(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 382 974**

(51) Int. CI.:
G10L 11/00 (2006.01)
G06F 17/30 (2006.01)
G11B 27/00 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Número de solicitud europea: **02747417 .0**
(96) Fecha de presentación: **20.06.2002**
(97) Número de publicación de la solicitud: **1405222**
(97) Fecha de publicación de la solicitud: **07.04.2004**

(54) Título: **Procedimiento y dispositivo para generar una huella digital y procedimiento y dispositivo para identificar una señal de audio**

(30) Prioridad:
10.07.2001 DE 10133333

(45) Fecha de publicación de la mención BOPI:
15.06.2012

(45) Fecha de la publicación del folleto de la patente:
15.06.2012

(73) Titular/es:
M2ANY GMBH
LICHTENBERGSTRASSE 8
85748 GARCHING, DE

(72) Inventor/es:
HERRE, Jürgen;
ALLAMANCHE, Eric;
HELLMUTH, Oliver;
KASTNER, Thorsten y
CREMER, Markus

(74) Agente/Representante:
Arizti Acha, Monica

DESCRIPCIÓN

Procedimiento y dispositivo para generar una huella digital y procedimiento y dispositivo para identificar una señal de audio.

La presente invención se refiere a la caracterización o identificación de señales de audio con respecto a su contenido, en particular a la generación y utilización de diferentes huellas digitales para una señal de audio.

En los últimos años ha aumentado considerablemente la disponibilidad de material de audio multimedia, es decir de datos de audio. Este desarrollo se produjo por una serie de factores técnicos. Estos factores técnicos comprenden por ejemplo la amplia disponibilidad de Internet, la amplia disponibilidad de ordenadores de gran capacidad y la amplia disponibilidad de procedimientos de gran capacidad para la compresión de datos, es decir codificación de fuente, de datos de audio. Como ejemplo para esto se indica MPEG 1/2 Layer 3, que también se denomina MP3.

Las enormes cantidades de datos audiovisuales, que están disponibles por ejemplo en Internet a nivel mundial, requieren conceptos que permitan evaluar, catalogar o gestionar estos datos según criterios de contenido. Existe la necesidad de buscar y encontrar datos multimedia de manera dirigida mediante la indicación de criterios útiles.

Esto requiere la utilización de las denominadas técnicas “basadas en contenido” que de los datos audiovisuales extraen denominadas características, que en la técnica se denominan también “propiedades” (*features*), que representan propiedades de contenido características importantes de la señal de interés. Basándose en tales características o combinaciones de tales características pueden derivarse relaciones de similitud o puntos en común entre las señales de audio. Esta operación se produce en general mediante la comparación o relación de los valores de característica extraídos a partir de diferentes señales, que en este caso también se denominarán “fragmentos”.

La patente estadounidense n.º 5.918.223 da a conocer un procedimiento para el análisis basado en contenido, almacenamiento, recuperación y segmentación de información de audio. Un análisis de datos de audio genera un conjunto de valores numéricos, que también se denomina vector de característica, y que puede utilizarse para clasificar y ordenar según su rango la similitud entre fragmentos de audio individuales, que normalmente se almacenan en una base de datos multimedia o en la *World Wide Web*.

El análisis posibilita además la descripción de clases de fragmentos de audio definidas por el usuario basándose en un análisis de un conjunto de fragmentos de audio, siendo todos componentes de una clase definida por el usuario. El sistema puede encontrar secciones de sonido individuales dentro de una pieza de música más larga, lo que permite que el registro de audio se segmente automáticamente en una serie de segmentos de audio más cortos.

Como características para la caracterización o clasificación de fragmentos de audio con respecto a su contenido se utiliza la intensidad de sonido de un fragmento, la altura de sonido (“*Brightness*”), el ancho de banda y los denominados coeficientes cepstrales en las frecuencias de Mel (MFCC) en intervalos periódicos en el fragmento de audio. Los valores por cada bloque o trama se almacenan y se someten a una primera derivación. A continuación se calculan magnitudes estadísticas específicas, como por ejemplo el valor medio o la desviación estándar, y concretamente de cada una de estas características incluyendo las primeras derivaciones de las mismas, para describir una variación a lo largo del tiempo. Este conjunto de magnitudes estadísticas forma el vector de característica. El vector de característica del fragmento de audio se almacena en una base de datos en asociación con el archivo original, pudiendo acceder un usuario a la base de datos, para llamar a fragmentos de audio correspondientes.

El sistema de base de datos puede cuantificar la distancia en un espacio n dimensional entre dos vectores n dimensionales. Además es posible generar clases de fragmentos de audio, especificando un conjunto de fragmentos de audio, que pertenece a una clase. Clases a modo de ejemplo son gorjeos de pájaros, música rock, etc. El usuario puede examinar la base de datos de fragmentos de audio utilizando procedimientos específicos. El resultado de una búsqueda es una lista de archivos de sonido, que se enumeran de manera ordenada según su distancia con respecto al vector n dimensional especificado. El usuario puede examinar la base de datos con respecto a características de similitud, con respecto a características acústicas o psicoacústicas, con respecto a características subjetivas o con respecto a ruidos especiales, como por ejemplo el zumbido de abejas.

La publicación “Multimedia Content Analysis”, Yao Wang entre otros, IEEE señal Processing Magazine, noviembre de 2000, páginas 12 a 36, da a conocer un concepto similar para caracterizar fragmentos multimedia. Como características para clasificar el contenido de un fragmento multimedia se proponen características de dominio de tiempo o características de dominio de frecuencia. Éstas comprenden el volumen de sonido, la altura de sonido como frecuencia fundamental de una forma de señal de audio, características espectrales, como por ejemplo el contenido de energía de una banda con respecto al contenido de energía total, frecuencias límite en el recorrido espectral, etc. Además de características a corto plazo, que se refieren a las magnitudes mencionadas por cada bloque de valores de muestreo de la señal de audio, también se proponen características a largo plazo, que se refieren a un periodo de tiempo más largo del fragmento de audio.

Para la caracterización de fragmentos de audio se proponen diferentes categorías, como por ejemplo ruidos de animales, ruidos de campanas, ruidos de una muchedumbre, risas, ruidos de maquinaria, instrumentos musicales, voz masculina, voz femenina, ruidos de teléfono o ruidos de agua.

En la elección de las características utilizadas es problemático que el esfuerzo de cálculo para la extracción de una característica debe de ser moderado, para alcanzar una caracterización rápida, que sin embargo simultáneamente debe ser la característica para el fragmento de audio, de tal manera que dos fragmentos diferentes también presenten características distinguibles entre sí.

Para identificar una señal de audio, tal como se ha explicado, se extrae una identificación de la señal de audio o una denominada propiedad, que también se denomina huella digital. Se exigen dos requisitos diferentes con respecto a la forma de la propiedad. Uno de los requisitos con respecto a la huella digital consiste en que la huella digital debe señalizar la señal de audio de la manera más unívoca posible. El otro requisito con respecto a la huella digital consiste en que la huella digital debe contener la menor información posible, es decir, que la huella digital debe utilizar el menor espacio de almacenamiento posible. Estos dos requisitos son contradictorios. De la manera más sencilla puede reconocerse en que la mejor “huella digital” para una señal de audio es la propia señal de audio, es decir la secuencia de valores de muestreo, que representa la señal de audio. Una huella digital de este tipo perjudicaría sin embargo considerablemente el segundo requisito, porque la huella digital de la señal de audio requeriría demasiada memoria, lo que por un lado haría imposible el almacenamiento de muchas huellas digitales para muchas señales de audio en una base de datos de reconocimiento de música. Otra desventaja es que los algoritmos de correspondencia, que deben comparar una huella digital de búsqueda con una pluralidad de huellas digitales de base de datos almacenadas requieren más tiempo de cálculo cuanto mayor es la huella digital de búsqueda o la huella digital de base de datos.

El otro extremo consistiría, por ejemplo, en tomar sólo un valor medio sobre todos los valores de muestreo de un fragmento. Este valor medio exige sólo muy poco espacio de almacenamiento y por tanto es el más adecuado tanto para una base da datos de música grande como para algoritmos de correspondencia. La fuerza de identificación de una huella digital de este tipo sería sin embargo poco robusta con respecto a una para una modificación irrelevante para una persona.

En general no existe un compromiso óptimo entre fuerza de identificación por un lado y volumen de datos de la huella digital por otro, sino que habitualmente se halla de manera empírica o depende de las circunstancias de la aplicación respectiva con respecto al espacio de almacenamiento disponible y la capacidad de transmisión disponible. Este modo de proceder tiene la desventaja de que los diferentes tipos de huellas digitales sólo son adecuados de manera óptima para una aplicación especial, sin embargo para otras aplicaciones son poco adecuados en mayor o menor medida. En este contexto se hace referencia a que una identificación de señal de audio o caracterización sólo es de especial interés cuando existen bases de datos de características muy grandes, cuyas huellas digitales pueden compararse con una huella digital de búsqueda, para o bien identificar una señal de audio directamente, o bien caracterizar la señal de audio en la medida en que se emita una medida de similitud de la señal de audio con respecto a una o varias de las señales de audio en la base de datos. Cuando se determina que si bien un tipo determinado de huella digital era favorable para una aplicación, no obstante ya no era favorable para la otra aplicación, para alcanzar un compromiso óptimo entre fuerza de identificación por un lado y espacio de almacenamiento por otro, para la gran cantidad de señales de audio, cuyas huellas digitales están almacenadas en la base de datos, tiene que realizarse por tanto un nuevo procesamiento de extracción de características para crear una base de datos de características nueva que sea un compromiso óptimo para las aplicaciones actuales. Por un lado los fragmentos originales ni siquiera están disponibles para una nueva extracción de características (para una base de datos de audio se utilizan por ejemplo 500.000 fragmentos de audio). Por otro lado esto da como resultado, en caso de que sea posible, un gran esfuerzo para el procesamiento de extracción de características, para llenar o “entrenar” la “nueva” base de datos.

Este problema se agrava en particular porque si bien por Internet está disponible una red mundial, que en principio tiene una capacidad de almacenamiento casi ilimitada, no obstante es imposible comunicar siempre a diferentes “productores de huellas digitales” la huella digital que es más favorable para una aplicación, de manera que siempre esté disponible suficiente material de base de datos de huella digital para poder realizar una identificación de señal de audio o caracterización útil.

Otra problemática consiste en que las huellas digitales también deberían transmitirse a través de los canales de transmisión más diferentes. Un canal de transmisión con una capacidad de transmisión muy reducida es por ejemplo un canal de transmisión exterior de un teléfono móvil. En este caso, además de la fuerza de identificación y la capacidad de almacenamiento para la base de datos el ancho de banda del canal de transmisión también desempeña un papel decisivo. No tendría sentido generar una huella digital con mucha fuerza de identificación, que entonces no podría transmitirse o sólo en una medida reducida por el canal de transmisión de banda estrecha. La huella digital óptima para una aplicación tal viene dictada por tanto adicionalmente por el canal de transmisión, por el que se transmitirá la huella digital por ejemplo de una base de datos de búsqueda.

El objetivo de la presente invención consiste en crear un concepto de huella digital flexible y adaptable a diferentes requisitos.

Este objetivo se soluciona mediante un procedimiento para generar una huella digital según la reivindicación 1, un procedimiento para caracterizar una señal de audio según la reivindicación 11, una representación de huella digital según la reivindicación 15, un dispositivo para generar una huella digital según la reivindicación 16 o un dispositivo para caracterizar una señal de audio según la reivindicación 17.

La presente invención se basa en el conocimiento de que puede obtenerse un concepto de huella digital lo más universal posible al haberse generado la huella digital según un modo de una pluralidad de modos de huella digital predeterminados genera, refiriéndose todos los modos de huella digital al mismo tipo de huella digital, proporcionando sin embargo los modos de huella digital huellas digitales diferentes, que por un lado se distinguen en su cantidad de datos y por otro lado en su intensidad de identificación para caracterizar una señal de audio. Los modos de huella digital están predeterminados según la invención de tal manera, que una huella digital según un modo de huella digital, que tiene una primera fuerza de identificación, puede convertirse en una huella digital según un modo de huella digital, que tiene una segunda fuerza de identificación, sin utilizar la propia señal de audio. Esta huella digital “escalable” permite, por ejemplo, proporcionar a una base de datos huellas digitales con mucha fuerza de identificación y complejidad de datos. La huella digital compleja de base de datos, que se ha generado según uno de los modos de huella digital predeterminados, puede convertirse fácilmente en una huella digital “más baja”, que tiene una menor fuerza de identificación, sin someter a la propia señal de audio a una extracción de características.

Preferiblemente la huella digital con una fuerza de identificación mayor de la huella digital de búsqueda y de la huella digital de base de datos se convierte siempre de tal manera, que se comparan dos huellas digitales entre sí, que también son comparables entre sí. Si se proporciona un escalado con muchos modos de huella digital diferentes según la invención, así la base de datos es adecuada para procesar tantas huellas digitales diferentes, es decir, huellas digitales generadas según modos de huella digital diferentes, como modos de huella digital están presentes en total.

Esto tiene la ventaja de que utilizando la misma base de datos de huella digital pueden procesarse huellas digitales con muy poca fuerza de identificación, pero también huellas digitales con mucha fuerza de identificación, de manera que dependiendo de los modos de huella digital predeterminados permitidos para cada aplicación puede hallarse un modo de huella digital adecuado, mientras que sin embargo siempre puede utilizarse la misma base de datos de huella digital.

Este concepto tiene además la ventaja de que los productores de huellas digitales para bases de datos de audio no tienen que generar continuamente otras huellas digitales para aplicaciones variables, más bien se genera una vez una huella digital escalable, que puede utilizarse entonces debido a su característica de escalabilidad para una pluralidad de aplicaciones. Por otro lado a los usuarios de tales bases de datos de búsqueda se les concede suficiente flexibilidad en el sentido de que, si las circunstancias lo requieren, pueden generar una huella digital con muy poca fuerza de identificación pero que se transmite muy rápidamente, mientras que el mismo usuario en otra aplicación, en la que depende más de la fuerza de identificación y menos de la cantidad de datos de la huella digital, puede generar una huella digital con mucha fuerza de identificación. El usuario puede recurrir en ambos casos de aplicación a la misma base de datos, lo que simplifica esencialmente todo el concepto de la descripción de señal de audio basada en el contexto y con ello lo hace más fácil para el usuario. En particular la facilidad para el usuario y la facilidad de manejo son decisivas para una amplia implantación comercial en el mercado.

Preferiblemente se emplea una escalabilidad con respecto a la frecuencia y/o una escalabilidad con respecto al tiempo. La escalabilidad con respecto a la frecuencia se consigue según un ejemplo de realización preferido de la presente invención, porque los modos de huella digital contienen en cada caso información de huella digital separada para subbandas separadas de la señal de audio, y porque los modos de huella digital se distinguen entre sí, en que comprenden información de huella digital separada para una cantidad diferente de subbandas. Esta determinación de las subbandas es igual para todos los modos de huella digital. Si se genera una base de datos con huellas digitales con mucha fuerza de identificación, es decir, si el modo de huella digital, en el que se han generado las huellas digitales para la base de datos, contiene información de huella digital separada para, por ejemplo, 20 subbandas de la señal de audio, entonces otros modos de huella digital con menos fuerza de identificación llevarían a huellas digitales, que por ejemplo, contendrían información de huella digital separada para sólo 15, 10, 5 o sólo una única subbanda. Todas las huellas digitales, que se han generado según los distintos modos de huella digital, son fácilmente compatibles hacia abajo, dado que en una operación de correspondencia con la base de datos de la huella digital de base de datos únicamente se toma la información de huella digital para las subbandas, que también están contenidas en la huella digital de búsqueda. En un caso extremo se utilizaría entonces para una operación de correspondencia de 20 informaciones de huella digital separadas de una huella digital únicamente la información de huella digital de una única subbanda, cuando la huella digital de búsqueda comprende igualmente sólo información de huella digital para una única subbanda.

Una alternativa preferida adicional es la escalabilidad a lo largo del tiempo. Una huella digital con relativamente poca fuerza de identificación contiene por ejemplo información de huella digital de 10 bloques de valores de muestreo de la señal de audio, mientras que una huella digital con más fuerza de identificación comprende información de huella digital por cada bloque de valores de muestreo de la señal de audio. En el caso de longitudes de bloque del mismo tamaño para ambas huellas digitales para la conversión hacia abajo un número correspondiente de informaciones de huella digital consecutivas en el tiempo se combina con una fuerza de identificación elevada, para generar a partir de éstas una huella digital convertida, que identifica la misma cantidad de valores de muestreo que la huella digital con poca

fuerza de identificación. Los modos de huella digital se distinguen a este respecto en el sentido de que un modo de huella digital con muy poca fuerza de identificación procesa muchos bloques de valores de muestreo en una única huella digital, mientras que un modo de huella digital con mucha fuerza de identificación genera, por ejemplo, por bloque una huella digital propia. En el caso de un tamaño de bloque determinado, sin que la propia señal de audio tenga que someterse de nuevo a una extracción de características, a partir de la correspondiente cantidad de modos de huella digital con fuerza de identificación puede generarse una huella digital con poca fuerza de identificación, para poder llevar a cabo por ejemplo una comparación de base de datos.

Ejemplos de realización preferidos de la presente invención se ilustran en detalle a continuación haciendo referencia a los dibujos adjuntos. Muestran:

la figura 1 un diagrama de bloques para generar una huella digital;

la figura 2 un diagrama de bloques de un dispositivo según la invención para caracterizar una señal de audio;

la figura 3a una representación de la división de una señal de audio en distintas subbandas;

la figura 3b una representación esquemática de distintas representaciones de huella digital, que pueden generarse mediante modos de huella digital diferentes a partir de la distribución de subbandas mostrada en la figura 3a;

la figura 4a una representación de una distribución de bloques de una señal de audio a lo largo del tiempo;

la figura 4b una vista general esquemática de diversas representaciones de huella digital, que pueden generarse según distintos modos de huella digital a partir de la división en bloques mostrada en la figura 4a; y

la figura 5 un diagrama de bloques esquemático de un sistema de reconocimiento de patrones.

A continuación se hace alusión a la figura 5, que muestra una vista general esquemática de un sistema de reconocimiento de patrones, en el que la presente invención puede utilizarse ventajosamente. En principio en el caso de un sistema de reconocimiento de patrones mostrado en la figura 5 se distingue entre dos modos de funcionamiento, en concreto el modo 50 de entrenamiento y el modo 52 de clasificación.

En el modo de entrenamiento los datos “se introducen por entrenamiento”, es decir, se añaden al sistema y a continuación se recogen en una base 54 de datos.

En el modo de clasificación se intenta comparar y ordenar una señal que va a caracterizarse con los registros presentes en la base 54 de datos.

El sistema de reconocimiento de patrones comprende un medio 56 para el procesamiento de señales, un medio 58 conectado aguas abajo para la extracción de características, un medio 60 para el procesamiento de características, un medio 62 para generación de agrupamiento, y un medio 64 para realizar una clasificación, para por ejemplo como resultado del modo 52 de clasificación llegar a una indicación tal sobre el contenido de la señal que va a caracterizarse, que la señal sea idéntica a la señal xy, que se ha introducido por entrenamiento en un modo de entrenamiento anterior.

A continuación se hará alusión a la funcionalidad de los bloques individuales de la figura 5.

El bloque 56 forma conjuntamente con el bloque 58 un extractor de características, mientras que el bloque 60 representa un procesador de características. El bloque 56 transforma una señal de entrada a un formato objetivo normalizado, tal como por ejemplo, la cantidad de canales, la tasa de muestreo, la resolución (en bits por valor de muestreo), etc. Esto es útil y necesario en este sentido, dado que no deberían hacerse suposiciones sobre la fuente, de la que proviene la señal de entrada.

El medio 58 para la extracción de características sirve para restringir la habitualmente gran cantidad de información a la salida del medio 56 a una pequeña cantidad de información. Las señales que van a examinarse tienen en la mayoría de los casos una tasa de transmisión de datos elevada, es decir, una alta cantidad de valores de muestreo por espacio de tiempo. La restricción a una pequeña cantidad de información debe tener lugar de modo que la esencia de la señal original, es decir, la singularidad de la misma, no se pierda. En el medio 58 se extraen de la señal propiedades características preestablecidas, como en general por ejemplo intensidad de sonido, frecuencia fundamental, etc. y/o, según la presente invención, características de tonalidad o la SFM. Las características de tonalidad así obtenidas deben contener por así decirlo la esencia de la señal examinada.

En el bloque 60 pueden procesarse los vectores de características calculados previamente. Un procesamiento sencillo consiste en la normalización de los vectores. Posibles procesamientos de características son transformaciones lineales, como por ejemplo la transformación de Karhunen-Loeve (KLT) o el análisis discriminante lineal (LDA), que se conocen en la técnica. Transformaciones adicionales en particular también no lineales son igualmente aplicables para el procesamiento de características.

El generador de clases sirve para combinar los vectores de características procesados en clases. Estas clases corresponden a una representación compacta de la señal correspondiente. El clasificador 64 sirve finalmente para clasificar un vector de características generado de una clase predefinida o una señal predefinida.

La figura 1 muestra esquemáticamente un dispositivo para generar una huella digital de una señal de audio, como puede estar presente por ejemplo en el bloque 58 de la figura 5. Para la generación de una huella digital de una señal de audio se utilizan informaciones que definen una pluralidad de modos de huella digital predeterminados, estando almacenadas estas informaciones de modo para modos de huella digital compatibles entre sí por medio de un medio 10. Los modos de huella digital definidos por las informaciones de modo almacenadas en el medio 10 se refieren todas al mismo tipo de huella digital, proporcionando sin embargo los modos de huella digital huellas digitales diferentes, que por un lado se distinguen en su cantidad de datos y por otro lado en su intensidad de identificación para identificar la señal de audio. Los modos de huella digital están predeterminados según la invención de modo que una huella digital según un modo de huella digital, que tiene una primera fuerza de identificación, puede convertirse en una huella digital según un modo de huella digital, que tiene una segunda fuerza de identificación, sin utilizar la propia señal de audio. Preferiblemente se prefiere una capacidad de conversión de la huella digital con una fuerza de identificación mayor a la huella digital con menor fuerza de identificación. Dependiendo del caso de aplicación también es posible, sin embargo, garantizar una compatibilidad ascendente por ejemplo por interpolación etc.

El dispositivo según la invención comprende además un medio 12 para ajustar un modo de huella digital a partir de la pluralidad de modos de huella digital predeterminados. En un medio 14 para calcular la huella digital según el modo de huella digital ajustado por el medio 12 se calcula finalmente una huella digital de una señal de audio alimentada a través de una entrada 16 y se emite por una salida 18. El medio 14 para calcular la huella digital según el modo de huella digital ajustado por el medio 12 está conectado con el medio 10 de almacenamiento, para dependiendo del modo de huella digital aplicar las correspondientes normas de cálculo.

A continuación se hace alusión con más detalle al medio 14 para calcular la huella digital según un modo de huella digital ajustado. Se ha comprobado que como característica, que por un lado es robusta y por otro lado puede escalarse de forma favorable, se adecua bien la tonalidad de una señal de audio.

Para el cálculo de la medida de tonalidad de un fragmento pueden aplicarse distintos procedimientos. Una señal temporal que va a caracterizarse puede transformarse por medio de un medio en el dominio espectral, para a partir de un bloque de valores de muestreo temporales generar un bloque de coeficientes espectrales. Como se expondrá posteriormente, para cada coeficiente espectral o para cada componente espectral puede determinarse un valor de tonalidad propio, para por ejemplo por medio de una determinación Sí/No clasificar, si una componente espectral es tonal o no. Utilizando los valores de tonalidad para las componentes espectrales y la energía o potencia de las componentes espectrales, puede calcularse entonces la medida de tonalidad para la señal de una pluralidad de maneras diferentes.

Debido al hecho de que se obtiene una medida de tonalidad cuantitativa, también es posible indicar distancias o similitudes entre dos fragmentos con indicación de tonalidad, pudiendo clasificarse los fragmentos como similares, si sus medidas de tonalidad se distinguen sólo por una diferencia menor que un umbral predeterminado, mientras que otros fragmentos pueden clasificarse como no similares, si sus índices de tonalidad se distinguen por una diferencia, que es mayor que un umbral de no similitud. Además de la diferencia entre dos medidas de tonalidad para la determinación de la distancia de tonalidad entre dos fragmentos pueden utilizarse magnitudes adicionales, como por ejemplo, el valor absoluto de la diferencia entre dos valores, el cuadrado de una diferencia, el cociente entre dos medidas de tonalidad menos uno, la correlación entre dos medidas de tonalidad, la métrica de distancia entre dos medidas de tonalidad, que son vectores n dimensionales, etc.

Se indica que la señal que va a caracterizarse no tiene que ser necesariamente una señal temporal, sino que la misma también puede ser por ejemplo una señal con codificación MP3, que se compone de una secuencia de palabras de código de Huffman, que se han generado a partir de valores espectrales cuantificados.

Los valores espectrales cuantificados se generaron a partir de los valores espectrales originales por cuantificación, seleccionándose la cuantificación de tal manera, que el ruido de cuantificación introducido por la cuantificación se encuentre por debajo del umbral de enmascaramiento psicoacústico. En un caso de este tipo puede utilizarse directamente el flujo de datos de MP3 codificados, para calcular por ejemplo por medio de un decodificador de MP3 los valores espectrales. No es necesario realizar antes de la determinación de la tonalidad una transformación en el dominio de tiempo y entonces de nuevo una transformación en el dominio espectral, sino que los valores espectrales calculados dentro del decodificador de MP3 pueden tomarse inmediatamente, para calcular la tonalidad por componente espectral o la SFM (SFM = Spectral Flatness Measure = Medida para la planeidad espectral). Cuando por tanto para la determinación de la tonalidad se utilizan componentes espectrales, y cuando la señal que va a caracterizarse es un flujo de datos de MP3, el medio 40 está configurado como un decodificador, sin embargo, sin el banco de filtros inverso.

La medida para la planeidad espectral (SFM) se calcula mediante la siguiente ecuación.

$$SFM = \frac{\left[\prod_{n=0}^{N-1} X(n)\right]^{\frac{1}{N}}}{\frac{1}{N}\sum_{n=0}^{N-1} X(n)}$$

En esta ecuación X(n) representa el cuadrado del valor absoluto de una componente espectral con el índice n, mientras que N representa la cantidad total de los coeficientes espectrales de un espectro. Por la ecuación puede verse que la SFM es igual al cociente de la media geométrica de las componentes espectrales con respecto a la media aritmética de las componentes espectrales. Como se conoce, la media geométrica es siempre menor o como mucho igual a la media aritmética, de manera que la SFM tiene un intervalo de valores, que se encuentra entre 0 y 1. A este respecto un valor cercano a 0 indica una señal tonal y un valor cercano a 1 una señal más bien de tipo ruido con un desarrollo espectral plano. Se indica que la media aritmética y la media geométrica sólo son iguales, cuando todos los X(n) son idénticos, lo que corresponde a una señal completamente atonal, es decir de tipo ruido o pulsante. Por el contrario si en el caso extremo únicamente una componente espectral es muy grande desde el punto de vista de su valor, mientras que otras componentes espectrales X(n) son muy pequeñas desde el punto de vista de su valor, así la SFM tendrá un valor cercano a 0, lo que indica una señal muy tonal.

La SFM está descrita en “Digital Coding of Waveforms”, Englewood Cliffs, NJ, Prentice-Hall, N. Jayant, P. Noll, 1984, y se definió originalmente como medida para la obtención de codificación máxima que va a alcanzarse a partir de una reducción de redundancia. A partir de la SFM puede determinarse entonces la medida de tonalidad.

Una posibilidad adicional para determinar la tonalidad de los valores espectrales consiste en la determinación de picos en el espectro de densidad de potencia de la señal de audio, como está descrito en MPEG-1 Audio ISO/IEC 11172-3, anexo D1 “Psychoacoustic Model 1”. A este respecto se determina el nivel de una componente espectral. Además se determinan los niveles de dos componentes espectrales que rodean una componente espectral. Una clasificación de la componente espectral como tonal tiene lugar cuando el nivel de la componente espectral es mayor en un factor predeterminado que un nivel de una componente espectral circundante. El umbral predeterminado se toma en el estado de la técnica como de 7 dB, pudiendo utilizarse para la presente invención sin embargo cualquier otro umbral predeterminado. De esta manera para cada componente espectral puede indicarse si es tonal o no.

Una posibilidad adicional para la determinación de la tonalidad de una componente espectral consiste en la evaluación de la capacidad de predicción temporal, es decir la capacidad de pronóstico, de la componente espectral. A este respecto se remite de nuevo a MPEG-1 Audio ISO/IEC 11172-3, anexo D2 “Psychoacoustic Model 2”. En general se transforma un bloque actual de valores de muestreo de la señal que va a caracterizarse en una representación espectral, para obtener un bloque actual de componentes espectrales. A continuación se predicen las componentes espectrales del bloque actual de componentes espectrales utilizando información a partir de valores de muestreo de la señal que va a caracterizarse, que preceden al bloque actual, es decir utilizando información del pasado. Además se determina un error de predicción, a partir del que entonces puede deducirse una medida de tonalidad.

Una posibilidad adicional para la determinación de la tonalidad está descrita en la patente estadounidense n.º 5.918.203. De nuevo se utiliza una representación de valor real positiva del espectro de la señal que va a caracterizarse. Esta representación puede comprender los valores absolutos, los cuadrados de los valores absolutos, etc. de las componentes espectrales. En un ejemplo de realización se comprimen los valores absolutos o cuadrados de los valores absolutos de las componentes espectrales en primer lugar de manera logarítmica y después se filtran con un filtro con característica diferenciadora, para obtener un bloque de componentes espectrales filtradas de manera diferenciada.

En otro ejemplo de realización los valores absolutos de las componentes espectrales se filtran en primer lugar con un filtro con característica diferenciadora, para obtener un numerador, y después se filtra con un filtro con característica integradora, para obtener un denominador. El cociente de un valor absoluto filtrado de manera diferenciada de una componente espectral y el valor absoluto filtrado de forma integrada de la misma componente espectral arroja entonces el valor de tonalidad para esta componente espectral.

Mediante estas dos formas de proceder se suprimen las modificaciones lentas entre valores absolutos contiguos de componentes espectrales, mientras que se potencian las modificaciones bruscas entre valores absolutos contiguos de componentes espectrales en el espectro. Las modificaciones lentas entre valores absolutos contiguos de componentes espectrales indican componentes de señal atonales, mientras que las modificaciones bruscas indican componentes de señal tonales. Las componentes espectrales comprimidas de manera logarítmica y filtradas de manera diferencia o los cocientes pueden utilizarse entonces a su vez para calcular una medida de tonalidad para el espectro contemplado.

Aunque en el texto anterior se habló de calcular un valor de tonalidad por componente espectral, se prefiere con vistas a un menor esfuerzo de cálculo, por ejemplo sumar siempre los cuadrados de valores absolutos de dos componentes espectrales contiguas y calcular después para cada resultado de la suma un valor de tonalidad mediante uno de los procedimientos mencionados. Puede utilizarse todo tipo de agrupamiento aditivo de cuadrados de valores absolutos o

valores absolutos de componentes espectrales para calcular valores de tonalidad para más de una componente espectral.

Una posibilidad adicional para la determinación de la tonalidad de una componente espectral consiste en comparar el nivel de una componente espectral con un valor medio de niveles de componentes espectrales en una banda de frecuencia. La anchura de la banda de frecuencia, en la que se encuentra la componente espectral, cuyo nivel se compara con el valor medio por ejemplo, de los valores absolutos o cuadrados de valores absolutos de las componentes espectrales, puede seleccionarse dependiendo del requisito. Una posibilidad consiste por ejemplo en seleccionar la banda de modo que sea estrecha. Alternativamente la banda también podría seleccionarse de modo que sea ancha, o también según puntos de referencia psicoacústicos. De este modo puede minimizarse la influencia de caídas de potencia momentáneas en el espectro.

Aunque en lo anterior se determinó la tonalidad de una señal de audio mediante sus componentes espectrales, esto también puede producirse en el dominio de tiempo, es decir utilizando los valores de muestreo de la señal de audio. Para ello podría realizarse un análisis LPC de la señal, para estimar una ganancia de predicción para la señal. La ganancia de predicción es inversamente proporcional a la medida SFM y también es una medida para la tonalidad de la señal de audio.

En un ejemplo de realización preferido de la presente invención no sólo se indica un valor por cada espectro a corto plazo, sino que la medida de tonalidad es un vector multidimensional de valores de tonalidad. Así, por ejemplo, el espectro a corto plazo puede dividirse en cuatro zonas o bandas de frecuencia contiguas entre sí y preferiblemente sin solapamiento, determinándose para cada banda de frecuencia un valor de tonalidad. De este modo, para un espectro a corto plazo de la señal que va a caracterizarse, se obtiene un vector de tonalidad de 4 dimensiones. Para permitir una mejor caracterización, se preferiría además tratar, como se describió anteriormente, por ejemplo cuatro espectros a corto plazo consecutivos, de modo que en total se obtiene una medida de tonalidad, que es un vector de 16 dimensiones o en general un vector de n x m dimensiones, siendo n la cantidad de componentes de tonalidad por cada trama o bloque de valores de muestreo, mientras que m es la cantidad de bloques o espectros a corto plazo considerados. La medida de tonalidad sería entonces, como se ha explicado, un vector de 16 dimensiones. Para considerar mejor el transcurso en el tiempo de la señal que va a caracterizarse, se prefiere además calcular varios vectores de este tipo por ejemplo de 16 dimensiones y a continuación procesarlos de manera estadística, para por ejemplo calcular la varianza, el valor medio o momentos centrales de orden mayor a partir de todos los vectores de tonalidad de n x m dimensiones de un fragmento con una determinada longitud, para de este modo indicar este fragmento.

En general de este modo puede calcularse la tonalidad a partir de partes de todo el espectro. De este modo es posible determinar la tonalidad/tipo de ruido de un subespectro o varios subespectros y de este modo conseguir una caracterización más fina del espectro y así de la señal de audio.

Además pueden calcularse estadísticas a corto plazo a partir de valores de tonalidad, como por ejemplo el valor medio, la varianza y momentos centrales de orden mayor, como medida de tonalidad. Éstos se determinan por medio de técnicas estadísticas mediante una secuencia temporal de valores de tonalidad o vectores de tonalidad y proporcionan por tanto una esencia sobre una sección más larga de un fragmento.

Además también pueden utilizarse diferencias de vectores de tonalidad consecutivos en el tiempo o valores de tonalidad filtrados de manera lineal, pudiendo utilizarse como filtros lineales por ejemplo filtros IIR o filtros FIR.

También en el cálculo de la SFM se prefiere por motivos de ahorro de tiempo de cálculo sumar o promediar por ejemplo dos cuadrados del valor absoluto contiguos respecto a la frecuencia y realizar el cálculo de SFM sobre esta representación espectral positiva y de valor real aproximada. Esto lleva además a una mayor robustez con respecto a caídas de frecuencia de banda estrecha así como a un menor esfuerzo de cálculo.

De nuevo, con referencia a la figura 1, a continuación se hace alusión en más detalle al medio 12 para ajustar un modo de huella digital de los modos de huella digital predeterminados. El medio 12 se encarga de elegir y ajustar el modo de huella digital más adecuado para una aplicación determinada a partir de la pluralidad de modos de huella digital predefinidos. La elección puede realizarse o bien de manera empírica o bien de manera automática mediante operaciones de correspondencia de comprobación establecidas de manera fija. Con tales operaciones de correspondencia de comprobación se procesan por ejemplo varias señales de audio conocidas según diferentes modos de huella digital, para generar diferentes huellas digitales con fuerza de identificación. Con estas diferentes huellas digitales, que sin embargo presentan todas el mismo tipo de huella digital, concretamente por ejemplo la tonalidad o una medida de tonalidad de la señal de audio, en una base de datos se realiza entonces una operación de correspondencia de patrones. Mediante un umbral preestablecido de salidas de datos para las huellas digitales individuales entonces puede elegirse uno de los modos de huella digital predeterminados, que por ejemplo cumpla con un criterio umbral.

Alternativamente el medio 12 puede elegir independientemente de valores umbral, por ejemplo dictado por un canal de transmisión el modo de huella digital, que proporciona una huella digital, que por su cantidad de datos por ejemplo, todavía puede transmitirse por un canal de transmisión con banda limitada. Según la ocupación del canal o la capacidad disponible del canal entonces puede ajustarse o bien un modo de huella digital con fuerza de identificación o, cuando el

canal está muy ocupado o muy limitado con respecto a la banda, o bien un modo de huella digital con una identificación relativamente reducida.

Lo mismo es válido para el caso de que la huella digital no tenga que transmitirse, sino almacenarse. Entonces según los recursos de almacenamiento disponibles puede ajustarse un modo de huella digital que necesita más capacidad de almacenamiento y así con más fuerza de identificación o un modo de huella digital que ahorra más capacidad de almacenamiento pero con relativamente menor capacidad de identificación mediante el medio 12.

La figura 2 muestra un diagrama de bloques de un dispositivo según la invención para caracterizar una señal de audio. Un dispositivo de este tipo comprende un medio para generar una huella digital de búsqueda en uno de los modos de huella digital predeterminados. Este medio se indica en la figura 2 con el número de referencia 20 y preferiblemente está configurado como se ha descrito en relación con la figura 1. El dispositivo para caracterizar una señal de audio comprende además una base 22 de datos, en la que están almacenadas huellas digitales de base de datos, que también se han calculado en uno de los modos de huella digital predeterminados.

El dispositivo mostrado en la figura 2 comprende además un medio 24 para comparar la huella digital de búsqueda, que se ha generado mediante el medio 20, con las huellas digitales de base de datos. En primer lugar en un medio 24a se determina si la huella digital de búsqueda y la huella digital de base de datos que va a compararse con la misma tienen la misma fuerza de identificación, es decir, si se han generado mediante el mismo modo de huella digital, o si la huella digital de búsqueda se ha generado según otro modo de huella digital que la huella digital de base de datos. Si se determina que una de las huellas digitales tiene una fuerza de identificación mayor que la otra, en un medio 24b se realiza una conversión, de manera que tras la conversión tanto la huella digital de búsqueda como la huella digital de base de datos tienen la misma fuerza de identificación, es decir, son comparables o están presentes según el mismo modo de huella digital. Sólo cuando se ha cumplido este requisito un medio 24c realizará una comparación de ambas huellas digitales. Una comparación tal dará entonces dado el caso como resultado 26, que la señal de audio, representada por la huella digital de búsqueda, corresponde a la señal de audio, representada por la huella digital de base de datos actual. Alternativamente el resultado 26 puede consistir también en que se determina una similitud con una cierta probabilidad, es decir una medida de similitud.

Preferiblemente el medio 24a está dispuesto para determinar la huella digital que tiene la mayor fuerza de identificación. Esta huella digital se escala entonces de manera descendente hasta la fuerza de identificación, es decir hasta el modo de huella digital de la huella digital, que tiene la fuerza de identificación menor de las dos huellas digitales. Alternativamente, cuando se desea por ejemplo por motivo de una búsqueda rápida, ambas huellas digitales pueden escalarse de manera descendente en un modo de huella digital, que proporciona huellas digitales más débiles con respecto a la identificación de lo que lo son la huella digital de búsqueda y la huella digital de base de datos.

Según la aplicación también puede ser necesario escalar de manera ascendente la huella digital con la fuerza de identificación menor por ejemplo por medio de interpolación, proporcionando esta alternativa sin embargo sólo resultados útiles, cuando el tipo de huella digital permite una interpolación.

Como ya se ha explicado existen requisitos contrarios en la determinación del modo de huella digital. Por un lado es de gran interés conseguir una reducción de datos lo mayor posible, es decir, un tamaño de huella digital reducido, para poder mantener el mayor número posible de huellas digitales de este tipo en la memoria de un ordenador y para hacer que el procesamiento posterior sea más eficaz.

Por otro lado a medida que disminuye el tamaño de huella digital aumenta el riesgo de que entre los fragmentos registrados en la base de datos no se pueda ya distinguir de manera correcta. Esto se aplica en particular en el caso de una base de datos de fragmentos de audio grande, que por ejemplo puede comprender 500.000 títulos así como en el caso de aplicaciones en las que los fragmentos de audio están sometidos a distorsiones importantes antes de la operación de reconocimiento, por ejemplo en el caso de una transmisión acústica de la señal o en el caso de una compresión con pérdidas.

Evidentemente sería posible definir por este motivo formatos de huella digital más compactos, que sean menos robustos y formatos menos compactos, que ofrezcan propiedades discriminantes correspondientemente mejores. Sin embargo, como se ha explicado, esto hace necesario que las bases de datos de huella digital complejas tengan que crearse varias veces, es decir una vez en cada formato y mantenerse disponibles, especialmente porque una descripción en un primer tipo de huella digital en general no puede compararse con una huella digital de otro tipo.

Para solucionar este problema, la presente invención proporciona un formato de descripción escalable universal que según la aplicación proporciona de manera flexible un compromiso diferente entre la intensidad de identificación y la compacidad de la huella digital, sin que se pierda la propiedad de la posibilidad de comparar las huellas digitales. Preferiblemente esto se consigue mediante una escalabilidad en dos dimensiones, siendo una dimensión la escalabilidad en la cantidad de bandas y la otra dimensión la escalabilidad en el tiempo. En general la escalabilidad en la cantidad de bandas se basa en una descomposición espectral de la señal de audio. El dominio de frecuencia de la señal de audio o una cantidad parcial del mismo, por ejemplo de 250 Hz a 4 kHz, se divide en bandas de frecuencia, en las que se calculan las huellas digitales basándose en las características observadas, por ejemplo la medida de tonalidad. Mediante la separación de frecuencia en cada una de las bandas existe una información independiente sobre

la característica de la señal. Todas las huellas digitales utilizan la misma división en bandas y comienzan preferiblemente en la misma frecuencia límite inferior. Un modo de huella digital, que proporciona una huella digital compacta con una fuerza de identificación menor, contiene sin embargo menos bandas y de este modo un intervalo de frecuencia menor que un modo de huella digital más generoso, que sin embargo tiene una forma menos compacta. A pesar de ello los dos tipos de descripción pueden compararse de manera útil sin un nuevo procesamiento de la señal de audio en sus bandas de frecuencia comunes.

Una forma de realización preferida consiste en la utilización de una división en bandas al menos parcialmente logarítmica, que para frecuencias no demasiado bajas, por ejemplo para frecuencias mayores que 500 Hz, se encuentra cerca de la escala de frecuencias o resolución de frecuencia utilizada por el oído humano. Se prefiere utilizar la división logarítmica mencionada anteriormente sólo a partir de, por ejemplo, 500 Hz, y dividir las bandas por debajo de 500 Hz por ejemplo para que tengan la misma anchura, como por ejemplo en cinco bandas de 100 Hz cada una. Esta división corresponde aproximadamente a la escala Bark.

A continuación, mediante las figuras 3a y 3b se proporciona un ejemplo para la escalabilidad en la cantidad de bandas. Por motivos de representación, tal como se muestra en la figura 3a, una señal de audio está dividida en cuatro subbandas 30a a 30d. La figura 3b muestra diferentes representaciones de huella digital, tal como pueden generarse mediante diferentes modos de huella digital. Cada representación de huella digital de la figura 3b contiene una sección 31 de identificador, que indica para cuántas subbandas está contenida información de huella digital, es decir según qué modo de huella digital se ha generado la huella digital en cuestión. El modo de huella digital n.º 4 proporciona la representación de huella digital que necesita más capacidad de almacenamiento, aunque también la que tiene más fuerza de identificación, porque la representación de huella digital contiene información de huella digital (IHD) tanto para la subbanda 1 como para las tres subbandas 2 a 4 restantes. El modo de huella digital n.º 3 proporciona por el contrario una representación de huella digital algo más compacta, aunque ya con una fuerza de identificación menor, porque la información de huella digital separada sólo está contenida para las primeras tres subbandas. El modo de huella digital n.º 2 proporciona una representación relativamente compacta aunque con una fuerza de identificación aún menor, porque la información de huella digital sólo está contenida para las dos subbandas más inferiores. El modo de huella digital n.º 1 proporciona por el contrario una representación de huella digital muy compacta aunque con la menor fuerza de identificación, porque sólo está contenida información de huella digital de la subbanda 30a más baja.

A continuación, mediante la figura 3b se hace alusión a la función del bloque 24b de la figura 2, es decir, a la conversión de huella digital de un modo de huella digital a otro modo de huella digital. Sólo a modo de ejemplo supóngase que se ha generado una huella digital de base de datos según el modo de huella digital n.º 4. La base de datos contiene por tanto huellas digitales con una fuerza de identificación considerable. Una huella digital de búsqueda se ha generado por ejemplo según el modo de huella digital n.º 2. Después de que el medio 24a de la figura 2 por ejemplo mediante el identificador 31 de huella digital de la figura 3b haya determinado que la huella digital de búsqueda y la huella digital de base de datos se han generado según diferentes modos de huella digital, la huella digital con la mayor fuerza de identificación, es decir la huella digital de base de datos se somete a una conversión. La conversión consiste en el ejemplo de realización mostrado en la figura 3b en que la información de huella digital de la tercera subbanda y la información de huella digital de la cuarta subbanda del modo de huella digital de base de datos ya no se siguen teniendo en cuenta, es decir, que en la operación de correspondencia no desempeñan ningún papel. Por tanto sólo se comparan entre sí información de huella digital de la primera subbanda e información de huella digital de la segunda subbanda. Alternativamente la huella digital de base de datos, que se ha generado según el modo de huella digital n.º 4, y la huella digital de búsqueda, que se ha generado según el modo de huella digital n.º 2, también podrían convertirse en el modo de huella digital n.º 1, lo que es particularmente ventajoso cuando se desea una operación de correspondencia rápida.

Se hace referencia a que no es esencial que la huella digital de base de datos tenga más fuerza de identificación que la huella digital de búsqueda. Si, por ejemplo, sólo existe una base de datos con menos fuerza de identificación, anterior, mientras que las huellas digitales de búsqueda son huellas digitales con mayor fuerza de identificación, entonces puede procederse a la inversa, de modo que las huellas digitales de búsqueda se conviertan en una forma con menos fuerza de identificación aunque más compacta y que entonces se realice la operación de correspondencia.

Aunque en la figura 3a están ilustradas las subbandas 1 a 4 (30a a 30d sin solapamiento), se indica que ya un pequeño solapamiento de las subbandas lleva a una robustez mayor de los cambios de altura de sonido. Para aumentar la robustez de la representación con respecto a los cambios de señal, que contienen un cambio de la altura de sonido de señal, por ejemplo una conversión de la tasa de muestreo o una modificación de la altura de sonido de una señal, que se reproduce algo más rápido o algo más lento, se prefiere un determinado solapamiento de banda. Con un cambio de altura de sonido aparece el problema de que partes de una señal, que con una señal sin modificar se encuentran en una determinada banda de frecuencia (n), mediante el cambio de frecuencia, por ejemplo una extensión o contracción del espectro, llegan a disponerse en ciertos casos en la banda n-1 o n+1, de modo que de manera evidente aparecen otros valores de característica y la tasa de reconocimiento disminuye de manera correspondiente. En el ejemplo de realización preferido de la presente invención este efecto se reduce porque entre bandas adyacentes existe un determinado dominio de frecuencia, por ejemplo en el que se utiliza una cantidad de líneas DFT en ambas bandas de frecuencia. Como punto de referencia se prefiere un solapamiento de los dominios de frecuencia de por ejemplo un 10%, pudiendo seleccionarse más grande, cuando son de esperar extensiones o contracciones más intensas del espectro.

Una dimensión adicional de la escalabilidad viene dada por el tiempo. Mediante el uso del valor medio y la varianza para combinar una cantidad n de valores de característica individuales puede ajustarse la granularidad en el tiempo de la huella digital. Una descripción compacta selecciona un valor mayor para n y de este modo una combinación en el tiempo mayor que una descripción más generosa, aunque menos compacta. Para ilustrar esto, a continuación se hace alusión a las figuras 4a y 4b. La figura 4a muestra un procesamiento por bloques de una señal de audio u(t) a lo largo del tiempo t, estando representados por motivos de claridad cuatro bloques 40a a 40d consecutivos en el tiempo. Los bloques 40a a 40d tienen todos la misma longitud, es decir, la misma cantidad de valores de muestreo. Un modo de huella digital n.º 3 proporcionará una representación con una fuerza de identificación considerable, ya que para cada bloque 1 a bloque 4 se calcula y almacena información de huella digital propia. Por el contrario, el modo de huella digital n.º 2 ya proporciona una representación de huella digital con menos fuerza de identificación, aunque más compacta desde el punto de vista del almacenamiento, porque siempre la información de huella digital se forma a partir de dos bloques consecutivos, es decir, por un lado por el bloque 1 y el bloque 2 y por otro lado por el bloque 3 y el bloque 4. La representación de huella digital más favorable desde el punto de vista del almacenamiento, aunque con la fuerza de identificación menor proporciona finalmente el modo de huella digital n.º 1, que comprende información de huella digital de todos los bloques 1 a 4.

Cuando en una base de datos está almacenada información de huella digital, que se ha generado según el modo de huella digital n.º 3, y cuando la huella digital de búsqueda se ha generado según el modo de huella digital n.º 2, entonces se convierte la huella digital de base de datos, y concretamente de tal manera, que se combinan los dos primeros bloques y entonces se comparan con la primera información de huella digital de la huella digital de búsqueda, repitiéndose este modo de proceder para los siguientes bloques 3 y 4 en el tiempo. En este punto también sería posible de nuevo convertir tanto la huella digital de base de datos como la huella digital de búsqueda en una representación de huella digital según el modo de huella digital n.º 1.

En aplicaciones reales se prefiere combinar información de huella digital de n bloques de tal manera, que la representación de huella digital contenga el valor medio/o la varianza de la información de huella digital de los bloques individuales. El valor medio y la varianza están definidos de la siguiente manera:

Valor medio:

$$M_n(F) = 1/n \cdot \sum_{i=0}^{n-1} F_i$$

Varianza:

$$V_n(F) = 1/n \cdot \sum_{i=0}^{n-1} (F_i - M_n(F))^2 = \left[1/n \cdot \sum_{i=0}^{n-1} F_i^2 \right] - M_n(F)^2$$

En las ecuaciones mencionadas anteriormente n es un índice, que indica cuánta información de huella digital F_i de cuántos bloques o bandas, etc. se combinan para formar el valor medio M_n de los mismos. En la definición anterior de la varianza se indica que la varianza de un bloque o banda, que no tiene gran valor informativo, es igual a cero.

Haciendo referencia a la figura 4b, la información de huella digital del bloque 1 de la representación de huella digital, que se ha generado mediante el modo de huella digital n.º 3, comprenderá el valor medio y/o la varianza de características de audio. Lo mismo valdría para la información de huella digital para el bloque 2 de la representación de huella digital, que se ha generado mediante el modo de huella digital n.º 3. Ahora, para convertir las dos informaciones de huella digital para el bloque 1 y el bloque 2 de la representación de huella digital según el modo de huella digital n.º 3 en información de huella digital de la representación de huella digital, que se ha generado según el modo de huella digital n.º 2, tal como se representa mediante la línea 42, la información de huella digital de la representación de huella digital debe convertirse según el modo de huella digital n.º 3 de la siguiente manera:

Valor medio:

$$M_{2n}(F) = 0.5 \cdot \left[M_n(F) + M'_n(F) \right]$$

Varianza:

$$V_{2n}(F) = 0.5 \cdot \left[V_n(F) + V'_n(F) + 0.5 \cdot \left(M_n(F) - M'_n(F) \right)^2 \right]$$

Los valores medios y las varianzas pueden compararse entre sí cuando la granularidad en el tiempo de una representación de huella digital es un múltiplo de número entero de la granularidad de las demás representaciones de huella digital. Las ecuaciones anteriores son válidas para un factor a modo de ejemplo de 2. En la ecuación las magnitudes N_n y V_n representan los valores correspondientes de valor medio o varianza para la información de huella digital del bloque 1 según el modo de huella digital n.º 3, mientras que M_n' y V_n' representan los valores de valor medio o varianza para el bloque 2 de la representación de huella digital según el modo de huella digital n.º 3 de la figura 4b. Para el caso de que la varianza se utilice como información de huella digital, también debe estar presente el valor medio por ejemplo como información de huella digital adicional para garantizar la escalabilidad.

Se indica que de manera análoga también puede compararse la información de huella digital de la representación de huella digital según la información de huella digital según el modo de huella digital n.º 1

De este modo pueden compararse representaciones de huella digital de diferente granularidad en el tiempo, es decir, según diferentes modos de huella digital, por ejemplo mediante la transformación de la representación más fina en una más basta.

La representación de huella digital según la invención puede definirse por ejemplo como la denominada serie escalable, tal como se describe en el párrafo 4.2 del documento ISO/IEC JTC 1/SC 29/WG11 (MPEG), Information technology - multimedia content description interface- parte 4: Audio, 27. 10. 2000.

REIVINDICACIONES

1. Procedimiento para generar una huella digital de una señal de audio utilizando información (10) de modo, que define una pluralidad de modos de huella digital predeterminados, refiriéndose todos los modos de huella digital al mismo tipo de huella digital, proporcionando sin embargo los modos de huella digital huellas digitales diferentes que pueden escalarse con respecto al tiempo y/o la frecuencia, que por un lado se distinguen en su cantidad de datos y por otro lado en su intensidad de identificación para caracterizar la señal de audio, distinguiéndose entre sí los modos de huella digital, porque comprenden información de huella digital por separado para una cantidad diferente de subbandas, o presentando la huella digital escalable información de huella digital para una cantidad de bloques temporales que depende del modo de huella digital, con las etapas siguientes:

ajustar (12) un modo de huella digital predeterminado a partir de la pluralidad de modos de huella digital predeterminados; y

calcular (14) una huella digital escalable según el modo de huella digital predeterminado ajustado aplicando normas de cálculo según la información de modo para el modo de huella digital ajustado.

2. Procedimiento según la reivindicación 1, en el que la huella digital puede convertirse según un modo de huella digital con una fuerza de identificación mayor en una huella digital según un modo de huella digital con una fuerza de identificación menor.

3. Procedimiento según la reivindicación 1 o 2, que presenta además la etapa siguiente:

transmitir o almacenar la huella digital generada a través de un canal de transmisión con capacidad de transmisión limitada o a un medio de almacenamiento con capacidad de almacenamiento limitada,

ajustándose en la etapa del ajuste (12) de un modo de huella digital de los modos de huella digital predeterminados en función del canal de transmisión o de la capacidad de almacenamiento.

4. Procedimiento según una de las reivindicaciones anteriores, en el que el tipo de huella digital se refiere a propiedades de tonalidad de la señal de audio.

5. Procedimiento según una de las reivindicaciones anteriores, en el que la señal de audio puede descomponerse en una cantidad predeterminada de bandas de frecuencia (30a a 30d) predefinidas,

en el que cada modo de huella digital comprende la generación de información de huella digital por cada banda de frecuencia predefinida, distinguiéndose los modos de huella digital en la cantidad de información de huella digital, de modo que un primer modo de huella digital como huella digital por separado para cada banda de frecuencia comprende una primera cantidad de información de huella digital para una primera cantidad de bandas de frecuencia, y un segundo modo de huella digital como huella digital por separado para cada banda de frecuencia comprende una segunda cantidad de información de huella digital para una segunda cantidad de bandas de frecuencia, distinguiéndose la primera cantidad de la segunda cantidad, y siendo las bandas de frecuencia predefinidas iguales para todos los modos de huella digital.

6. Procedimiento según la reivindicación 5, en el que la descomposición de la señal de audio en las bandas de frecuencia predefinidas presenta al menos en parte una división en bandas logarítmica.

7. Procedimiento según la reivindicación 5 o 6, en el que dos bandas de frecuencia adyacentes según la frecuencia tienen una zona de solapamiento, perteneciendo las componentes espectrales en la zona de solapamiento a ambas bandas de frecuencia adyacentes.

8. Procedimiento según una de las reivindicaciones 5 a 7, en el que la banda de frecuencia, que comprende la frecuencia más baja, está contenida en todos los modos de huella digital, distinguiéndose los modos de huella digital en la cantidad de las bandas de frecuencia que siguen a las frecuencias mayores.

9. Procedimiento según una de las reivindicaciones 1 a 8,

en el que la señal de audio puede descomponerse en bloques (40a a 40d) consecutivos en el tiempo de una longitud predeterminada,

en el que en la generación de una huella digital se determina información de huella digital por cada bloque, distinguiéndose los modos de huella digital en la cantidad de bloques, que se representan por información de huella digital, y siendo la longitud de los bloques para todos los modos de huella digital igual.

10. Procedimiento según la reivindicación 9, en el que en un primer modo de huella digital como información de huella digital están comprendidos el valor medio y/o la varianza a partir de una primera cantidad predefinida de bloques, y un segundo modo de huella digital comprende el valor medio y/o la varianza a partir de una segunda cantidad

predefinida de bloques, encontrándose la primera cantidad predefinida y la segunda cantidad predefinida en una relación de número entero.

11. Procedimiento para caracterizar una señal de audio, con las etapas siguientes:

generar una huella digital de búsqueda de la señal de audio utilizando información (10) de modo, que define una pluralidad de modos de huella digital predeterminados, refiriéndose todos los modos de huella digital al mismo tipo de huella digital, proporcionando sin embargo los modos de huella digital huellas digitales diferentes que pueden escalarse con respecto al tiempo y/o la frecuencia, que por un lado se distinguen en su cantidad de datos y por otro lado en su intensidad de identificación para caracterizar la señal de audio, distinguiéndose entre sí los modos de huella digital, porque comprenden información de huella digital por separado para una cantidad diferente de subbandas, o presentando la huella digital escalable información de huella digital para una cantidad de bloques temporales que depende del modo de huella digital, con las etapas siguientes:

ajustar (12) un modo de huella digital predeterminado a partir de la pluralidad de modos de huella digital predeterminados; y

calcular (14) una huella digital escalable según el modo de huella digital predeterminado ajustado aplicando normas de cálculo según la información de modo para el modo de huella digital ajustado;

comparar (24) la huella digital calculada con una pluralidad de huellas digitales almacenadas, que representan señales de audio conocidas, para caracterizar la señal de audio, habiéndose generado las huellas digitales almacenadas según uno de la pluralidad de modos de huella digital, presentando la etapa de comparación (24) las subetapas siguientes:

examinar (24a), si la huella digital de búsqueda y la huella digital de base de datos se han generado según modos de huella digital diferentes;

convertir (24b) la huella digital de búsqueda y/o la huella digital de base de datos, de modo que las huellas digitales que van a compararse estén presentes según el mismo modo de huella digital; y

realizar (24c) la comparación utilizando las huellas digitales presentes en el mismo modo de huella digital.

12. Procedimiento según la reivindicación 11, en el que cada modo de huella digital comprende la generación de información de huella digital por cada banda de frecuencia predefinida, distinguiéndose los modos de huella digital en la cantidad de información de huella digital, de modo que un primer modo de huella digital como huella digital por separado para cada banda de frecuencia comprende una primera cantidad de información de huella digital para una primera cantidad de bandas de frecuencia, y un segundo modo de huella digital como huella digital por separado para cada banda de frecuencia comprende una segunda cantidad de información de huella digital para una segunda cantidad de bandas de frecuencia, distinguiéndose la primera cantidad de la segunda cantidad, presentando la etapa de conversión (24b) la supresión de información de huella digital para subbandas.

13. Procedimiento según la reivindicación 11,

en el que la señal de audio puede descomponerse en bloques (40a a 40d) consecutivos en el tiempo de una longitud predeterminada,

en el que en la generación de una huella digital se determina información de huella digital por cada bloque, distinguiéndose los modos de huella digital en la cantidad de bloques, que se representan por información de huella digital, y siendo la longitud de los bloques para todos los modos de huella digital igual, y

en el que la etapa de conversión (24b) presenta la etapa de combinar la información de huella digital de bloques consecutivos en el tiempo.

14. Procedimiento según la reivindicación 13,

en el que la información de huella digital comprende un valor medio y/o una varianza, y

en el que existe una relación de número entero entre los bloques combinados en la huella digital de búsqueda y los bloques combinados en la huella digital de base de datos.

15. Representación de huella digital para una señal de audio, con las características siguientes:

una huella digital que puede escalarse con respecto al tiempo y/o la frecuencia, estando configurada la huella digital según uno de una pluralidad de modos de huella digital predeterminados, refiriéndose todos los modos de huella digital al mismo tipo de huella digital, proporcionando sin embargo los modos de huella digital huellas digitales diferentes que pueden escalarse con respecto al tiempo y/o la frecuencia, que por un lado se distinguen en su cantidad de datos y por otro lado en su intensidad de identificación para caracterizar la señal de audio, presentando la huella digital escalable información de huella digital separada para subbandas

separadas de la señal de audio, distinguiéndose los modos de huella digital entre sí, porque comprenden información de huella digital separada para una cantidad diferente de subbandas, o presentando la huella digital escalable información de huella digital para una cantidad de bloques temporales dependientes del modo de huella digital; y

un indicador (31), que indica el modo de huella digital, que se basa en la huella digital.

16. Dispositivo para generar una huella digital de una señal de audio utilizando información (10) de modo, que define una pluralidad de modos de huella digital predeterminados, refiriéndose todos los modos de huella digital al mismo tipo de huella digital, proporcionando sin embargo los modos de huella digital huellas digitales diferentes que pueden escalarse con respecto al tiempo y/o la frecuencia, que por un lado se distinguen en su cantidad de datos y por otro lado en su intensidad de identificación para caracterizar la señal de audio, distinguiéndose los modos de huella digital entre sí, porque comprenden información de huella digital separada para una cantidad diferente de subbandas, o presentando la huella digital escalable información de huella digital para una cantidad de bloques temporales dependientes del modo de huella digital, con las características siguientes:

un medio para ajustar (12) un modo de huella digital predeterminado a partir de la pluralidad de modos de huella digital predeterminados; y

un medio para calcular (14) una huella digital escalable según el modo de huella digital predeterminado ajustado aplicando normas de cálculo según la información de modo para el modo de huella digital ajustado.

17. Dispositivo para caracterizar una señal de audio, con las características siguientes:

un medio para generar una huella digital de búsqueda de la señal de audio utilizando información (10) de modo, que define una pluralidad de modos de huella digital predeterminados, refiriéndose todos los modos de huella digital al mismo tipo de huella digital, proporcionando sin embargo los modos de huella digital huellas digitales diferentes que pueden escalarse con respecto al tiempo y/o la frecuencia, que por un lado se distinguen en su cantidad de datos y por otro lado en su intensidad de identificación para caracterizar la señal de audio, distinguiéndose los modos de huella digital entre sí, porque comprenden información de huella digital separada para una cantidad diferente de subbandas, o presentando la huella digital escalable información de huella digital para una cantidad de bloques temporales dependientes del modo de huella digital, con las características secundarias siguientes:

un medio para ajustar (12) un modo de huella digital predeterminado a partir de la pluralidad de modos de huella digital predeterminados; y

un medio para calcular (14) una huella digital escalable según el modo de huella digital predeterminado ajustado aplicando normas de cálculo según la información de modo para el modo de huella digital ajustado;

un medio para comparar la huella digital calculada con una pluralidad de huellas digitales almacenadas, que representan señales de audio conocidas, para caracterizar la señal de audio, habiéndose generado las huellas digitales almacenadas según uno de la pluralidad de modos de huella digital, con las características secundarias siguientes:

un medio para examinar (24a), si la huella digital de búsqueda y la huella digital de base de datos se han generado según modos de huella digital diferentes;

un medio para convertir (24b) la huella digital de búsqueda y/o la huella digital de base de datos, de modo que las huellas digitales que van a compararse están presentes según el mismo modo de huella digital; y

un medio para realizar (24c) la comparación utilizando las huellas digitales presentes en el mismo modo de huella digital.

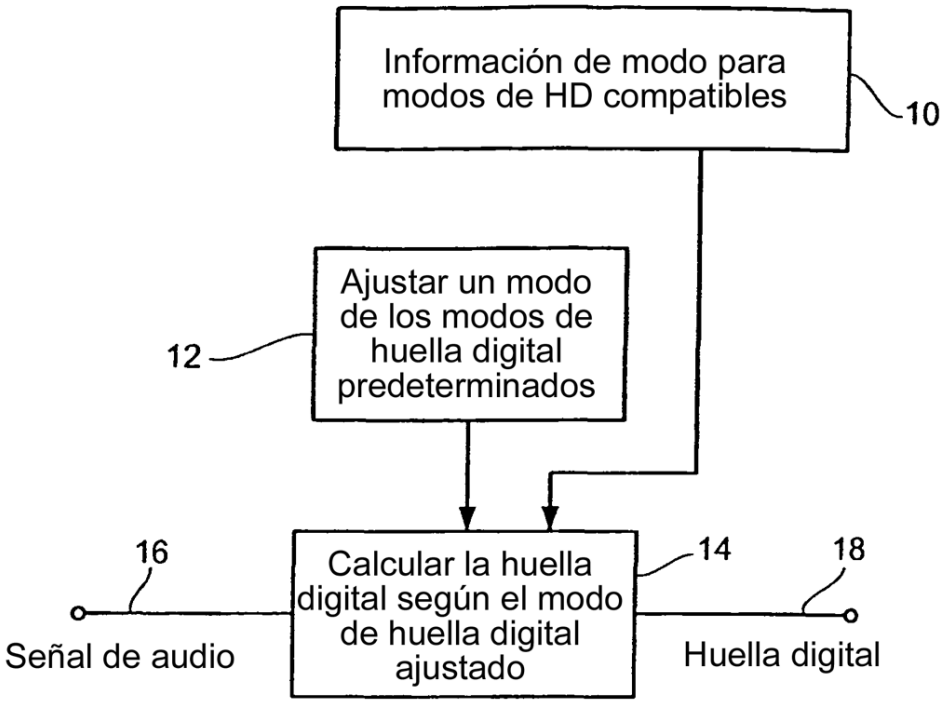
Información de modo para modos de HD compatibles
10
Ajustar un modo de los modos de huella digital predeterminados
12
16
Calcular la huella digital según el modo de huella digital ajustado
14
18
Señal de audio
Huella digital

Fig. 1

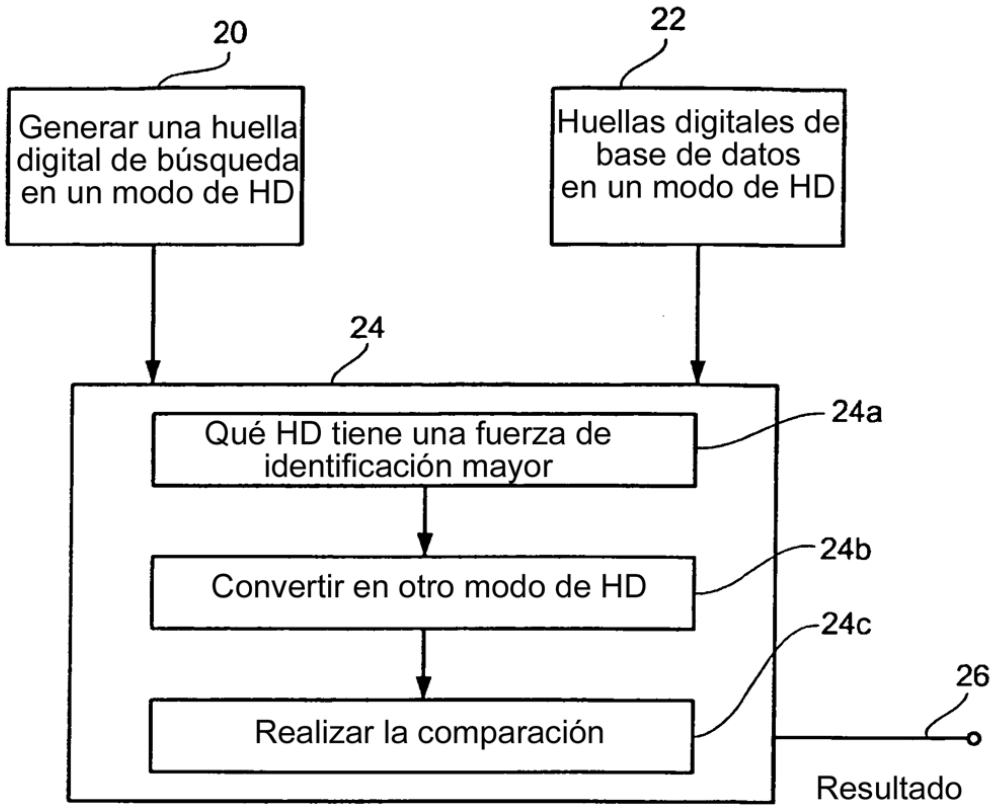
20
Generar una huella digital de búsqueda en un modo de HD
22
Huellas digitales de base de datos en un modo de HD
24
Qué HD tiene una fuerza de identificación mayor
24a
Convertir en otro modo de HD
24b
Realizar la comparación
24c
26
Resultado

Fig. 2

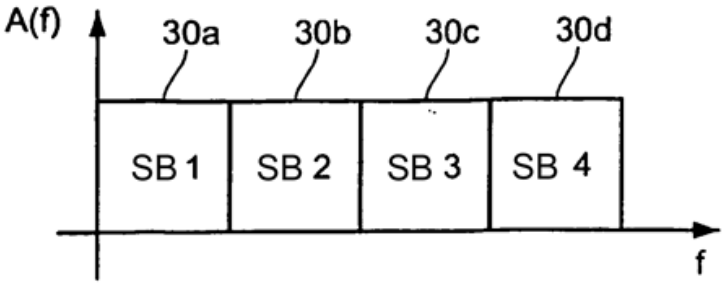
A(f)
30a
30b
30c
30d
SB 1
SB 2
SB 3
SB 4
f

Fig. 3a

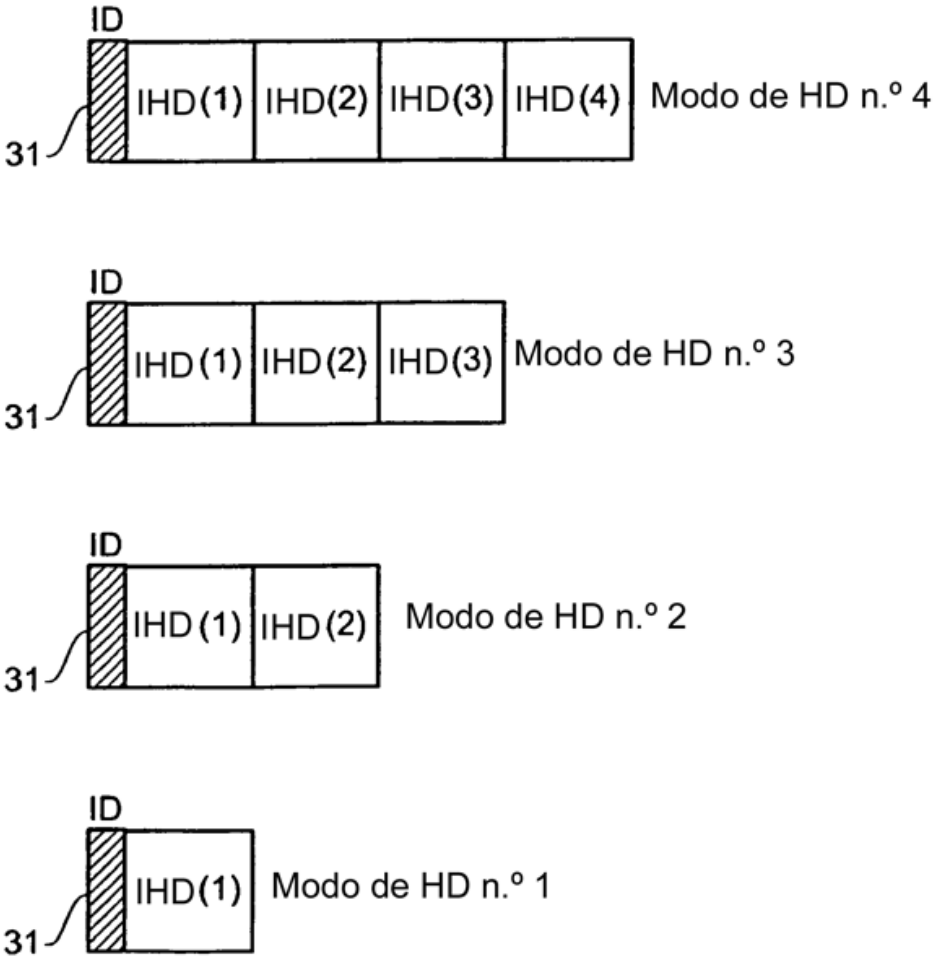
ID
IHD(1)
IHD(2)
IHD(3)
IHD(4)
Modo de HD n.º 4
31
ID
IHD(1)
IHD(2)
IHD(3)
Modo de HD n.º 3
31
ID
IHD(1)
IHD(2)
Modo de HD n.º 2
31
ID
IHD(1)
Modo de HD n.º 1
31

Fig. 3b

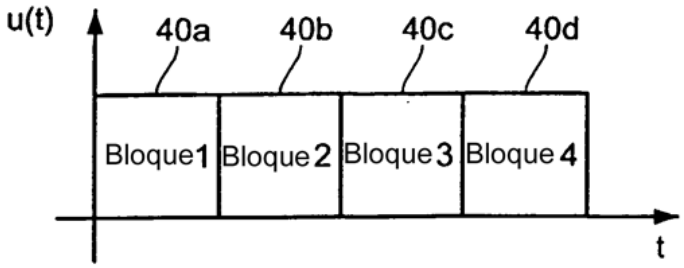
u(t)
40a
40b
40c
40d
Bloque 1
Bloque 2
Bloque 3
Bloque 4
t

Fig. 4a

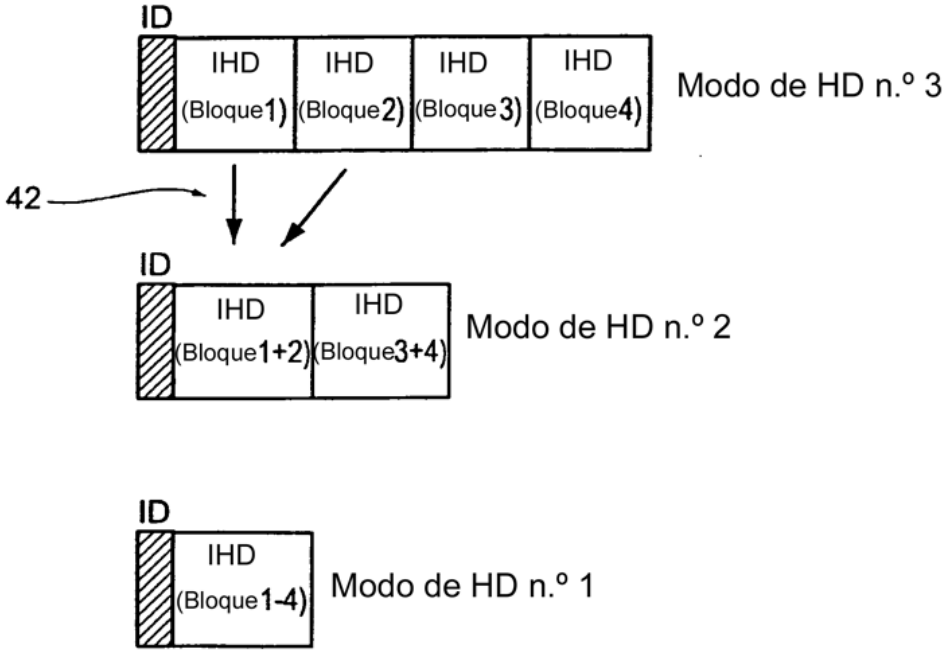
ID
IHD (Bloque 1)
IHD (Bloque 2)
IHD (Bloque 3)
IHD (Bloque 4)
Modo de HD n.º 3
42
ID
IHD (Bloque 1+2)
IHD (Bloque 3+4)
Modo de HD n.º 2
ID
IHD (Bloque 1-4)
Modo de HD n.º 1

Fig. 4b

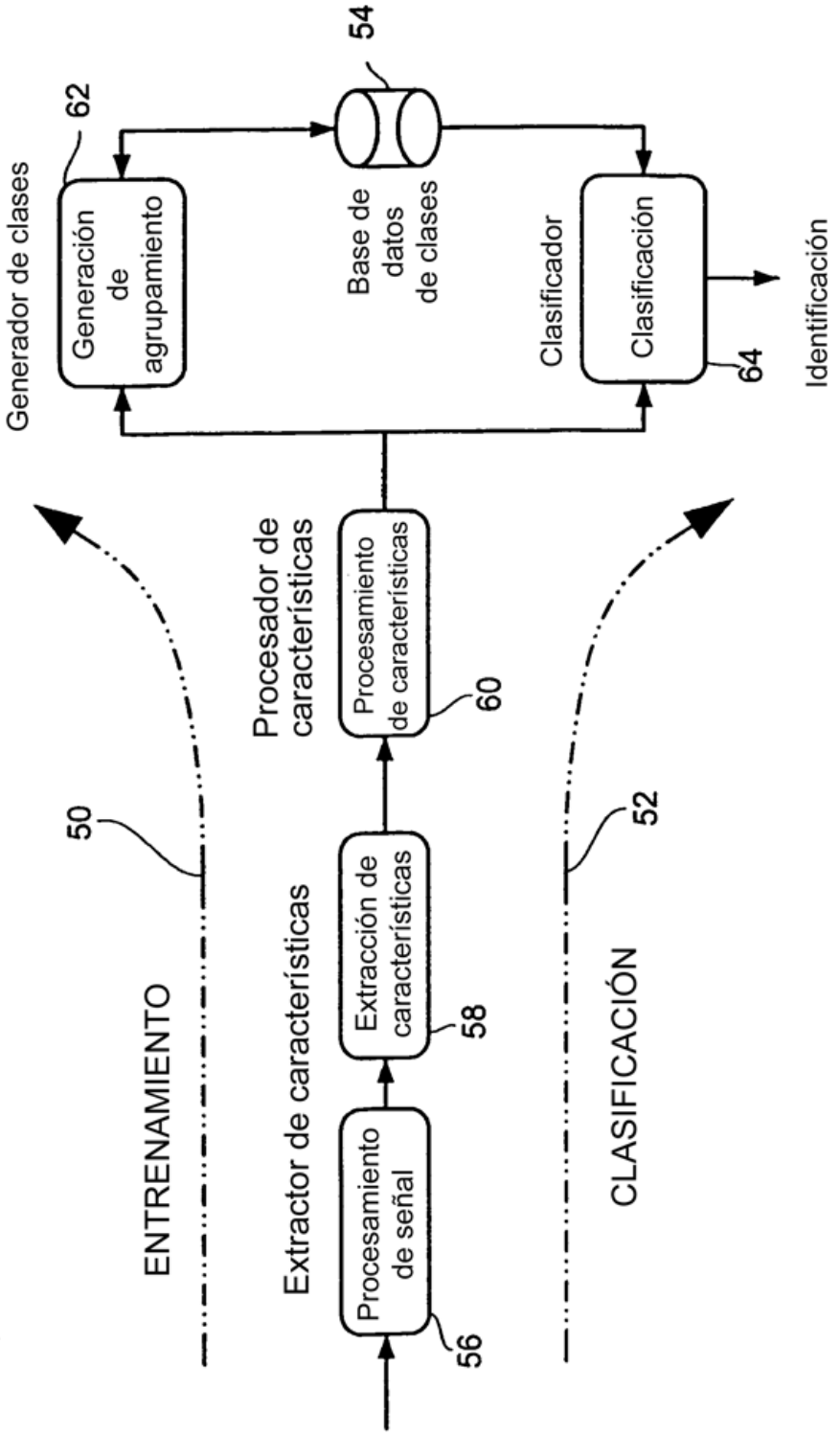
Generador de clases
62
Generación de agrupamiento
54
Base de datos de clases
Clasificador
Clasificación
64
Identificación
Procesador de características
Procesamiento de características
60
Extractor de características
Extracción de características
58
Procesamiento de señal
56
50
ENTRENAMIENTO
52
CLASIFICACIÓN

Fig. 5