La variable λ utilizada en el cálculo del rango de inserción es un número positivo que permite satisfacer que $h(i) \gg L$, Yang *et. al* [28] sugieren un valor de $\lambda = 0.6$.

Cada uno de los bins deberá tener un tamaño M que está dado por:

$$M = \frac{2\lambda\mu}{L} \tag{4.4}$$

En este esquema, la marca se inserta utilizando la forma del histograma, que se representa con una relación entre cuatro bins vecinos que conforman un grupo en el histograma. Esta relación se denota como β_k y se calcula de la siguiente forma:

$$\beta_k = \frac{h(k) + h(k+2)}{h(k+1) + h(k+3)} \tag{4.5}$$

donde β_k es la relación entre las frecuencias de los bins k, $k+1$, $k+2$ y $k+3$. En la siguiente sección se explicará con más detalle cómo estas relaciones permiten insertar la marca.

4.1.3. Algoritmos de inserción y detección

Inserción

La figura 4.2 muestra un diagrama de bloques del algoritmo de inserción. Primero, se aplica una transformada *wavelet* no-decimada al audio original. De los coeficientes resultantes, únicamente se toman aquellos de baja frecuencia y con éstos se construye un histograma, como se describe en la sección 4.1.2.

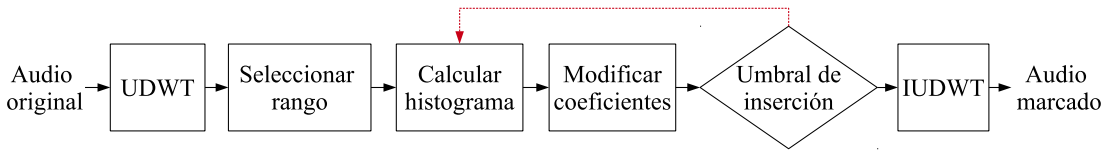


Figura 4.2: Diagrama en bloques del algoritmo de inserción de Yang.

Con los bins del histograma, se toman grupos de cuatro bins consecutivos y en cada uno de estos grupos se inserta un bit de la marca W. Dicha inserción se

logra al modificar los valores de los coeficientes que caen en los bins del grupo que se esté marcando.

Sean BIN_a, BIN_b, BIN_c y BIN_d cuatro bins consecutivos, y sus frecuencias representadas por N_a, N_b, N_c y N_d, respectivamente. Las siguientes reglas se emplean para insertar un bit de la marca, dependiendo de su valor:

$$\begin{cases} \frac{N_a+N_c}{N_b+N_d} \geq T & \text{si} \quad w(i)=1 \\ \frac{N_b+N_d}{N_a+N_c} \geq T & \text{si} \quad w(i)=0 \end{cases} \tag{4.6}$$

donde T es un umbral de inserción seleccionado, que se utiliza para controlar la robustez de la marca y la distorsión durante el proceso de inserción. A continuación se analizarán los dos casos de inserción; es decir, cuando se inserta un "1" o un "0".

Si se inserta un "1" y se cumple que $\frac{N_a+N_c}{N_b+N_d} \geq T$, no es necesario realizar ninguna operación; si no se cumple esta condición, se seleccionan aleatoriamente dos coeficientes, uno en el BIN_b y otro en el BIN_d para ser modificados de tal forma que ahora pertenezcan al BIN_a y BIN_c, respectivamente. Este proceso se repetirá hasta que se cumpla la condición que $\frac{N'_a+N'_c}{N'_b+N'_d} \geq T$. Las reglas para modificar los coeficientes son:

$$\begin{aligned} f'_b &= f_b(i) - M \quad 1 \leq i \leq I_b \\ f'_d &= f_d(i) - M \quad 1 \leq i \leq I_d \end{aligned} \tag{4.7}$$

donde $f_b(i)$ y $f_d(i)$ son los i-ésimos coeficientes UDWT sin modificar en BIN_b y BIN_d, $f'_b(i)$ y $f'_d(i)$ son las versiones modificadas correspondientes, M es el ancho de cada bin (fórmula 4.22), I_b e I_d se calculan de la siguiente manera:

$$I_b \geq \frac{I \cdot N_b}{N_b+N_d}, \quad I_d \geq \frac{I \cdot N_d}{N_b+N_d} \tag{4.8}$$

donde $I \geq \frac{T \cdot (N_b + N_d) - N_a - N_c}{1+T}$.

Si se inserta un "0" y se cumple que $\frac{N_b+N_d}{N_a+N_c} \geq T$, no es necesario realizar ninguna operación; si no se cumple esta condición, se seleccionan aleatoriamente dos coeficientes, uno en el BIN_a y otro en el BIN_c para ser modificados de tal forma que ahora pertenezcan al BIN_b y BIN_d, respectivamente. Este proceso se repetirá hasta que se cumpla la condición que $\frac{N'_b+N'_d}{N'_a+N'_c} \geq T$. Las reglas para modificar los coeficientes son:

$$\begin{aligned} f'_a &= f_a(i) + M \quad 1 \leq i \leq I_a \\ f'_c &= f_c(i) + M \quad 1 \leq i \leq I_c \end{aligned} \tag{4.9}$$

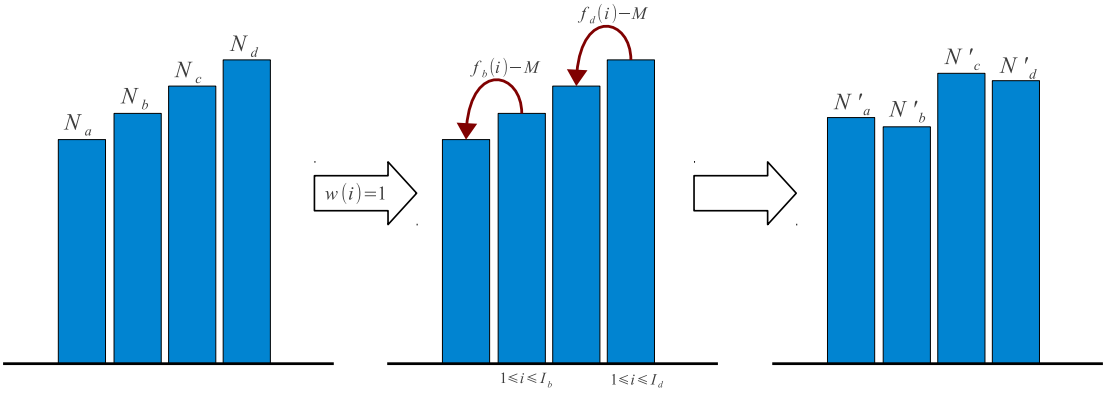
donde $f_a(i)$ y $f_c(i)$ son los i-ésimos coeficientes UDWT sin modificar en BIN_a y BIN_c, $f'_a(i)$ y $f'_c(i)$ son las versiones modificadas correspondientes, I_a e I_c se calculan de la siguiente manera:

$$I_a \geq \frac{I \cdot N_a}{N_a+N_c}, \quad I_c \geq \frac{I \cdot N_c}{N_a+N_c} \tag{4.10}$$

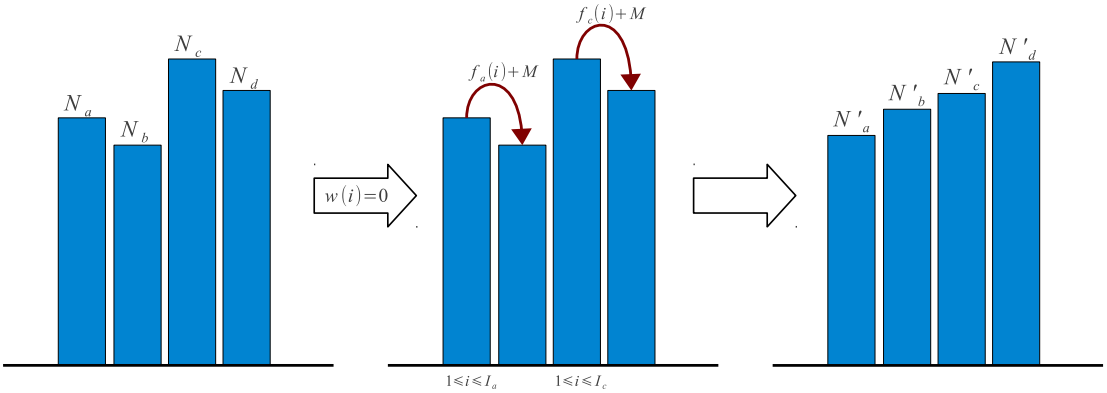
donde $I \geq \frac{T \cdot (N_a + N_c) - N_b - N_d}{1+T}$.

La figura 4.3 muestra un ejemplo con los dos casos de inserción. La figura 4.3a muestra las modificaciones que se realizan a los coeficientes que se encuentran dentro de los bins BIN_b y BIN_d (fórmulas 4.7 y 4.8). La figura 4.3b muestra las modificaciones que se realizan a los coeficientes que se encuentran dentro de los bins BIN_a y BIN_c (fórmulas 4.9 y 4.10).

Cuando todos los bits de la marca W han sido insertados en los coeficientes UDWT de bajas frecuencias, se realiza la transformada *wavelet* no-decimada in-



(a) Insertando "1"



(b) Insertando "0"

Figura 4.3: Ejemplo para los dos casos de inserción.

versa, tomando los coeficientes modificados y los coeficientes de altas frecuencias, para obtener el audio marcado $A^* = \{a^*(i), i = 1, 2, \ldots, L_A\}$.

Detección

La figura 4.4 muestra un diagrama en bloques del algoritmo de detección. Este proceso es ciego, ya que no se requiere información adicional para reconstruir la marca. El procedimiento se puede enumerar de la siguiente manera:

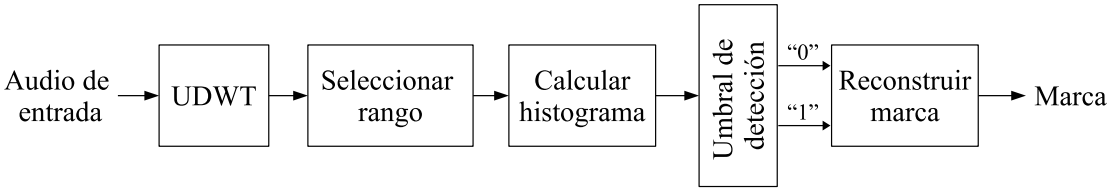


Figura 4.4: Diagrama en bloques del algoritmo de detección de Yang.

1. Se aplica la transformada *wavelet* discreta no-decimada al audio marcado A^*.

2. A partir de estos coeficientes, se obtiene un histograma H^* considerando sólo la banda de frecuencias bajas F^*. Este histograma se construye como se describe en la sección 4.1.2.

3. Se toman grupos de cuatro bins consecutivos y las frecuencias de cada uno de ellos se denotan como N_a^*, N_b^*, N_c^* y N_d^*. La marca se reconstruye con las siguientes reglas:

$$w^*(i) = \begin{cases} 1, & \text{si } \frac{N_a^*+N_c^*}{N_b^*+N_d^*} \geq 1 \\ 0, & \text{En caso contrario} \end{cases} \tag{4.11}$$

4.1.4. Discusión del esquema

El esquema base tiene algunas ventajas y desventajas que se describirán en esta sección. Los autores presentan como una ventaja utilizar las bajas frecuencias de los coeficientes UDWT, ya que éstos son invariantes a cambios; es decir, aunque un audio sea atacado, la forma del histograma se mantendrá. Sin embargo, no se cuenta con los elementos suficientes para replicar los resultados reportados por los autores y no es posible comprobar o descartar esta afirmación; en la sección 4.2.1 se explica uno de los problemas encontrados con esta transformada y que

causan que no se logre recuperar completamente la marca insertada aún sin haber aplicado ningún tipo de ataque a los audios marcados.

Otra de las ventajas que dicen tener los autores en su esquema, es que se garantiza que la mayor parte de la marca permanezca aún después de los ataques, dado que la inserción se realiza en las frecuencias bajas, donde se concentra la mayor parte de la información [5]. Sin embargo, por las mismas razones que en el punto anterior, no es posible confirmar o rechazar esta afirmación, puesto que no se tiene información suficiente para replicar los resultados reportados. Una de las ventajas más significativas del esquema es que, por modificar sólo un rango de valores, no se tiene un impacto perceptual significativo en los audios marcados; esta característica es de suma importancia para los esquemas de marcas de agua en audio y en éste se cumple.

Las desventajas más importantes de este esquema son las siguientes. El desempeño del mismo depende fuertemente del rango de valores que se seleccione para realizar la inserción de la marca, con las fórmulas empleadas para obtener este rango el esquema falla con algunos audios, por lo que no hay manera de garantizar el éxito de la recuperación de la marca para *cualquier* audio. Otra de las desventajas que tiene el esquema es que se asume que los coeficientes UDWT presentarán una distribución normal (figura 4.5), donde la media estadística (μ) representa el pico de la distribución; sin embargo, en la práctica esto no siempre ocurre, ya que es muy poco probable que los coeficientes de distintos audios presenten este tipo

[5] En las aplicaciones donde los audios son canciones, la mayor parte de la información se encuentra en las bajas frecuencias, porque muchos de los instrumentos en las canciones producen sonidos en este rango de frecuencias.

de distribución y la mayor parte de veces, la media estadística no corresponde con el pico de la distribución.

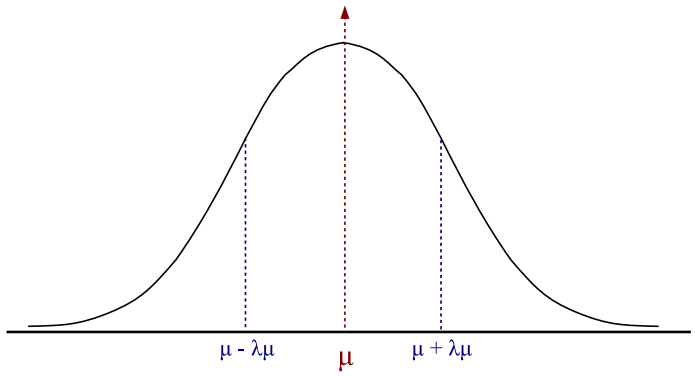


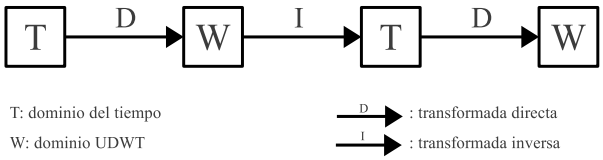
Figura 4.5: Suposición de una distribución normal en los coeficientes UDWT.

4.2. Esquema propuesto

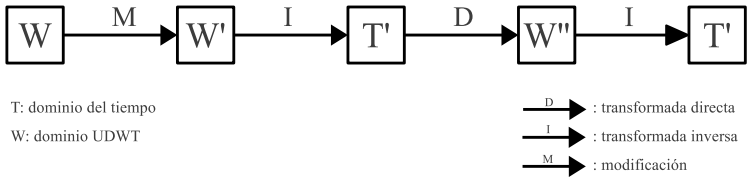
4.2.1. Dominios UDWT y DWT

Durante los experimentos que se realizaron para probar el esquema base, se encontraron comportamientos no esperados, por lo que se decidió analizar con más detalle la transformada UDWT. La problemática que se presentaba con esta transformada se muestra en la figura 4.6.

Al aplicar las transformadas directa e inversa, sin hacer modificaciones a los coeficientes en dominio UDWT, se tuvieron las mismas representaciones para T y W (figura 4.6a). Sin embargo, se presentaron resultados inesperados al realizar modificaciones en los coeficientes UDWT (figura 4.6b). A partir de los coeficientes W, se realizó una modificación W', al aplicar la transformada inversa se obtuvo



(a) Conversión sin modificaciones



(b) Conversión con modificaciones

Figura 4.6: Problemática encontrada al convertir al dominio UDWT.

una representación en el dominio del tiempo T', se aplicó una transformada directa y se obtuvieron los coeficientes W'' (distintos de W'), nuevamente se aplicó una transformada inversa y se regresó a la representación T'. Este simple experimento sirvió para concluir que, no existe una representación única en el dominio UDWT para una señal dada en el dominio del tiempo.

El problema hallado en el dominio UDWT ocurre porque esta transformada es una representación sobredeterminada de la señal original y contiene coeficientes relativos a muchas bases; por lo tanto, la operación inversa no es única [17]. Dicha sobredeterminación provoca que los coeficientes en la inserción sean distintos en la detección, con lo que el esquema falla aún sin ataques.

Como solución a la sobredeterminación del dominio UDWT, se planteó el uso de la transformada DWT. Con la transformada DWT se evita el problema de la sobredeterminación, ya que ésta es una transformada ortogonal; es decir, se

pueden obtener los mismos coeficientes en el dominio de la transformada, aún después de haber realizado las modificaciones durante el proceso de inserción. La idea general es sustituir la transformada UDWT por la DWT, de los coeficientes DWT tomar únicamente los de baja frecuencia, construir un histograma y aplicar los algoritmos propuestos por Yang *et al.*

A continuación se dará una explicación más amplia de la transformada DWT y cómo se modificó el esquema de marcas de agua para utilizar estos coeficientes.

4.2.2. Transformada DWT

La transformada *wavelet* discreta se basa en los filtros L, H y un operador de decimación binaria B_0 [17]. El filtro L es un filtro pasa bajas, definido por una secuencia denotada $\{l_n\}$. Se asume que el filtro satisface la relación de ortogonalidad interna:

$$\sum_n l_n l_{n+2j} = 0 \tag{4.12}$$

para todos los enteros $j \neq 0$, y que tiene una suma de cuadrados $\sum l_n^2 = 1$.
El filtro pasa bajas H está definido con la secuencia:

$$h_n = (-1)^n l_{1-n} \tag{4.13}$$

para toda n. El filtro H cumple las mismas relaciones de ortogonalidad interna que L y además, los filtros obedecen la relación de ortogonalidad mutua:

$$\sum_n l_n h_{-2j} = 0 \tag{4.14}$$

para todos los enteros j.

Los filtros que se construyen de la manera que se acaba de describir se conocen como *filtros espejo en cuadratura*. El operador de decimación binaria B_0 simplemente elige cada miembro impar de una secuencia, tal que:

$$(B_0 x)_j = x_{2j} \tag{4.15}$$

para todos los enteros j. Dadas las propiedades de ortogonalidad interna y mutua de los filtros, se tiene que el mapeo de una secuencia x al par de secuencias $(B_0 H_x, B_0 L_x)$ es una transformada ortogonal.

La transformada *wavelet* discreta deriva de un análisis de multiresolución, que se realiza de la siguiente manera. Supóngase que los datos originales están dados por la secuencia c:

$$c_n^J = c_n \quad \text{para} \quad n = 0, 1, \ldots, 2^J - 1 \tag{4.16}$$

Se define recursivamente para $j = J-1, J-2, \ldots, 0$ la aproximación A_j y el detalle D_j en el nivel j por:

$$A_j = B_0 L A_{j+1} \quad \text{y} \quad D_j = B_0 H A_{j+1} \tag{4.17}$$

Tanto A_j como D_j son secuencias de longitud 2^j. Se puede ver en la ecuación 4.17 que la aproximación en cada nivel se alimenta al siguiente nivel para encontrar la aproximación y el detalle.

Dado que el mapeo $(B_0 L, B_0 H)$ es una transformada ortogonal, es muy sencillo invertirla para encontrar A_{j+1} en términos de A_j y D_j, escribir la transformada como una matriz y calcular la transpuesta. La transformada inversa, R_0, está dada por:

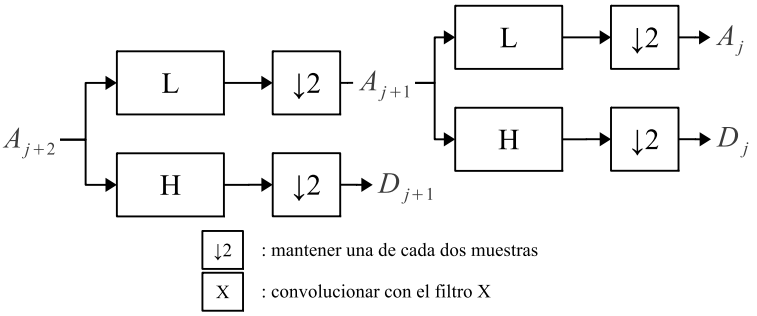
$$A_{j+1} = R_0(A_j, D_j) \quad \text{para cada} \quad j \tag{4.18}$$

La transformada *wavelet* discreta se obtiene al continuar este proceso, calculando el detalle y la aproximación en cada nivel, hasta alcanzar el nivel cero, de tal manera que la secuencia original se transforma ortogonalmente a la secuencia de secuencias $D_{J-1}, D_{J-2}, \ldots, D_0, A_0$ de longitud total 2^J.

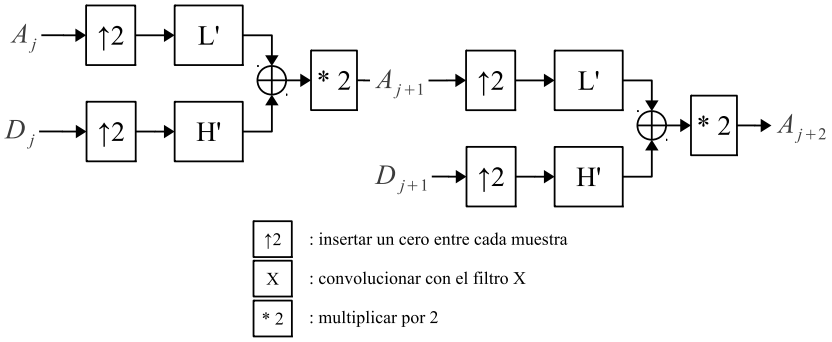
En la figura 4.7 se puede apreciar un diagrama de cómo se calculan tanto la transformada *wavelet* discreta como la transformada inversa [15], donde L y H son los filtros anteriormente descritos. En la figura 4.7a se muestra la transformada directa, las operaciones descritas en la ecuación 4.17 están representadas en la figura como la convolución con los filtros (L y H) y la decimación (B_0) como $\downarrow 2$.

En el ejemplo, se empieza con la señal original (A_{j+2}) que se convoluciona con los filtros y posteriormente se hace la decimación, obteniendo las señales A_{j+1} y D_{j+1} que son la aproximación y el detalle del primer nivel, la aproximación se alimenta al siguiente nivel para obtener las señales A_j y D_j que son la aproximación y el detalle del segundo nivel. En la figura 4.7b se ejemplifica la transformada inversa dada en la ecuación 4.18; las señales A_j y D_j se sobremuestrean ($\uparrow 2$) y posteriormente se convolucionan con los filtros L' y H', respectivamente, ambas señales se suman y multiplican por dos para obtener la señal A_{j+1}; con esta última y la señal D_{j+1} se aplica el mismo proceso para obtener la señal original A_{j+2}.

En la siguiente subsección se detallará cómo a partir de los coeficientes obtenidos con la transformada DWT se construye un histograma, que permite aplicar los algoritmos de inserción y detección propuestos por Yang, *et al.*



(a) DWT



(b) IDWT

Figura 4.7: Diagrama de la transformada DWT.

4.2.3. Construcción del histograma

Al igual que en el dominio de la transformada UDWT, sean $A = \{a(i), i = 1, 2, \ldots, L_A\}$ el audio original con L_A muestras, $F = \{f(i), i = 1, 2, \ldots, L_F\}$ los coeficientes de baja frecuencia, esta vez en el dominio DWT y $W = \{w(i), i = 1, 2, \ldots, L_W\}$ una marca binaria con L_W elementos. El histograma en dominio DWT se representa como:

$$H = \{h(i) | i = 1, 2, \ldots, L\} \tag{4.19}$$

donde H es un vector que denota el histograma construido a partir de los coeficientes en el vector F y L es el número de bins del histograma, tal que

$L \geq 4L_W$. $h(i) \geq 0$ denota el número de coeficientes en el i-ésimo bin y satisface que $\sum_{i=1}^{L} h(i) = L_A$.

Para calcular el histograma, se selecciona un rango de inserción $B = [Moda - K \cdot MAD, Moda + K \cdot MAD]$, donde $Moda$ es el elemento más frecuente de los valores absolutos en el vector F. Supóngase que la función $mode()$ permite obtener el valor más frecuente de un vector dado X, la $Moda$ que se utiliza para el cálculo del histograma está dada de la siguiente manera:

$$Moda = mode(|X|) \tag{4.20}$$

donde $|X|$ son los valores absolutos del vector. La variable K es un número positivo que permite controlar el tamaño del rango, de manera experimental se determinó que un rango adecuado para K es de $[2 : 4]$; la medida de dispersión MAD es la Desviación Media Absoluta (MAD, por sus siglas en inglés) y está dada por:

$$MAD = \frac{1}{N} \sum_{i=1}^{N} |x_i - mode(|X|)| \tag{4.21}$$

donde N es la longitud del vector X. Cada uno de los bins del histograma deberá tener un tamaño M que está dado por:

$$M = \frac{2 \cdot K \cdot MAD}{L} \tag{4.22}$$

De forma similar al esquema de Yang, en éste la marca se inserta utilizando la forma del histograma, que se representa con la relación en un grupo de cuatro bins vecinos en el hisotgrama (ecuación 4.5). A continuación se explicarán los algoritmos de inserción y detección que se utilizan en este esquema.

4.2.4. Algoritmos de inserción y detección

Inserción

En la figura 4.8 se muestra un diagrama de bloques del algoritmo de inserción del esquema propuesto. Primero se aplica una transformada *wavelet* discreta al audio original. De los coeficientes resultantes, únicamente se toman aquellos de baja frecuencia y con éstos se construye un histograma, como se describe en la sección 4.2.3.

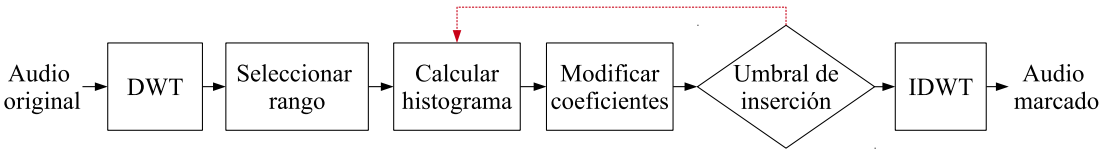


Figura 4.8: Diagrama en bloques del algoritmo de inserción propuesto.

Los pasos para modificar los coeficientes y el umbral de inserción son iguales que en el esquema de Yang, *et al.* Se forman grupos de cuatro bins consecutivos y en cada uno de estos grupos se inserta un bit de la marca W. Dicha inserción se logra al modificar los coeficientes según las ecuaciones 4.6, 4.7, 4.8, 4.9 y 4.10, según sea el bit que se está insertando.

Una vez que todos los bits de la marca W han sido insertados en los coeficientes DWT de bajas frecuencias, se realiza una transformada *wavelet* discreta inversa, tomando tanto los coeficientes modificados como los de altas frecuencias, para obtener el audio marcado $A^* = \{a^*(i) = 1, 2, \ldots, L_A\}$.

Detección

La figura 4.9 muestra un diagrama en bloques del algoritmo de detección. Para extraer la marca, este proceso utiliza el audio marcado y además el rango de inserción que se utilizó cuando se insertó la marca. El procedimiento se puede enumerar de la siguiente manera:

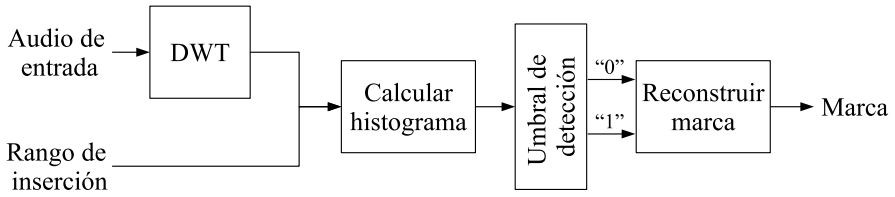


Figura 4.9: Diagrama en bloques del algoritmo de detección propuesto.

1. Se aplica la transformada *wavelet* discreta al audio marcado A^*.

2. Tomando únicamente los coeficientes de bajas frecuencias F^* y considerando los rangos de inserción, se construye un histograma H^*. Al tener los rangos de inserción, únicamente se cuentan los coeficientes de bajas frecuencias que caen dentro de los bins correspondientes y de esta manera se obtienen las frecuencias.

3. Se toman grupos de cuatro bins consecutivos y las frecuencias de cada uno de ellos se denotan como N_a^*, N_b^*, N_c^* y N_d^*. La marca se reconstruye con la regla dada en la ecuación 4.11.

4.2.5. Discusión del esquema

Aunque este esquema soluciona algunos de los inconvenientes del esquema propuesto por Yang, *et al.*, tiene algunas ventajas y desventajas que se describirán en esta sección.

La primera ventaja es que se soluciona el problema de la sobredeterminación que existe en el dominio UDWT, al utilizar en su lugar la transformada DWT, misma que permite obtener los mismos coeficientes en la inserción y en la detección, sin aplicar ataques; y coeficientes muy parecidos aún después de los ataques. Otra ventaja es que, por la forma de calcular el rango de inserción, se logra encontrar el pico de la distribución, lo que permite tener un histograma simétrico y que contribuye con la robustez del esquema. Finalmente, se mantiene la ventaja de poco impacto perceptual en los audios marcados, debido a que solamente se utiliza un rango de las bajas frecuencias.

No obstante, este esquema tiene como desventaja que, al construir el histograma a partir de los rangos utilizados en la inserción, se logra robustez ante los ataques de desincronización pero no ante los ataques comunes de procesamiento de señales. En general, los ataques de desincronización eliminan muestras o modifican su posición en el tiempo; esto repercute en el número de muestras dentro de cada bin del histograma, pero la forma del histograma reconstruido es muy parecida a la del histograma en la inserción, por lo que se tiene robustez ante este tipo de ataques. Los ataques comunes de procesamiento de señales por su parte, no cambian de posición las muestras sino que modifican sus coeficientes; esto hace que las muestras se recorran a otros rangos y el histograma reconstruido se recorra

con respecto al histograma de la inserción, por lo que el esquema propuesto no tiene mucha robustez contra este tipo de ataques.

En el siguiente capítulo se describirán los experimentos realizados para evaluar ambos esquemas y se explicará con más detalle el porqué de estos resultados.

Capítulo 5

Experimentación y resultados

5.1. Banco de pruebas

Aunque existen algunos *benchmarks* para evaluar el desempeño de los esquemas de marcas de agua para audio, aún no se han desarrollado *corpus* donde se tengan audios para probar los algoritmos. Por lo tanto, se construyó un banco de pruebas, recopilando audios con las características adecuadas para evaluar el esquema propuesto.

Se buscaron audios en formato FLAC, que posteriormente se convirtieron a formato WAVE de 16 bits por muestra, con frecuencia de muestreo de 44.1 kHz. Los audios se convirtieron con la herramienta de software libre y código abierto Audacity®1.3.13. Además de hacer la conversión de formatos, con esta herramienta también se recortaron los audios para tener pistas con duración de 20 s. Se eligieron estas características y duración de los audios, porque fue la información reportada por Yang *et al.* [28] y se buscó tener condiciones similares para comparar los esquemas.

El banco de pruebas que se construyó, consiste en 1 canción por género y 5 géneros distintos: jazz, orquesta, pop, rock y vocal. Las canciones fueron las siguientes: "Aquarela do Brasil" de Trío da Paz para jazz, "Allegro de Concierto de Violin N.4" de Mozart para orquesta, "So long Jimmy" de James Blunt para pop, "Ain't it fun" de Guns and Roses para rock y "Baby it's cold outside" de Norah Jones para vocal. Estos géneros son los más comunes para evaluar esquemas de marcas de agua para audio, ya que cada uno presenta características distintas en cuando a los instrumentos y las frecuencias en las canciones. Con respecto a las marcas, se crearon marcas aleatorias de 60 bits que fueron generadas con la herramienta Matlab (R2011b).

5.2. Métricas de evaluación

Como se mencionó en la subsección 2.2.3 del capítulo 2, se necesita conocer el rendimiento de los esquemas de marcas de agua para audio. Para los esquemas que se evalúan en este trabajo, se requiere conocer su desempeño con respecto a la robustez de la marca insertada y la fidelidad existente entre el audio marcado y el audio original.

Para evaluar la robustez de la marca, se emplea la *tasa de bits erróneos* (BER), esta métrica es la razón entre el número de bits erróneos extraídos del audio marcado o atacado y el total de bits insertados; la fórmula para calcularla es la siguiente:

$$BER = \frac{B}{L} * 100\,\% \tag{5.1}$$

donde B es el número de bits detectados erróneamente y se define como:

$$B = \sum_{n=1}^{L} \begin{cases} 1, & W'_n \neq W_n \\ 0, & W'_n = W_n \end{cases} \tag{5.2}$$

L es la longitud de la marca, W_n corresponde al n-ésimo bit de la marca insertada y W'_n al n-ésimo bit de la marca recuperada. Un resultado igual a uno significa que todos los bits detectados fueron erróneos, mientras que un valor igual a cero significa que se logró una recuperación perfecta.

Para evaluar la fidelidad de los esquemas se utiliza la *relación señal a ruido pico* (PSNR), esta métrica permite medir la relación señal a ruido máxima en una señal de audio, comparando qué tan distinta es una señal modificada con respecto a una señal original y se calcula de la siguiente manera:

$$PSNR(A, A^*) = 10 \log_{10} \frac{A^2_{\text{pico}}}{\sigma_e^2} \tag{5.3}$$

donde σ_e^2 se define como:

$$\sigma_e^2 = \left(\frac{1}{L_A}\right) \sum_{i=1}^{L_A} \Big(a(i) - a^*(i)\Big)^2 \tag{5.4}$$

donde L_A es la longitud del audio original, $a(i)$ es la magnitud del audio original A en el tiempo i; $a^*(i)$ es la magnitud del audio marcado A^* en el tiempo i y A^2_{pico} denota el valor máximo de la señal original elevado al cuadrado. Un valor de PSNR mayor indica que el audio tiene mayor similitud con el audio original y un valor menor indica lo contrario.

5.3. Experimentos

5.3.1. Evaluación sin ataques

Estos experimentos sirvieron para evaluar el impacto perceptual del esquema, lo que afecta la fidelidad de los audios marcados. Además, dichos experimentos sirvieron para probar la robustez del esquema propuesto bajo un escenario ideal, donde no existen ataques durante la transmisión.

La figura 5.1 muestra un diagrama con el procedimiento general para llevar a cabo estos experimentos. Primero se toma un audio original y se le aplica el algoritmo de inserción, con lo que se obtiene un audio marcado y a éste se le aplica el algoritmo de detección que proporciona dos resultados, uno es el BER que permite medir la robustez y el segundo es el PSNR que mide la fidelidad.

Figura 5.1: Procedimiento general de la evaluación sin ataques.

5.3.2. Evaluación con ataques

Estos experimentos sirvieron para probar la robustez del esquema con ataques de distintos tipos. También se evaluó la fidelidad de los audios después de haber aplicado cada uno de los ataques.

La figura 5.2 muestra un diagrama con el procedimiento general para llevar

a cabo estos experimentos. Primero se toma un audio original y se le aplica el algoritmo de inserción con lo que se obtiene un audio marcado, éste se ataca y posteriormente se le aplica el algoritmo de detección que proporciona dos resultados, uno es el BER que permite medir la robustez de la marca ante ese ataque en particular y el segundo es el PSNR que mide la fidelidad.

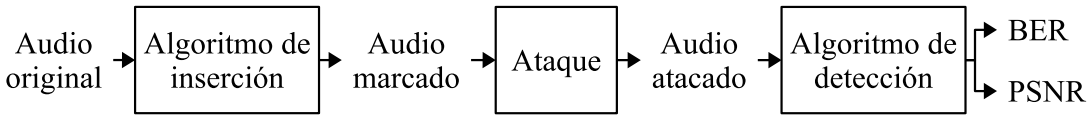


Figura 5.2: Procedimiento general de la evaluación con ataques.

El esquema propuesto se probó sometiendo los audios marcados ante dos tipos de ataques:

1. Ataques comunes de procesamiento de señales:

   a) Compresión MP3. Se realizó una compresión MPEG con la herramienta CoolEdit Pro®, se aplicaron tasas de 32, 64 y 128 kbps.

   b) Ruido aditivo. Se agregó ruido blanco gaussiano con el *benchmark* Stirmark, utilizando intensidades de ruido de 100, 200, 300 y 400 dB.

   c) Re-muestreo. Se modificó la tasa de muestreo de los audios a 16 y 32 kHz, y se regresó a la tasa original de 44.1 kHz; este ataque se aplicó con CoolEdit.

   d) Filtrado pasa bajas. Se aplicaron filtros pasa bajas con frecuencias de 1, 2, 3, 4, 6 y 8 kHz, empleando el *benchmark* Stirmark.

2. Ataques de desincronización:

   a) Recortes (10 %). Se removieron 10 % de las muestras del audio al inicio del mismo, para ello se utilizó el *benchmark* Stirmark.

   b) Jittering (300,500,1000,2000,3000). Con este ataque se recorta una muestra cada 300, 500, 1000, 2000 y 3000 muestras, para esto se empleó el *benchmark* Stirmark.

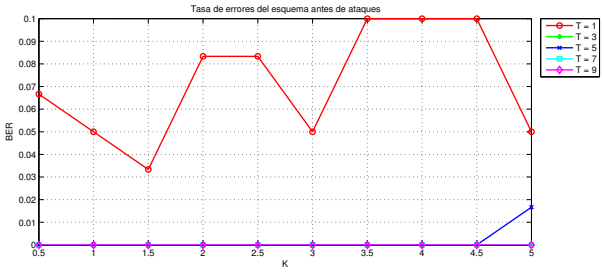
   c) Modificación de la escala de tiempo ($\pm$ 1,2,3,4 %). Los porcentajes positivos realizan un incremento en la velocidad del audio, mientras que los porcentajes negativos decrementan la velocidad, este ataque se realizó con la herramienta CoolEdit.

5.4. Resultados

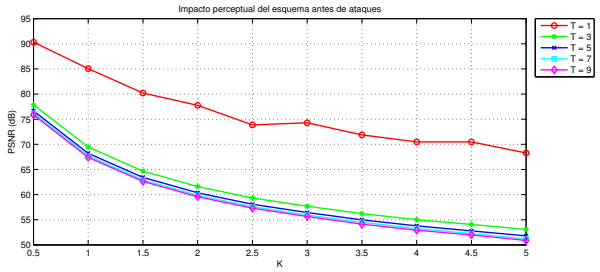
5.4.1. Sin ataques

Como se mencionó anteriormente, estos experimentos sirvieron para evaluar el impacto perceptual del esquema en los audios marcados. Asimismo, estas pruebas ayudaron a determinar el impacto que tienen la fuerza de inserción (T) y los valores para los rangos de inserción (K) en la calidad de los audios marcados y la tasa de errores del esquema sin ataques.

Se probaron umbrales de inserción T=1–9 y rangos de inserción K=0.5–5, aplicando el procedimiento de la figura 5.1. Las figuras 5.3 a 5.7 presentan los resultados de BER y PSNR obtenidos con estas configuraciones de T y K.

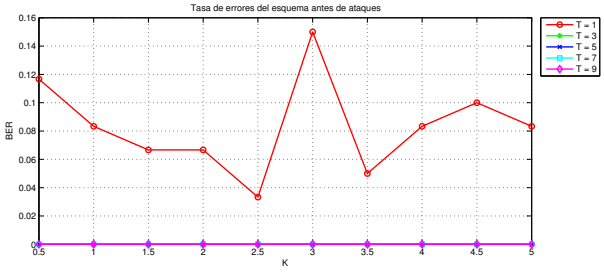


(a) BER.

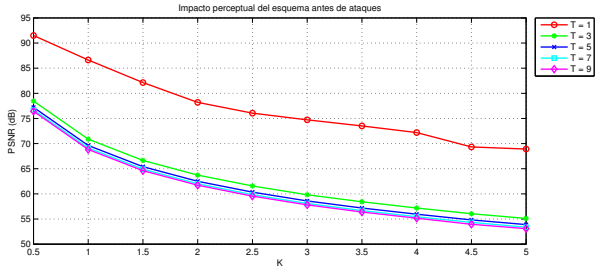


(b) PSNR.

Figura 5.3: Desempeño del esquema propuesto para el audio de Jazz.

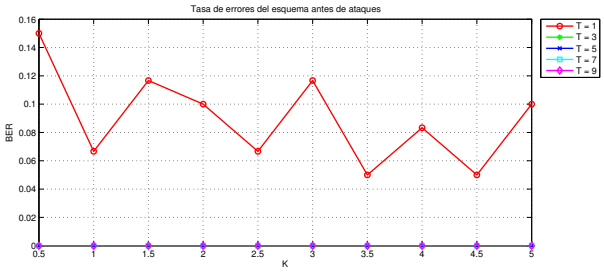


(a) BER.

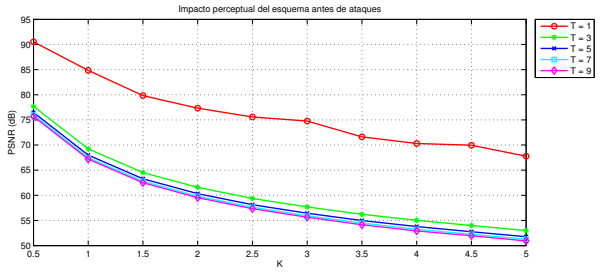


(b) PSNR.

Figura 5.4: Desempeño del esquema propuesto para el audio de Orquesta.



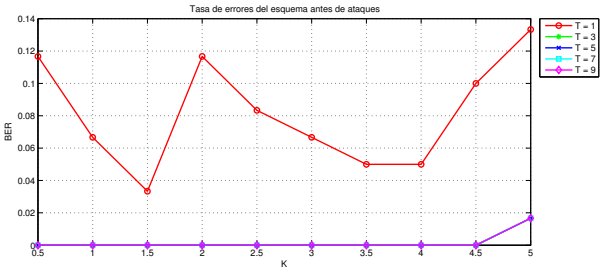
(a) BER.



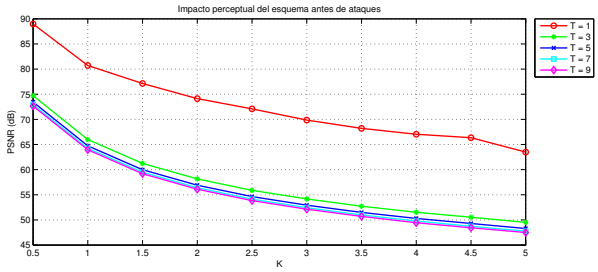
(b) PSNR.

Figura 5.5: Desempeño del esquema propuesto para el audio de Pop.

Estos experimentos ayudaron a restringir los valores de los parámetros para

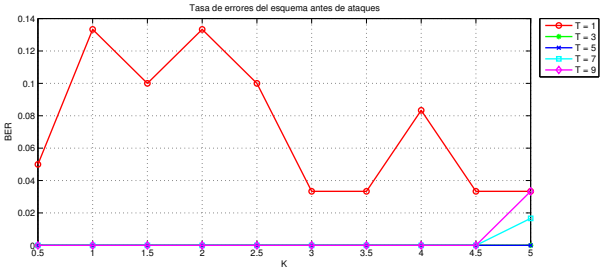


(a) BER.

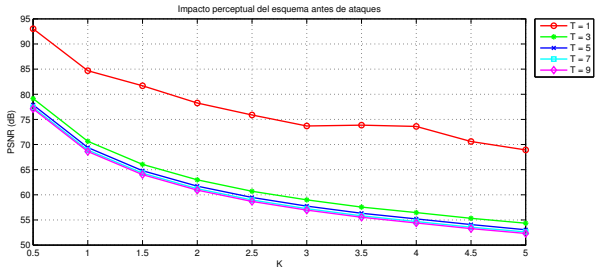


(b) PSNR.

Figura 5.6: Desempeño del esquema propuesto para el audio de Rock.



(a) BER.



(b) PSNR.

Figura 5.7: Desempeño del esquema propuesto para el audio de Vocal.

utilizar en las pruebas con ataques. Se descartaron umbrales para $T \leq 1$ y rangos de inserción para $K \geq 5$. Se comprobó también que el impacto perceptual que produce el esquema en los audios marcados es pequeño, pues aún con valores de T y K mayores, se mantiene un PSNR arriba de 50 dB.

A continuación se presenta el desempeño que tuvo el esquema, atacando los audios marcados.

5.4.2. Con ataques

Ataques comunes de procesamiento de señales

Se ralizaron pruebas con audios que fueron marcados previamente, utilizando distintos rangos de inserción y posteriormente fueron sometidos a los ataques comunes de procesamiento de señales mencionados previamente. A continuación se analizarán los resultados de cada uno de los ataques.

Compresión MP3. En la tabla 5.1 se muestran los mejores resultados de BER obtenidos en cada uno de los audios, utilizando las distintas configuraciones del ataque; además, se indican la fuerza de inserción (T) y el rango de inserción (K) con los que se obtuvieron los resultados, finalmente se indica el PSNR obtenido después de aplicar los ataques. La figura 5.8 muestra el comportamiento del PSNR para los ataques MP3; en todos los audios se puede observar que los valores aumentan conforme se incrementa la calidad de compresión.

Tabla 5.1: Resultados de los ataques MP3.

Audio	Ataque	T	K	BER	PSNR (dB)
Jazz	MP3 32 kbps	5	2	0.4667	41.6054
	MP3 64 kbps	5	4	0.3500	43.0515
	MP3 128 kbps	7	2	0.3167	48.9223
Orquesta	MP3 32 kbps	5	3	0.4333	40.2940
	MP3 64 kbps	5	4	0.4667	44.9428
	MP3 128 kbps	7	2	0.4167	49.8045
Pop	MP3 32 kbps	5	4	0.3667	39.0819
	MP3 64 kbps	5	3	0.4167	42.5483
	MP3 128 kbps	7	2	0.2000	48.5449
Rock	MP3 32 kbps	5	3	0.3833	30.9624
	MP3 64 kbps	5	2	0.4167	34.9601
	MP3 128 kbps	7	3	0.4167	40.6910
Vocal	MP3 32 kbps	7	4	0.3833	39.8392
	MP3 64 kbps	5	3	0.3167	43.5195
	MP3 128 kbps	5	4	0.3333	47.8737

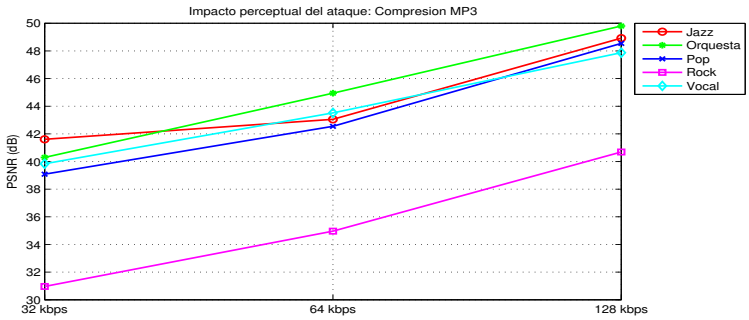


Figura 5.8: Comportamiento del PSNR para los ataques MP3.

Ruido aditivo. Similar al caso anterior, en la tabla 5.2 se muestran los mejores resultados de BER obtenidos en cada audio para los ataques de ruido aditivo; se indican los parámetros T, K y los valores de PSNR correspondientes. La figura 5.9 muestra cómo los valores de PSNR disminuyen conforme se aumenta la cantidad de ruido que se le agrega al audio.

Tabla 5.2: Resultados de los ataques de Ruido.

Audio	Ataque	T	K	BER	PSNR (dB)
Jazz	Ruido aditivo 100	5	4	0.1833	49.9664
	Ruido aditivo 200	5	2	0.5167	45.7769
	Ruido aditivo 300	5	2	0.5167	42.3193
	Ruido aditivo 400	5	4	0.5167	39.7420
Orquesta	Ruido aditivo 100	5	4	0.1500	52.3858
	Ruido aditivo 200	5	4	0.5833	48.0258
	Ruido aditivo 300	5	2	0.4833	45.1395
	Ruido aditivo 400	5	4	0.5000	42.5143
Pop	Ruido aditivo 100	5	4	0.0833	51.4005
	Ruido aditivo 200	5	4	0.4500	47.5845
	Ruido aditivo 300	5	4	0.4833	44.5836
	Ruido aditivo 400	5	4	0.4000	42.2547
Rock	Ruido aditivo 100	5	3	0.0667	50.1164
	Ruido aditivo 200	5	4	0.3833	45.3938
	Ruido aditivo 300	5	4	0.4667	42.6161
	Ruido aditivo 400	5	4	0.3667	40.4049
Vocal	Ruido aditivo 100	5	4	0.1167	51.7240
	Ruido aditivo 200	5	4	0.6000	47.3029
	Ruido aditivo 300	5	4	0.4667	44.1414
	Ruido aditivo 400	5	4	0.5000	41.8625

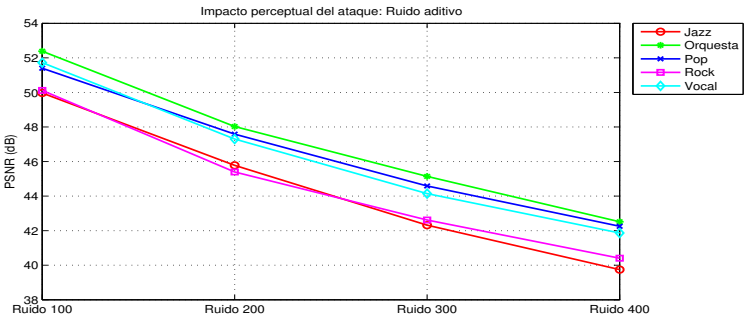


Figura 5.9: Comportamiento del PSNR para los ataques de Ruido.

Re-muestreo. En la tabla 5.3 se presentan los mejores resultados de BER obtenidos en cada audio para los ataques de re-muestreo; se indican los parámetros T, K y los valores de PSNR correspondientes. La figura 5.10 muestra cómo los valores de PSNR aumentan al disminuir la severidad de 16 a 32 kHz, excepto para el caso del Rock donde el ataque tiene un impacto perceptual tal, que los valores de PSNR no son los esperados.

Tabla 5.3: Resultados de los ataques de Re-muestreo.

Audio	Ataque	T	K	BER	PSNR (dB)
Jazz	Re-muestreo 16 kHz	7	3	0.4000	32.9372
	Re-muestreo 32 kHz	7	3	0.4000	31.2871
Orquesta	Re-muestreo 16 kHz	7	2	0.4167	25.1593
	Re-muestreo 32 kHz	7	4	0.5667	24.8995
Pop	Re-muestreo 16 kHz	5	3	0.2500	27.6318
	Re-muestreo 32 kHz	7	4	0.3500	27.2034
Rock	Re-muestreo 16 kHz	7	2	0.5500	18.9791
	Re-muestreo 32 kHz	7	4	0.3667	19.9622
Vocal	Re-muestreo 16 kHz	5	4	0.2833	32.1079
	Re-muestreo 32 kHz	5	4	0.3333	30.4146

Filtrado pasa bajas. En la tabla 5.4 se presentan los mejores resultados de BER obtenidos en cada audio para los ataques de filtrado pasa bajas; se indican los parámetros T, K y los valores de PSNR correspondientes. La figura 5.11 muestra cómo los valores de PSNR incrementan al aumentar los umbrales de los filtros.

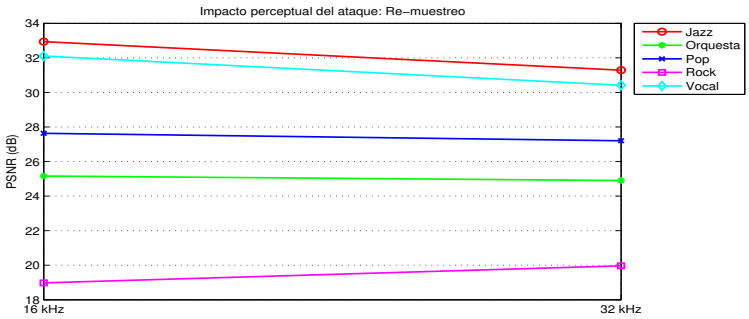


Figura 5.10: Comportamiento del PSNR para los ataques de Re-muestreo.

Tabla 5.4: Resultados de los ataques de Filtrado.

Audio	Ataque	T	K	BER	PSNR (dB)
Jazz	Pasa bajas 1 kHz	5	4	0.4667	23.2201
	Pasa bajas 2 kHz	5	4	0.4833	27.8682
	Pasa bajas 3 kHz	5	3	0.4833	30.8469
	Pasa bajas 4 kHz	5	4	0.4333	33.0901
	Pasa bajas 6 kHz	5	3	0.3333	36.6508
	Pasa bajas 8 kHz	5	3	0.2833	39.6062
Orquesta	Pasa bajas 1 kHz	5	4	0.6000	18.5579
	Pasa bajas 2 kHz	7	3	0.4667	21.3634
	Pasa bajas 3 kHz	5	3	0.4167	23.7873
	Pasa bajas 4 kHz	5	2	0.4500	25.9677
	Pasa bajas 6 kHz	5	4	0.4500	29.8105
	Pasa bajas 8 kHz	7	3	0.3667	33.2027
Pop	Pasa bajas 1 kHz	5	3	0.3500	20.1408
	Pasa bajas 2 kHz	5	2	0.4167	23.1500
	Pasa bajas 3 kHz	7	3	0.4167	25.5280
	Pasa bajas 4 kHz	5	4	0.4667	27.5628
	Pasa bajas 6 kHz	5	3	0.3833	31.0362
	Pasa bajas 8 kHz	7	4	0.2333	34.0191
Rock	Pasa bajas 1 kHz	7	4	0.6167	15.0294
	Pasa bajas 2 kHz	5	4	0.4833	16.9774
	Pasa bajas 3 kHz	7	4	0.4833	18.7038
	Pasa bajas 4 kHz	7	2	0.5333	20.3078
	Pasa bajas 6 kHz	7	4	0.4500	23.2878
	Pasa bajas 8 kHz	5	3	0.4167	26.0915
Vocal	Pasa bajas 1 kHz	7	3	0.4667	22.6435
	Pasa bajas 2 kHz	5	2	0.4833	25.9709
	Pasa bajas 3 kHz	7	2	0.4833	28.4312
	Pasa bajas 4 kHz	7	4	0.4333	30.4615
	Pasa bajas 6 kHz	5	4	0.4500	33.8714
	Pasa bajas 8 kHz	7	3	0.2167	36.8396

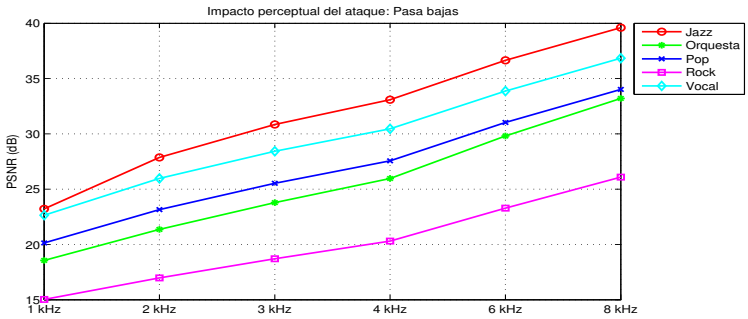


Figura 5.11: Comportamiento del PSNR para los ataques de Filtrado.

En las tablas 5.1 a 5.4 se puede observar que los valores de BER son elevados; es decir, el esquema no es muy robusto ante los ataques comunes de procesamiento de señales con los que se realizaron las pruebas. Lo anterior sucede por la manera de realizar la detección en el esquema propuesto.

Para realizar la inserción de la marca, se selecciona un rango de valores a modificar y en la detección se emplea ese mismo rango para identificar los bits de la marca. Esta estrategia asume que los ataques no afectan los valores de las muestras, sino sus posiciones en el tiempo; no obstante, los ataques comunes de procesamiento de señales modifican los valores de las muestras sin afectar su ubicación. Cuando estos valores cambian, la forma del histograma que se reconstruye es completamente distinta a la que se utilizó para realizar la inserción y por este motivo, la detección falla.

Aunque es deseable que los esquemas de marcas de agua para audio sean robustos ante ataques comunes de procesamiento de señales, el objetivo de esta investigación se centra en los ataques de desincronización, que se presentan a continuación.

Ataques de desincronización

Similar que en los ataques anteriores, para realizar éstos se tomaron audios previamente marcados utilizando distintos rangos de inserción y posteriormente fueron sometidos ante los ataques de desincronización mencionados previamente. En seguida se presentarán los resultados de cada uno de estos ataques.

Recortes. En la tabla 5.5 se presentan los resultados de robustez (BER) e impacto perceptual (PSNR), al quitar el 10 % de las muestras al inicio de los audios; asimismo, se muestran los parámetros T y K con los que se obtuvieron dichos valores. El esquema propuesto logra una detección perfecta de la marca en todos los audios, aplicando este ataque.

Tabla 5.5: Resultados del ataque Recortes.

Audio	Ataque	T	K	BER	PSNR (dB)
Jazz	Recortes 10 %	5	2	0.0000	34.1399
Orquesta	Recortes 10 %	5	2	0.0000	31.3681
Pop	Recortes 10 %	5	2	0.0000	29.4676
Rock	Recortes 10 %	5	2	0.0000	21.8692
Vocal	Recortes 10 %	5	2	0.0000	31.8896

Jittering. La tabla 5.6 muestra los resultados de BER y PSNR para los ataques de Jittering aplicados a los audios de prueba; además, se presentan los parámetros T y K con los que se obtuvieron estos valores. De manera similar que en el ataque anterior, se logra una detección perfecta en todos los audios.

Modificación de la escala de tiempo (TSM). En la tabla 5.7 se presentan los resultados de BER y PSNR para los audios de prueba, aplicando los ataques de modificación de escala de tiempo; asimismo, se indican los parámetros T y K con los que se alcanzaron dichos valores. Los ataques con porcentaje positivo realizan

Tabla 5.6: Resultados para los ataques de Jittering.

Audio	Ataque	T	K	BER	PSNR (dB)
Jazz	Jittering (1/300)	5	2	0.0000	37.3076
	Jittering (1/500)	5	2	0.0000	37.3330
	Jittering (1/1000)	5	2	0.0000	37.4655
	Jittering (1/2000)	5	2	0.0000	37.1451
	Jittering (1/3000)	5	2	0.0000	37.2703
Orquesta	Jittering (1/300)	5	2	0.0000	34.5318
	Jittering (1/500)	5	2	0.0000	34.5222
	Jittering (1/1000)	5	2	0.0000	34.5658
	Jittering (1/2000)	5	2	0.0000	34.4802
	Jittering (1/3000)	5	2	0.0000	34.5683
Pop	Jittering (1/300)	5	2	0.0000	32.0481
	Jittering (1/500)	5	2	0.0000	31.9918
	Jittering (1/1000)	5	2	0.0000	32.1334
	Jittering (1/2000)	5	2	0.0000	32.1733
	Jittering (1/3000)	5	2	0.0000	32.0608
Rock	Jittering (1/300)	5	2	0.0000	24.9939
	Jittering (1/500)	5	2	0.0000	24.9840
	Jittering (1/1000)	5	2	0.0000	24.9681
	Jittering (1/2000)	5	2	0.0000	24.8732
	Jittering (1/3000)	5	2	0.0000	24.9321
Vocal	Jittering (1/300)	5	4	0.0000	35.0584
	Jittering (1/500)	5	2	0.0000	35.1776
	Jittering (1/1000)	5	2	0.0000	35.1069
	Jittering (1/2000)	5	2	0.0000	35.2101
	Jittering (1/3000)	5	2	0.0000	35.1351

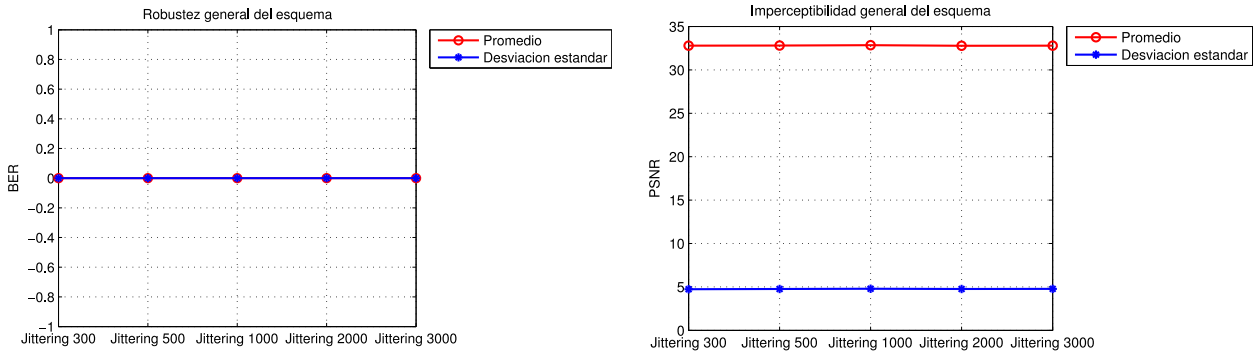
un incremento en el tiempo del audio, mientras que los ataques con porcentaje negativo lo decrementan. Aunque para estos ataques no se logra la recuperación perfecta en todos los audios, los resultados de BER son bajos, por lo que se tiene robustez ante estas operaciones. Al realizar estos experimentos, se tuvieron dos casos en los que el esquema falló. No se logró una detección adecuada de la marca para el audio de Rock con el ataque TSM +1 %, ni para el audio de Vocal con el ataque TSM -1 %, porque dichos ataques modificaron gravemente las muestras en estos audios y los errores en la detección fueron mayores al 3 %.

Tabla 5.7: Resultados para los ataques de TSM.

Audio	Ataque	T	K	BER	PSNR (dB)
Jazz	TSM +1 %	5	2	0.0000	24.6218
	TSM +2 %	7	2	0.0000	26.4189
	TSM +3 %	5	2	0.0000	24.0972
	TSM +4 %	5	2	0.0167	28.5442
	TSM -1 %	5	2	0.0000	21.8351
	TSM -2 %	7	2	0.0000	21.3610
	TSM -3 %	7	3	0.0000	20.4077
	TSM -4 %	5	2	0.0167	21.8576
Orquesta	TSM +1 %	5	2	0.1000	28.3174
	TSM +2 %	5	3	0.1333	20.0507
	TSM +3 %	5	2	0.0167	24.7336
	TSM +4 %	7	2	0.0333	27.6522
	TSM -1 %	5	2	0.0000	28.3142
	TSM -2 %	7	2	0.0000	27.6439
	TSM -3 %	5	2	0.0000	26.3861
	TSM -4 %	7	3	0.0000	21.1801
Pop	TSM +1 %	7	2	0.0000	31.0846
	TSM +2 %	7	4	0.0000	25.3826
	TSM +3 %	7	2	0.0333	31.1092
	TSM +4 %	7	2	0.0000	31.0554
	TSM -1 %	7	3	0.0000	35.2740
	TSM -2 %	7	2	0.0000	30.3543
	TSM -3 %	7	2	0.0167	26.6057
	TSM -4 %	7	3	0.0000	26.1156
Rock	TSM +1 %	—	—	Falló	—
	TSM +2 %	7	2	0.0833	26.0024
	TSM +3 %	7	4	0.0333	28.2952
	TSM +4 %	7	2	0.0500	27.5929
	TSM -1 %	7	2	0.0000	22.1259
	TSM -2 %	5	2	0.0000	21.0834
	TSM -3 %	7	2	0.0167	21.0759
	TSM -4 %	7	2	0.0000	21.6778
Vocal	TSM +1 %	7	2	0.0500	21.6414
	TSM +2 %	5	3	0.0000	21.4020
	TSM +3 %	7	2	0.0167	20.0408
	TSM +4 %	7	2	0.0000	20.7835
	TSM -1 %	—	—	Falló	—
	TSM -2 %	7	2	0.0000	28.2005
	TSM -3 %	7	2	0.1833	24.7773
	TSM -4 %	7	3	0.0000	28.1380

Resultados generales del esquema. En las figuras 5.12 a 5.14 se presentan los resultados generales de BER y PSNR que tiene el esquema propuesto ante los ataques de *Jittering* y TSM. Estos resultados generales se obtuvieron promediando los resultados obtenidos con los audios de distintos géneros, además de las desviaciones estándar de estos mismos resultados.

En la figura 5.12a se puede apreciar que para los ataques de *jittering* la detección es perfecta. La figura 5.12b muestra que el comportamiento de PSNR es estable, independientemente de los parámetros del ataque y se tienen desviaciones estándar bajas; es decir, que los resultados de PSNR para los distintos audios no difieren mucho de la media.



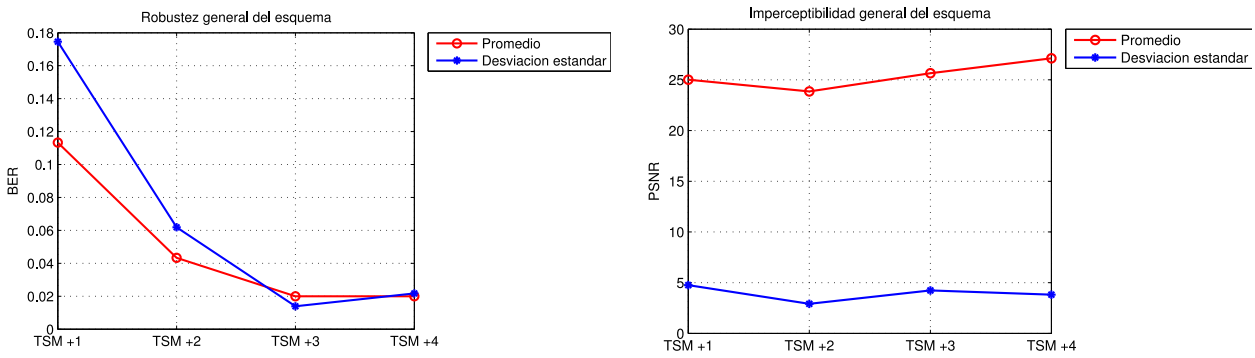
(a) BER.

(b) PSNR.

Figura 5.12: Desempeño general del esquema para los ataques de Jittering.

En la figura 5.13a se puede ver que los promedios de BER para los ataques de incremento en la velocidad de audio (TSM +) son valores bajos, por lo que, en términos generales se tiene una robustez aceptable ante estos ataques [6]. En la figura 5.13b se puede ver que el comportamiento del PSNR es relativamente estable, además que los valores de desviación estándar son bajos.

[6] Los primeros puntos de esta gráfica son más elevados porque para el cálculo del promedio y la desviación estándar se consideró el resultado del audio de Rock, con un BER mayor al 4 %.



(a) BER. **(b)** PSNR.

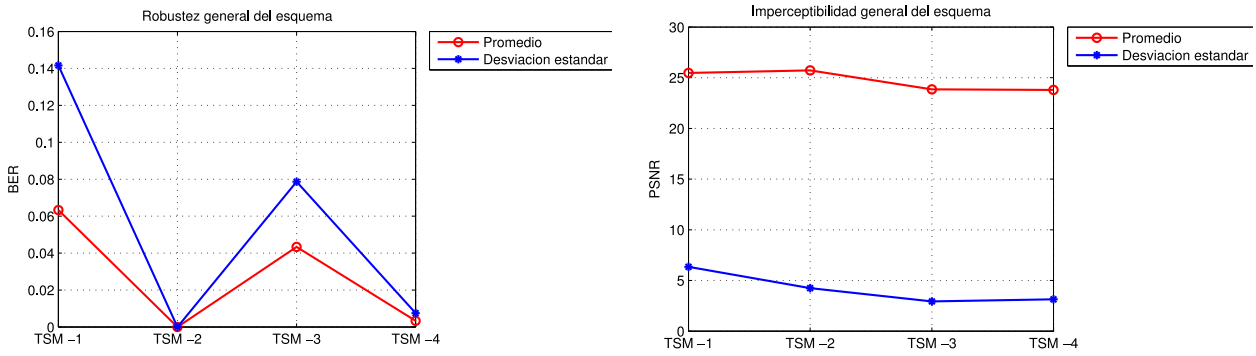
Figura 5.13: Desempeño general del esquema para los ataques de TSM(+).

En la figura 5.14a se puede apreciar que los promedios de BER para los ataques de decremento en la velocidad de audio (TSM-) son valores bajos y en términos generales, se tiene una robustez aceptable contra este tipo de ataques [7]. En la figura 5.14b se puede ver un comportamiento relativamente estable para los promedios de PSNR y que las desviaciones estándar son bajas.

Con estos resultados se puede observar que el comportamiento del esquema propuesto es aceptable con respecto a robustez ante ataques de desincronización y que el comportamiento del PSNR es relativamente establa ante estos mismos ataques. Los resultados tanto de BER como de PSNR son independientes a los audios utilizados para probar el esquema.

A continuación se realizará una comparativa entre los resultados obtenidos con el esquema propuesto y los resultados de los esquemas actuales de la literatura.

[7] Los primeros puntos de esta gráfica son más elevados porque para el cálculo del promedio y la desviación estándar se consideró el resultado del audio de Vocal, con un BER mayor al 3 %.



(a) BER. **(b)** PSNR.

Figura 5.14: Desempeño general del esquema para los ataques de TSM(-).

5.5. Comparativa con otros esquemas

En esta sección se comparan los resultados obtenidos con el esquema propuesto y los resultados reportados en distintos trabajos de la literatura actual. Es importante mencionar que no es posible afirmar si alguno de los trabajos tiene mejor desempeño que el resto, debido a que no se cuentan con las implementaciones de otros autores y con la implementación realizada del trabajo de Yang, *et al.* [28], no fue posible reproducir los resultados reportados. Sin embargo, en esta sección se pretende mostrar que el esquema propuesto da resultados muy parecidos a los resultados reportados por otros trabajos de la literatura actual, con respecto a robustez e imperceptibilidad.

En la tabla 5.8 se presentan los resultados reportados por diferentes autores y los obtenidos con el esquema propuesto; de este último, se consideraron los resultados para el audio de Jazz porque con éste se obtuvieron los mejores, de los otros trabajos también se tomaron los mejores resultados reportados.

Tabla 5.8: Comparativa entre esquemas de marcas de agua con sincronización.

Ataque	Audio Jazz		Resultados reportados							
	Propuesto		Yang[28]		Xiang[27]		Wang[25]		Yang[29]	
	BER	PSNR (dB)	BER	PSNR (dB)	BER	PSNR (dB)	BER	PSNR (dB)	BER	PSNR (dB)
Recortes 10 %	0.0000	34.1399	—	—	—	—	0.0000	34.6220	0.0000	—
Jittering (1/300)	0.0000	37.3076	0.0000	—	—	—	—	—	—	—
Jittering (1/500)	0.0000	37.3330	0.0000	—	0.0000	—	—	—	—	—
Jittering (1/1000)	0.0000	37.4655	0.0000	—	0.0000	—	—	—	—	—
Jittering (1/2000)	0.0000	37.1451	0.0000	—	0.0000	—	—	—	—	—
Jittering (1/3000)	0.0000	37.2703	0.0000	—	—	—	—	—	—	—
TSM +1 %	0.0000	24.6218	0.0000	—	—	—	0.0000	26.7450	0.1221	—
TSM +2 %	0.0000	26.4189	0.0000	—	—	—	—	—	—	—
TSM +3 %	0.0000	24.0972	0.0000	—	—	—	—	—	—	—
TSM +4 %	0.0167	28.5442	0.0000	—	—	—	—	—	—	—
TSM -1 %	0.0000	21.8351	0.0000	—	—	—	0.0000	27.2720	0.0000	—
TSM -2 %	0.0000	21.3610	0.0000	—	—	—	0.0240	27.1430	0.0342	—
TSM -3 %	0.0000	20.4077	0.0000	—	—	—	0.1300	26.9820	0.1670	—
TSM -4 %	0.0167	21.8576	0.0000	—	—	—	0.2070	26.8420	0.2461	—

En la tabla se observa que los mejores resultados con respecto a BER son los del trabajo de Yang, *et al.* [28]; sin embargo, no proporcionan mediciones del impacto perceptual que su esquema y los ataques tienen en los audios. Un escenario probable es que alcancen mayor robustez al sacrificar fidelidad en los audios. La robustez del esquema de Wang, *et al.* [25] es ligeramente menor que la del esquema propuesto; no obstante, su impacto perceptual en los audios es menor [8]. Con esto se pude decir que el esquema propuesto es competitivo con otros esquemas actuales, considerando la robustez ante ataques de desincronización y que el impacto perceptual del esquema es similar al de otros trabajos.

[8] En términos cualitativos, se puede decir que un audio con PSNR entre 20 y 30 dB tiene calidad razonable, entre 30 y 40 dB buena calidad y de más de 40 dB excelente calidad.

5.6. Discusión de los resultados

En esta sección se analizan los resultados presentados anteriormente, las situaciones en que se originan y la importancia de los mismos. Se comenzará estudiando los resultados que tiene el esquema propuesto sin ataques, posteriormente los resultados para los ataques comunes de procesamiento de señales, enseguida los resultados para los ataques de desincronización y finalmente, el desempeño del esquema en comparación con otros trabajos.

Los resultados sin ataques, además de ayudar a encontrar configuraciones adecuadas para los parámetros T y K que se utilizaron en las pruebas siguientes, permitieron verificar que el comportamiento del esquema propuesto es el adecuado.

Para los ataques comunes de procesamiento de señales se puede observar que el impacto perceptual en los audios es el esperado; no obstante, el esquema no es robusto ante este tipo de procesamientos. Como se explicaba anteriormente, esto ocurre por la estrategia de detección propuesta, donde se asume que los ataques afectarán la ubicación espacial de las muestras pero no su valor. Sin embargo, los ataques comunes de procesamiento de señales modifican las amplitudes de las muestras, repercutiendo en el histograma reconstruido y afectando severamente la detección de la marca.

Con los ataques de desincronización, por su parte, el esquema propuesto logra robustez, manteniendo un impacto perceptual aceptable. Se alcanza una recuperación perfecta para los ataques de recortes y jittering, y para los ataques de

modificación de la escala de tiempo se tienen muy pocos errores de detección.

Por último, aunque no es posible determinar si este trabajo de investigación mejora la robustez que tienen otros esquemas ante ataques de desincronización, se puede decir que el esquema propuesto es competitivo con otros trabajos, considerando robustez ante ataques de desincronización e impacto perceptual.

Resumiendo, el esquema propuesto logra gran robustez ante ataques de desincronización, obteniendo una recuperación perfecta en 12 de los 14 ataques de desincronización analizados y en los ataques restantes, solamente falla en detectar un bit de la marca; además, el impacto perceptual del esquema es mínimo en los audios marcados. No obstante, el esquema no es muy robusto ante los ataques comunes de procesamiento de señales evaluados y solamente se logran resultados de robustez aceptables para los ataques menos severos.

En el siguiente capítulo se darán las conclusiones generales de este trabajo de investigación, así como el trabajo futuro que podrá desarrollarse a partir del mismo.

Capítulo 6

Conclusiones y trabajo futuro

6.1. Conclusiones

El problema de sincronización en marcas de agua para audio no es trivial. Si bien es cierto que existen distintos enfoques que tratan de dar soluciones, aún no existe ninguna solución definitiva.

La presente investigación proporciona una solución más a esta problemática, que presenta ventajas y desventajas. Esto ocurre porque existen compromisos al momento de construir esquemas de marcas de agua y no es posible obtener resultados aceptables en todos los aspectos de un sistema de este tipo.

Este trabajo utiliza las bajas frecuencias del dominio DWT para construir un histograma, del cual se modifican las relaciones entre bins adyacentes para representar la inserción de un bit de la marca. La detección se realiza construyendo un histograma con los mismos rangos empleados en la inserción e interpretando estas relaciones para calcular la marca. Dicha estrategia permite alcanzar gran robus-

tez ante ataques de desincronización; sin embargo, su robustez es menor que la de otros trabajos con respecto a los ataques comunes de procesamiento de señales aquí estudiados.

A continuación se verá el cumplimiento de los objetivos de este trabajo de investigación. Recordando el objetivo general:

"Diseñar e implementar un esquema de marcas de agua en audio, robusto ante ataques de desincronización, que mejore el desempeño de los métodos propuestos en la literatura actual"

La primera parte del objetivo general se cumple, porque se diseñó e implementó un esquema de marcas de agua para audio y en el capítulo 5 se muestra que presenta una gran robustez ante los ataques de desincronización. Sin embargo, no se tienen los elementos suficientes para afirmar que se mejora el desempeño de otros métodos en la literatura actual. Esto ocurre porque no se cuenta con toda la información para reproducir los resultados de los otros trabajos y realizar una comparación con condiciones iguales entre los esquemas. No obstante, sí es posible decir que el esquema propuesto es competitivo con otros esquemas actuales, en términos de robustez ante ataques de desincronización e impacto perceptual en los audios.

El esquema de marcas de agua desarrollado en esta investigación es competitivo con otros esquemas de marcas de agua para audio, porque los resultados de robustez ante ataques de desincronización son muy parecidos; es decir, los valores de BER obtenidos por el esquema propuesto son casi iguales a los valores de BER

reportados por otros trabajos. Lo mismo ocurre con el impacto perceptual de los esquemas; los resultados de PSNR para el esquema propuesto son similares a los resultados de PSNR reportados por otros trabajos. Para realizar una comparación directa entre esquemas, sería necesario contar con sus implementaciones, marcar los mismos audios, aplicar los mismos ataques y evaluar los resultados para cada esquema.

Del esquema desarrollado en este trabajo, queda como experiencia que un histograma construido a partir de coeficientes de baja frecuencia en el dominio DWT, es robusto contra ataques que modifican la ubicación temporal de las muestras en el audio (ataques de desincronización); sin embargo, la estrategia de histogramas en dominio DWT no es la más apropiada contra ataques que cambien los coeficientes de las muestras (ataques comunes de procesamiento de señales), porque el histograma es frágil ante los desfases entre las muestras y los bins.

6.2. Trabajo futuro

Es posible modificar el esquema desarrollado en esta investigación, de tal manera que se mejore el desempeño o se adapte a las necesidades de ciertas aplicaciones. A continuación se mencionan estas posibles modificaciones:

1. Desarrollar una estrategia automática para adaptar el rango de inserción según las características del audio a marcar. Es posible realizar un análisis de las características estadísticas que presentan los coeficientes DWT, con lo que se determine un rango óptimo de inserción y detección. Estos rangos óptimos permitirían aumentar la capacidad de inserción del esquema, conservando la robustez ante ataques de desincronización.

2. Modificar el esquema para que se utilice un detector ciego; es decir, que no requiera los rangos de inserción para reconstruir el histograma y recuperar la marca. Con la incorporación de un detector ciego, el esquema propuesto podría utilizarse en aplicaciones comerciales para protección de derechos de autor o control de distribución de copias.

3. Incorporar técnicas de corrección de errores al esquema. Estas estrategias permiten mejorar la recuperación de la marca; sin embargo, limitan la capacidad de inserción porque se incluye información redundante. Esta estrategia ayudaría al esquema propuesto, si éste se utilizara en aplicaciones con canales de transmisión muy ruidosos.

Bibliografía

[1] Arnold, M., Schmucker, M., and Wolthusen, S. *Techniques and Applications of Digital Watermarking and Content Protection*. Computer Security Series. Artech House, 2003.

[2] Barni, M., and Bartolini, F. *Watermarking Systems Engineering*. Signal Processing and Communications Series. Marcel Dekker, 2004.

[3] Cox, I., Miller, M., Bloom, J., Fridrich, J., and Kalker, T. *Digital Watermarking and Steganography*, 2nd edition ed. The Morgan Kaufmann Series in Multimedia Information and Systems. Morgan Kaufmann Publishers, 2008.

[4] Cvejic, N., and Seppänen, T. *Digital Audio Watermarking Techniques and Technologies*. Information Science Reference, 2008.

[5] Erçelebi, E., and Batakçi, L. Audio watermarking scheme based on embedding strategy in low frequency components with a binary image. *Digit. Signal Process. 19* (March 2009), 265–277.

[6] Fan, M.-Q., and Wang, H.-X. Statistical characteristic-based robust audio watermarking for resolving playback speed modification. *Digital Signal Processing 21*, 1 (2011), 110–117.

[7] Hartung, F., and Ramme, F. Digital rights management and watermarking of multimedia content for m-commerce applications. *Communications Magazine, IEEE 38*, 11 (2000), 78–84.

[8] He, X. *Signal Processing, Perceptural Coding and Watermarking of Digital Audio: Advanced Technologies and Models*. IGI Global, 2012.

[9] Hernández Ávalos, P. A. *Esquema de Marcas de Agua Resistente ante Pérdida de Sincronización Temporal para Video*. PhD thesis, Instituto Nacional de Astrofísica, Óptica y Electrónica, 2012.

[10] International Intellectual Property Alliance. Estimated Levels of Copyright Piracy 2008-2009. `http://bit.ly/ekjKn3`, 2010.

[11] Kim, H. J., Choi, Y. H., Seok, J., and Hong, J. *Audio watermarking techniques*. Series on innovative intelligence. World Scientific Press, 2004, ch. 8, pp. 185–217.

[12] Kirovski, D., and Malvar, H. Spread-Spectrum Watermarking of Audio Signals. *IEEE Transactions on Signal Processing 51*, 4 (Apr. 2003), 102–133.

[13] Kutter, M., Bhattacharjee, S., and Ebrahimi, T. Towards second generation watermarking schemes. In *Image Processing, 1999. ICIP 99. Proceedings. 1999 International Conference on* (1999), vol. 1, pp. 320–323 vol.1.

[14] Lu, C.-S. *Multimedia Security: Steganography and Digital Watermarking Techniques for Protection of Intellectual Property*. IGI Publishing, Hershey, PA, USA, 2004.

[15] Mallat, S. G. A Theory for Multiresolution Signal Decomposition: The Wavelet Representation. *IEEE Trans. Pattern Anal. Mach. Intell. 11*, 7 (July 1989), 674–693.

[16] Martinez-Noriega, R., Nakano, M., and Yamaguchi, K. Self-synchronous time-domain audio watermarking based on coded-watermarks. In *Proceedings of the 2010 Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing* (Washington, DC, USA, 2010), IIH-MSP '10, IEEE Computer Society, pp. 135–138.

[17] Nason, G. P., and Silverman, B. W. The Stationary Wavelet Transform and some Statistical Applications. Springer-Verlag, pp. 281–300.

[18] Niu, P.-P., Wang, X.-Y., and Lu, M.-y. A New Digital Audio Watermarking Scheme Robust to Desynchronization Attacks. *2010 Fifth International Conference on Frontier of Computer Science and Technology (FCST)* (Aug. 2010), 233–238.

[19] Perez-Meana, H. *Advances in Audio and Speech Signal Processing*. Idea Group Publishing, 2007.

[20] Pinel, J., Girin, L., Baras, C., and Parvaix, M. A high-capacity watermarking technique for audio signals based on MDCT-domain quantization. In *Proceedings of the 20th International Congress on Acoustics (ICA 2010)* (Sydney, Australie, Aug. 2010), p. ICA2010.

[21] Podilchuk, C., and Delp, E. Digital watermarking: algorithms and applications. *Signal Processing Magazine, IEEE 18*, 4 (jul 2001), 33–46.

[22] Seitz, J. *Digital Watermarking for Digital Media.* Information Science Publishing, 2005.

[23] Steinebach, M., Petitcolas, F., Raynal, F., Dittmann, J., Fontaine, C., Seibel, S., Fates, N., and Ferri, L. StirMark benchmark: audio watermarking attacks. In *Information Technology: Coding and Computing, 2001. Proceedings. International Conference on* (apr 2001), pp. 49–54.

[24] Tao, Z., Zhao, H.-m., Wu, J., Gu, J.-h., Xu, Y.-s., and Wu, D. A Lifting Wavelet Domain Audio Watermarking Algorithm Based on the Statistical Characteristics of Sub-Band Coefficients. *ARCHIVES OF ACOUSTICS 35*, 4 (2010), 481–491.

[25] Wang, F.-H., Pan, J.-S., and Jain, L. C. *Innovations in Digital Watermarking Techniques*, 1st ed., vol. 232 of *Studies in Computational Intelligence*. Springer Publishing Company, Incorporated, 2009.

[26] Xiang, S. Robust Audio Watermarking by Using Low-Frequency Histogram. In *Digital Watermarking*, H.-J. Kim, Y. Shi, and M. Barni, Eds., vol. 6526 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg, 2011, pp. 134–147. 10.1007/978-3-642-18405-5_11.

[27] Xiang, S., Huang, J., and Yang, R. *Time-scale invariant audio watermarking based on the statistical features in time domain*, vol. 4437 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg, 2007, pp. 93–108.

[28] Yang, H. Y., Bao, D. W., Wang, X. Y., and Niu, P. P. A robust content based audio watermarking using UDWT and invariant histogram. *Multimedia Tools and Applications* (Nov. 2010), 1–24.

[29] Yang, H.-y., Wang, X.-y., and Ma, T.-x. A robust digital audio watermarking using higher-order statistics. *AEU: International Journal of Electronics & Communications 65*, 6 (2011), 560–568.

[30] Zhang, X., Yin, X., and Yu, Z. Robust Audio Watermarking Algorithm Based on Histogram Specification. *International Conference on IIHMSP '08* (2008), 163–166.