Índice

1. Audio Fingerprinting: Generalidades

1.1 Introducción

Los sistemas de huellas dactilares tienen más de 100 años de antigüedad. En 1893, Sir Francis Galton fue el primero en probar que no hay dos huellas iguales de dos seres humanos distintos. Aproximadamente 19 años después Scotland Yard aceptó un sistema diseñado por Sir Edward Henry para identificar huellas de gente. El sistema se basa en el modelo de surcos dérmicos en las yemas de los dedos y todavía es la base de todas las técnicas de huellas dactilares humanas de hoy en día. Este tipo de sistema de huellas forense ha existido sin embargo durante más de un siglo, ya que hace 2000 años, los emperadores chinos ya usaban las firmas con el pulgar para documentos importantes. La implicación es que ya esos emperadores (o, al menos, sus sirvientes administrativos) se dieron cuenta que cada huella era única. Conceptualmente, una huella puede ser como un resumen o firma "humana" que es única para cada ser humano. Es importante hacer notar que una huella dactilar humana difiere de un resumen textual en que no permite la reconstrucción de otros aspectos del original. Por ejemplo, una huella humana no da ninguna información sobre el color de los ojos o del pelo de la persona.

Los últimos años han visto un creciente interés científico e industrial en computar huellas de objetos multimedia. El creciente interés industrial es mostrado entre otros por un gran número de compañías (Auditude, Relatable, Audible Magic, Shazam…) y la reciente demanda de información basada en tecnologías de "audio fingerprinting" por parte de la Federación Internacional de la Industria Fonográfica (IFPI) y de la Asociación de la Industria de Grabación de América (RIAA).

1.2 Fundamentos

La tecnología de "Audio fingerprinting" (o "Huella Dactilar de Audio") es una firma compacta basada en el contenido que resume una grabación de audio [1]. "Audio fingerprinting" y, en general, todas las tecnologías CBID ("Content-Based Identification" o "Identificación basada en el contenido") extraen características acústicas relevantes de la señal acústica. Dichas características son únicas para cada señal de audio y, por analogía con las huellas dactilares humanas, son llamadas también huellas dactilares. Una grabación de música o un anuncio solo pueden ser reconocidos si esas características fueron previamente grabadas e introducidas en una base de datos especial.

.

Después de la adquisición de estas características ya no es necesario ningún procesamiento más de la señal. Cuando se implementa la tecnología de "audio fingerprinting", la señal de audio en sí no es modificada, en particular no se le añade ninguna información adicional. El reconocimiento del título se realiza basándose exclusivamente en el contenido, es decir, basándose sólo en características derivadas de la pista de audio.

Este enfoque difiere de otra solución alternativa existente para monitorizar contenido de audio, llamada "Audio Watermarking" (literalmente Marca de Agua de Audio). En ella, se realiza una investigación psicoacústica para que un mensaje arbitrario, la marca de agua, pueda ser incrustado en la grabación sin alterar la percepción del sonido.

Usando "fingerprints" y un algoritmo eficiente para buscar coincidencias en la base de datos, pueden identificarse como el mismo título versiones modificadas o distorsionadas de la misma canción. Dichas modificaciones incluyen, por ejemplo, distorsiones lineales tales como cambios de nivel o limitación de ancho de banda, como pueden darse en el caso de emisiones de radio. Otras modificaciones incluyen distorsiones no lineales, como, por ejemplo, codificación en formato MP3. También pueden reconocerse trozos de material de audio que están incompletos o, incluso, entre varias versiones de una grabación particular, si está grabada en estudio o en directo, y entre varias grabaciones en directo de la misma pieza.

El factor decisivo para la implementación de un proceso de "Audio Fingerprinting" es la selección de las características a investigar. Ser capaces de discernir entre un número elevado de títulos solo es posible si se seleccionan las características adecuadas. Eso sí, al mismo tiempo hay que tener en cuenta que la selección de las características influye directamente en el tamaño de la huella, y, por tanto en el tiempo necesario para identificar un título.

1.3 Propiedades

Los requisitos dependen fuertemente de la aplicación pero son útiles para evaluar y comparar diferentes tecnologías de "audio fingerprinting".

Una enumeración detallada de los requisitos que pueden ayudarnos a distinguir entre los distintos enfoques incluye:

- *Precisión*: El número de identificaciones correctas, identificaciones falsas (falsos positivos) e identificaciones que se pasan por alto.

- *Fiabilidad*: En la generación de "playlists" o listas de reproducción para organizaciones de control de copyright es de una importancia capital tener métodos para valorar si un elemento está presente o no en el conjunto de elementos a identificar. En esos casos, si una canción no ha sido emitida, no debería ser identificada como una coincidencia, incluso a costa de obviar verdaderas coincidencias. En otras aplicaciones, como el etiquetado automático de archivos .MP3, evitar falsos positivos no es una necesidad tan perentoria.

- *Robustez*: Habilidad para identificar con precisión un elemento, sin importar su nivel de compresión y distorsión o de interferencia en el canal de transmisión. Es también la habilidad de identificar títulos completos a partir de extractos de unos pocos segundos, lo cual requiere métodos para tratar con la falta de sincronización. Otras

fuentes de degradación son la ecualización, ruido de fondo, conversiones A/D y D/A, codificadores de audio (tales como GSM y MP3), etc.

- *Seguridad*: Vulnerabilidad de una solución a manipulaciones intencionadas. En contraste con el requerimiento de robustez, las manipulaciones con las que hay que tratar están diseñadas específicamente para engañar al algoritmo de identificación de "fingerprint".

- *Versatilidad*: Habilidad para identificar sea cual sea su formato. Habilidad para usar la misma base de datos para distintas aplicaciones.

- *Escalabilidad*: Actuación con bases de datos de títulos muy largas o con un gran número de identificaciones concurrentes. Esto afecta a la precisión y a la complejidad del sistema.

- *Complejidad*: Se refiere al coste computacional de la extracción de la huella, al tamaño de la misma, a la complejidad de la comparación, al coste de añadir nuevos elementos a la base de datos, etc.

- *Fragilidad*: Algunas aplicaciones, tales como sistemas de verificación de la integridad del contenido, pueden requerir la detección de cambios en el contenido. Esto es contrario al requerimiento de robustez, ya que la huella debería ser robusta a transformaciones en las que se preserva el contenido, pero no a otras distorsiones.

Mejorar un determinado requerimiento implica empeorar cualquier otro. Generalmente, la huella debería ser:

- *Un resumen perceptual de la grabación*. La huella debe retener el máximo de información acústicamente relevante. Este resumen debería permitir la discriminación entre un número elevado de huellas. Esto puede entrar en conflicto con otros requerimientos tales como la complejidad o la robustez.

- *Invariante a las distorsiones*. Esto deriva del requisito de robustez. Sin embargo, las aplicaciones que vigilan la integridad del contenido relajan esta restricción para las distorsiones que preservan el contenido, con la intención de detectar manipulaciones deliberadas.

- *Compacto*. Para la complejidad es interesante una representación de pequeño tamaño, ya que necesitamos almacenar y comparar un gran número de huellas (puede que millones, depende de la aplicación). Sin embargo, un tamaño de representación excesivamente corto puede no ser suficiente para discriminar entre grabaciones, afectando a la precisión, fiabilidad y robustez.

- *Fácilmente computable*. Por razones de complejidad, la extracción de la huella no debería consumir demasiado tiempo.

1.4 Modos de uso

1.4.1 Identificación

Independientemente del enfoque específico para extraer la firma compacta basada en el contenido, se puede observar una arquitectura común para describir la funcionalidad del "fingerprinting" cuando se usa para identificación.

El funcionamiento general mimetiza la manera en la que los humanos realizan la tarea. Como se ve en la fig. 1, se realiza en dos niveles, creación de la base de datos e identificación en si misma:

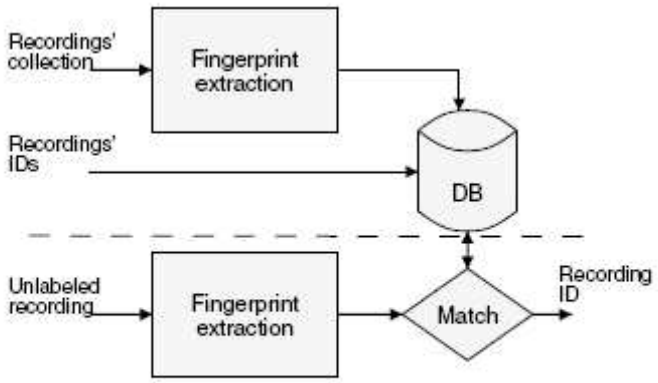

Fig. 1. Content-based audio identification framework

-Creación de la base de datos: La colección de trabajos que se pretende reconocer es presentada al sistema para la extracción de su huella. Las huellas son almacenadas en la base de datos y pueden ser enlazadas a una etiqueta o a otros metadatos relevantes para cada grabación.

-Identificación: El audio sin etiquetar es procesado para extraer la huella. Ésta es comparada con las huellas de la base de datos. Si se encuentra una coincidencia, se obtiene la etiqueta asociada al trabajo que estaba en la base de datos. También se puede realizar una medida de la fiabilidad de dicha coincidencia. La diferencia entre unos y otros enfoques del problema se encuentra en las características de audio que se observan, además de en los algoritmos de búsqueda e indexado que se usan. Entre las características a estudiar podemos nombrar: energía, volumen, centroide espectral, tasa de cruces por cero, tono, armonicidad, monotonía espectral, coeficientes Mel-Cepstrum, etc. Los algoritmos pueden ir desde la comparación directa, extrayendo un "hash" del archivo binario y usando alguna representación compacta como el CRC (código de redundancia cíclico) para almacenarlo en la base hasta otros algoritmos más complejos y robustos (a veces basados también en "hash") que estudiaremos más detalladamente más adelante.

1.4.2 Verificación de la integridad

La verificación de la integridad pretende detector la alteración de los datos. El funcionamiento general es similar al de identificación (ver fig.2). Primero, se extrae la huella del audio original. En la fase de verificación, la huella extraída de la señal de prueba es comparada con la huella de la original. Como resultado, se da como salida un informe indicando si la señal ha sido manipulada o no. Opcionalmente el sistema puede indicar el tipo de manipulación y donde ha ocurrido dentro del audio. Los datos de verificación, que deberían ser significativamente más pequeños que los datos de audio, pueden ser enviados junto a los anteriores (por ejemplo, en una cabecera) o almacenados en una base de datos. Una técnica, conocida como "auto-incrustación" evita la necesidad de una base de datos o una cabecera especialmente dedicada incrustando la firma basada en el contenido en los datos de audio usando "watermarking" [2].

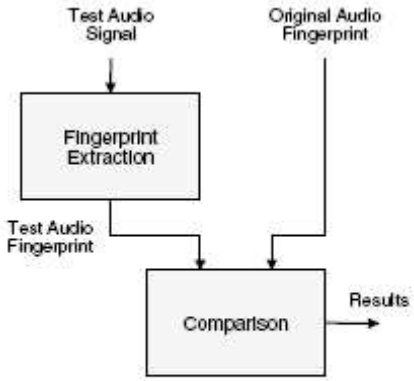

Fig. 2. Integrity verification framework

1.4.3 Apoyo al "Watermarking"

La técnica de "Audio Fingerprinting" puede asistir a la de "watermarking". La huella de audio puede ser usada para obtener claves secretas a partir del contenido presente. Como se describe en [3], usar la misma clave secreta para una serie de elementos de audio distintos puede comprometer la seguridad, ya que cada elemento puede filtrar información parcial sobre la clave. La codificación "hash" perceptual puede ayudar a generar claves dependientes de la entrada para cada pieza de audio. Un ámbito donde puede ser usado el "audio fingerprinting" para mejorar la seguridad de las marcas de agua es en el intento de hacer copias ilegales [4]. En estos casos se intenta estimar una marca de un contenido "marcado" y transplantarla a un contenido "no marcado". Además puede usarse "fingerprinting" para combatir intentos de inserción o borrado de contenido que causan una desincronización de la detección: usando la huella, el detector es capaz de encontrar "puntos ancla" en el flujo de audio y resincronizarse en ellos [3].

1.4.4 Recuperación y procesamiento de audio basados en el contenido

Obtener firmas compactas a partir de objetos multimedia complejos e índices potentes para buscar contenidos es un asunto esencial en Recuperación de Información Multimedia. La técnica de "fingerprinting" puede extraer información de la señal de audio en diferentes niveles de abstracción, desde descriptores de bajo nivel hasta descriptores de alto nivel. Especialmente las abstracciones de alto nivel para modelar audio nos dan la posibilidad de extender los modos de uso de dicha técnica a la navegación basada en el contenido, la búsqueda por similaridad, procesamiento basado en el contenido y otras aplicaciones de Recuperación de Información de Música. Adaptar los eficientes sistemas actuales de la identificación a la búsqueda de similaridad puede tener un significativo impacto en la industria musical. Antes, los proveedores de música on-line ofrecen una búsqueda por datos editoriales (artista, título, etc.). Con "fingerprinting" podría usarse la huella de una canción para encontrar no solo la versión original de dicha canción, sino otras versiones similares de la misma.

1.5 Escenarios de aplicación

Aquí vamos a presentar algunos usos particulares de la tecnología de "fingerprinting", la mayoría de ellos son casos particulares del modo de uso de identificación ya descrito. Se basan en la habilidad del "audio fingerprinting" de unir audio sin etiquetar a sus correspondientes meta-datos, sea cual sea el formato del audio.

1.5.1 Monitorización y Seguimiento del contenido de audio

Un sistema de monitorización a gran escala basado en huellas consiste en varios puntos de monitorización y en un lugar central de monitorización donde se encuentra el servidor de huellas. En los puntos de monitorización, se extraen las huellas de todos los canales de emisión (locales). El sitio central recoge las huellas de los puntos de monitorización. Posteriormente, el servidor, que contiene una enorme base de datos de huellas, produce las listas de reproducción de todos los canales de emisión.

<u>En el lado del distribuidor</u>

Los distribuidores de contenido pueden necesitar saber si tienen los derechos o no para difundir el contenido a los consumidores. La huella puede ayudar a identificar audio sin etiquetar en los archivos de canales de TV y radio. También puede identificar contenido sacado de CDs de fábrica y distribuidores en investigaciones anti-piratería (por ejemplo, avances de grabaciones sin masterizar en plantas de fabricación de CDs).

En el canal de transmisión

En muchos países, las emisoras de radio deben pagar derechos de autor por la música que emiten. Los poseedores de los derechos necesitan monitorizar las transmisiones para verificar si dichos derechos están siendo pagados apropiadamente o no. Incluso en los países donde las emisoras pueden emitir música gratuitamente, los poseedores están interesados en monitorizarlas con objetivos estadísticos. Los anunciadores también necesitan monitorizar las transmisiones de radio y TV para verificar si sus anuncios están siendo difundidos según lo acordado. Lo mismo se puede aplicar a emisoras web. Otros usos incluyen compilaciones para análisis estadístico o refuerzo de "leyes culturales" (por ejemplo, las canciones francesas en Francia). Para todos estos propósitos actualmente se están usando sistemas de monitorización basados en el "fingerprinting". El sistema "escucha" la radio y continuamente actualiza una lista de reproducción de canciones o anuncios emitidos por cada emisora. Por supuesto, debe estar disponible para el sistema una base de datos que contenga las huellas de todas las canciones y anuncios que se pretenden identificar, y dicha base de datos debe ser actualizada con las nuevas canciones que salgan. Ejemplos de proveedores comerciales de este servicio son: Broadcast Data System, Music Reporter, Audible Magic o Yacast.

Napster y otras comunidades parecidas, donde los usuarios intercambian música, han sido excelentes canales para la piratería musical. Después de una batalla judicial con la industria musical, a Napster se le prohibió facilitar la transferencia de música con copyright. La primera medida que se tomó conforme al mandato judicial fue la introducción de un sistema de filtrado basado en el análisis del nombre de los archivos, de acuerdo con listas de grabaciones de música con copyright aportadas por las compañías discográficas. Pero este simple sistema no solucionó el problema, ya que los usuarios demostraros ser extremadamente creativos para escoger títulos de canciones que "torearan" el sistema, pero permitiendo a los otros usuarios una identificación fácil. Además, el elevado número de canciones con títulos idénticos fue un factor adicional que redujo la eficiencia de dichos filtros. Los sistemas de monitorización basada en las huellas constituyen una solución bastante apropiada al problema. De hecho, Napster adoptó una nueva tecnología de "fingerprinting" y un nuevo sistema de filtrado de archivos basados en ella. Además, se puede encontrar contenido de audio en páginas web ordinarias. El "audio fingerprinting" combinado con un "web crawler" ("araña de la web", que inspecciona las páginas del World Wide Web de forma sistemática y automatizada) puede identificar este contenido e informar a los correspondientes poseedores de los derechos.

En el extremo del consumidor

En aplicaciones de monitorización de la política de uso, el objetivo es evitar un mal uso de las señales de audio por parte del consumidor. Podemos concebir un sistema donde una pieza de música es identificada por medio de una huella, y se mira en una base de datos para obtener información sobre sus derechos. La información dicta el comportamiento del dispositivo en el que se va a reproducir (por ejemplo, lectores y grabadores

de CD y DVD, reproductores de MP3 o incluso ordenadores), de acuerdo con la política de uso. Dichos dispositivos necesitan estar conectados a una red para poder tener acceso a la base de datos.

1.5.2 Servicios de valor añadido

La información del contenido se define como información sobre un extracto de audio que es relevante para el usuario o necesario para la aplicación pretendida. Dependiendo de la aplicación y del perfil del usuario, se pueden definir varios niveles de información de contenido. Estas son algunas de las situaciones que podemos imaginar:

- Información de contenido describiendo un extracto de audio, tales como descripciones rítmicas, armónicas y melódicas.
- Metadatos describiendo un trabajo musical, como fue compuesto y como fue grabado. Por ejemplo: compositor, año de composición, intérprete, fecha de la actuación, grabación en estudio/actuación en directo…
- Otras informaciones concernientes a un trabajo musical, tales como la imagen de la portada del álbum, precio del mismo, biografía del artista, información de los próximos conciertos, etc.

Se pueden definir distintos perfiles. Los usuarios normales estarían interesados en informaciones generales sobre un trabajo musical, tales como el título, compositor, sello discográfico y año de edición; los músicos podrían estar interesados en qué instrumentos se han utilizado, los ingenieros de sonido podrían estar interesados en informaciones sobre el proceso de grabación. La información de contenido puede ser estructurada por medio de un esquema de descripción de música (MusicDS), que es una estructura de metadatos usada para describir y anotar datos de audio. El estándar MPEG-7 propone un esquema de descripción para contenido multimedia basado en el metalenguaje XML, aportando un intercambio de datos fácil entre distintos equipos.

Algunos sistemas almacenan información de contenido en una base de datos que es accesible a través de Internet. La huella puede ser pues usada para identificar una grabación y obtener la correspondiente información de contenido, sin tener en cuenta el tipo de soporte, formato de archivo o cualquier otra particularidad de los datos de audio. Por ejemplo, *MusicBrainz, Id3man* o *Moodlogic* etiquetan automáticamente colecciones de archivos de audio; el usuario puede descargar un reproductor compatible que extrae las huellas y las manda a un servidor central desde donde se descargan los metadatos asociados a las grabaciones. *Gracenote,* quien ha estado proveyendo enlaces a metadatos de música basados en la tabla de contenidos (TOC) de un CD, empezó a ofrecer tecnología de "audio fingerprinting" para extender el enlace hasta el nivel de canción. Su método de identificación de audio se usa en combinación con clasificadores basados en texto para aumentar la precisión. Todo esto se puede usar para organizar una "biblioteca" musical dentro de cada ordenador. Hoy en día muchos usuarios de PCs tienen una biblioteca musical que contiene varios cientos, incluso miles de canciones. La música está almacenada

normalmente en formato comprimido (por ejemplo MP3) en sus discos duros. Cuando esas canciones se obtienen de otras fuentes, tales como "ripeadas" de un CD o descargadas de una red de intercambio de archivos, estas librerías habitualmente no están bien organizadas. Los metadatos son inconsistentes, incompletos y, a veces, incluso inexistentes. Asumiendo que la base de datos de huellas contiene metadatos correctos, la tecnología de "audio fingerprinting" puede hacer consistentes los metadatos de las canciones de la librería, permitiendo una fácil organización basada, por ejemplo en disco o artista.

Otro ejemplo es la identificación de una melodía mediante dispositivos móviles, por ejemplo un teléfono móvil; es una de las situaciones más exigentes en términos de robustez, ya que la señal de audio viaja a través de distorsión de radio, conversiones A/D y D/A, ruido de fondo y codificación GSM, distorsión del canal de comunicación móvil y sólo están disponibles unos pocos segundos de audio.

Otro posible ejemplo podrían ser las radios de los coches que ofrecen un botón de identificación.

1.5.3 Sistemas de verificación de Integridad

En algunas aplicaciones, la integridad de las grabaciones de audio debe ser establecida antes de que la señal pueda ser de hecho usada, es decir, uno debe asegurar que la grabación no ha sido modificada o que no está demasiado distorsionada: Si la señal sufre compresión con pérdidas, conversiones A/D o D/A, u otras transformaciones que preservan el contenido en el canal de transmisión, la integridad no puede ser comprobada por medio de funciones de "hash" estándar, ya que el mínimo cambio en un solo bit es suficiente para que la salida de la función cambie. Los métodos basados en la marca de agua pueden ofrecer falsas alarmas en este contexto. Los sistemas basados en "audio fingerprinting", a veces combinados con "watermarking" están siendo investigados para afrontar este problema [2]. Entre algunas posibles aplicaciones podemos nombrar: comprobar que los anuncios son emitidos con la calidad y longitud requeridas, verificar que una grabación supuestamente infractora es de hecho la misma que una cuyo dueño es conocido, etc.

1.6 Alternativas: Audio Watermarking

Durante siglos, el uso de documentos con marca de agua impresa para evitar falsificaciones ha sido una práctica habitual. Por analogía, el término "Watermarking" también es habitualmente usado para describir métodos que pretenden marcar imperceptiblemente documentos digitales (imágenes, audio o video) [5].

Las primeras investigaciones en "audio watermarking" datan de mediados de los años 90 [6]. La idea básica consiste en añadir una señal, la marca de agua, a la señal original de audio. La señal resultante debe ser

percibida por el oyente como idéntica a la original. La marca de agua transporta datos que pueden ser recuperados por un detector y ser usados para una multitud de objetivos.

Al igual que los sistemas de "audio fingerprinting", estos sistemas deben cumplir una serie de propiedades, que frecuentemente dependen de la aplicación y entran en conflicto unas con otras. En general podemos nombrar: inaudibilidad, robustez, capacidad, fiabilidad y baja complejidad. Como ya hemos dicho, dichos requisitos dependen de la aplicación. Así, por ejemplo, algunas aplicaciones (como el audio en Internet, con baja tasa de bit) pueden admitir que la marca de agua añada una pequeña degradación de la calidad de la señal, mientras que otras (como audio con alta tasa de bit) deben ser extremadamente rigurosas con este asunto.

En general, podemos ver este sistema como un sistema de comunicaciones: la marca de agua es la señal de información y la señal de audio hace el papel de ruido del canal. En los sistemas de comunicación convencionales la señal útil es normalmente más fuerte que el ruido, que normalmente se asume que es blanco y gaussiano. Aquí este no es el caso. Para evitar la distorsión audible, la señal de marca de agua debe ser mucho más débil (decenas de decibelios) que la señal de audio. Además, la señal de audio generalmente no es estacionaria y es fuertemente coloreada. En la literatura han sido propuestos varios enfoques para "audio watermarking":

- De Espectro Ensanchado: Como en los sistemas de comunicación de espectro ensanchado, la idea consiste en ensanchar la "marca de agua" para maximizar su potencia, consiguiendo mantenerla inaudible e incrementar su resistencia a ataques [6].

- "Echo-hiding": Se explotan las propiedades de enmascaramiento temporal para mantener la "marca de agua" inaudible. La marca es un "eco" de la señal original [7].

- En la Cadena de Bits: La marca es insertada directamente en el flujo de bits generado por un codificador de audio.

Se han propuesto muchas variantes de estos esquemas básicos. Por ejemplo, en vez de añadir la marca en el dominio del tiempo, hacerlo en el de la frecuencia, reemplazando directamente componentes espectrales [8].

Para asegurar la inaudibilidad de la señal se usan modelos psicoacústicos. Es bien sabido que cuando dos tonos están muy próximos en frecuencia, un tono, si es mucho más fuerte, puede enmascarar al otro. Los modelos se usan para generalizar el efecto de enmascaramiento en frecuencia a señales no tonales. A partir de una señal de audio, se calcula una curva llamada *umbral de enmascaramiento,* homogénea a una densidad espectral de potencia. La marca se construye modificando en frecuencia una señal aproximadamente blanca según el umbral de audición. Después de esta operación, el PSD de la marca siempre estará por debajo del umbral de audición y no se oirá en presencia de la señal de audio. Se puede conseguir esto con una diferencia de potencia entre la señal y la marca de unos 20 dB.

1.6.1 Semejanzas y diferencias con "audio fingerprinting"

- El "Audio Watermarking" modifica la señal de audio original incrustando una marca en ella, mientras que el "Audio Fingerprinting" no la cambia en absoluto, sino que la analiza y obtiene un "hash" (la huella) unívocamente asociada a la señal. En "watermarking" hay un compromiso entre la potencia de la "marca" (y su audibilidad) y su capacidad de detección. En "fingerprinting" no existe tal compromiso: el sistema "escucha" el audio, construye una descripción y busca en la base de datos una descripción que coincida.

- Necesidad de un catálogo de huellas: Un oyente humano sólo puede identificar una pieza de música si la ha oído antes, salvo que tenga acceso a algo más que a la señal de audio. Similarmente, los sistemas de "fingerprinting" requieren un conocimiento previo de las señales de audio para poder identificarlas, ya que no hay disponible ninguna información además de la señal en si misma durante la fase de identificación. Por tanto, debe construirse una base de datos, que contenga todas las canciones que se supone que debe identificar el sistema. Mientras más grande sea la base de datos, más requerimientos de memoria y coste computacional serán necesarios, aumentando, por tanto la complejidad del proceso de detección. En contraste a esto, para un sistema de "watermarking" no hace falta base de datos, ya que toda la información asociada a la señal está en la "marca" en sí misma. El detector comprueba la presencia de una marca y, si encuentra una, extrae los datos contenidos en ella. Además, tampoco hace falta ir actualizando el detector cuando aparecen nuevas canciones y la complejidad no cambia cuando llegan marcas nuevas.

- Para algunas aplicaciones, la necesidad de preprocesar las señales de audio es una gran desventaja de los sistemas de "watermarking". Por ejemplo, los sistemas de monitorización de distribución basados en "watermarking" solo serían capaces de detectar infracciones de copyright si las señales con copyright habían sido previamente marcadas, lo que significa que todo el material antiguo no marcado no sería protegido en absoluto. Además, el nuevo material tendría que ser marcado en todos sus formatos de distribución, e incluso la disponibilidad de un pequeño grupo de copias no marcadas podría comprometer la seguridad del sistema. Esto no es ninguna preocupación para los sistemas de "fingerprinting", ya que no hace falta ningún procesado previo.

- En la detección de la marca, la señal que contiene la información útil para la misma corresponde a una fracción muy pequeña de la potencia de entrada, ya que la marca es mucho más débil que la señal de audio original debido a la restricción de inaudibilidad. Además, el ruido que podría añadirse a la señal marcada (por compresión MP3 o transmisión analógica, por ejemplo), puede ser tan fuerte como la marca, o incluso más. En casos de perturbación severa de canal o de ataques de piratas, puede que la marca deje de ser detectable.

En contraste, la detección en sistemas de "fingerprinting" está directamente basada en la señal de audio en sí misma, la cual es suficientemente fuerte para resistir la mayoría de las perturbaciones de canal y es menos susceptible a ataques piratas. Estos sistemas son, por tanto, inherentemente más robustos. Mientras que el audio original en la base de datos suene "aproximadamente" igual que la pieza de música que el sistema está escuchando, sus huellas también serán similares. La definición de "aproximadamente" depende del proceso de extracción de la huella, y, por tanto, la robustez del sistema depende también de eso. La mayoría de sistemas usan un enfoque psicoacústico para sacar la huella. Haciendo esto, el audio a analizar puede sufrir una fuerte distorsión sin que ello provoque un decremento en la efectividad del sistema.

- La información contenida en la "marca de agua" puede no tener relación directa con la portadora de la señal de audio. Por ejemplo, una emisora de radio podría incrustar las últimas noticias en las canciones que emite a través de una marca; en la recepción las noticias aparecerían en una pequeña pantalla mientras suenan las canciones. Por el contrario, una huella está correlada con la señal de audio a partir de la cual se obtuvo; cualquier cambio en la señal de audio que sea perceptible al oído humano debería provocar un cambio en la huella. Este hecho está detrás de la mayoría de diferencias en las aplicaciones de cada uno de los dos enfoques: mientras que las marcas pueden transportar cualquier tipo de información, las huellas siempre representan la señal de audio.

  Esta independencia entre señal e información se deriva del hecho de que los sistemas de marcas solo tratan con información que ha sido previamente añadida, dado que no se provee la conexión a ninguna base de datos. Esta información pude estar relacionada o no con la señal de audio en la que se ha incrustado. Con "fingerprinting" se puede extraer información de la señal en diferentes niveles de abstracción, dependiendo de la aplicación y el escenario de uso. Las abstracciones de nivel más alto permiten extender las aplicaciones a navegación basada en el contenido, búsqueda por similaridad y otras aplicaciones de Recuperación de Información Musical.

En conclusión, ambas metodologías tienen muchas aplicaciones en común y también muchas específicas de cada una. El "audio watermarking", aunque en un principio estaba pensado para protección de copyright, también es útil para otros muchos propósitos, particularmente para transporte de información de propósito general. El "audio fingerprinting" se usa sobre todo para identificar señales de audio, no solo en aplicaciones de copyright, sino también en reconocimiento de anuncios, por ejemplo.

Se podría decir que el "watermarking" tiene un rango más amplio de aplicaciones que "fingerprinting". Sin embargo, éste último es inherentemente más robusto, lo cual significa que resistirá distorsiones más fuertes, lo que le hace particularmente atractivo en cuestiones de copyright. Y el hecho de pueda reconocer audio a partir de extractos hace que sea una solución flexible para estos temas. Eso sí, la protección absoluta contra la piratería no es más que una mera ilusión. Puesto que sus puntos fuertes

son normalmente complementarios, el uso de ambas técnicas combinadas da lugar a aplicaciones muy interesantes.

.